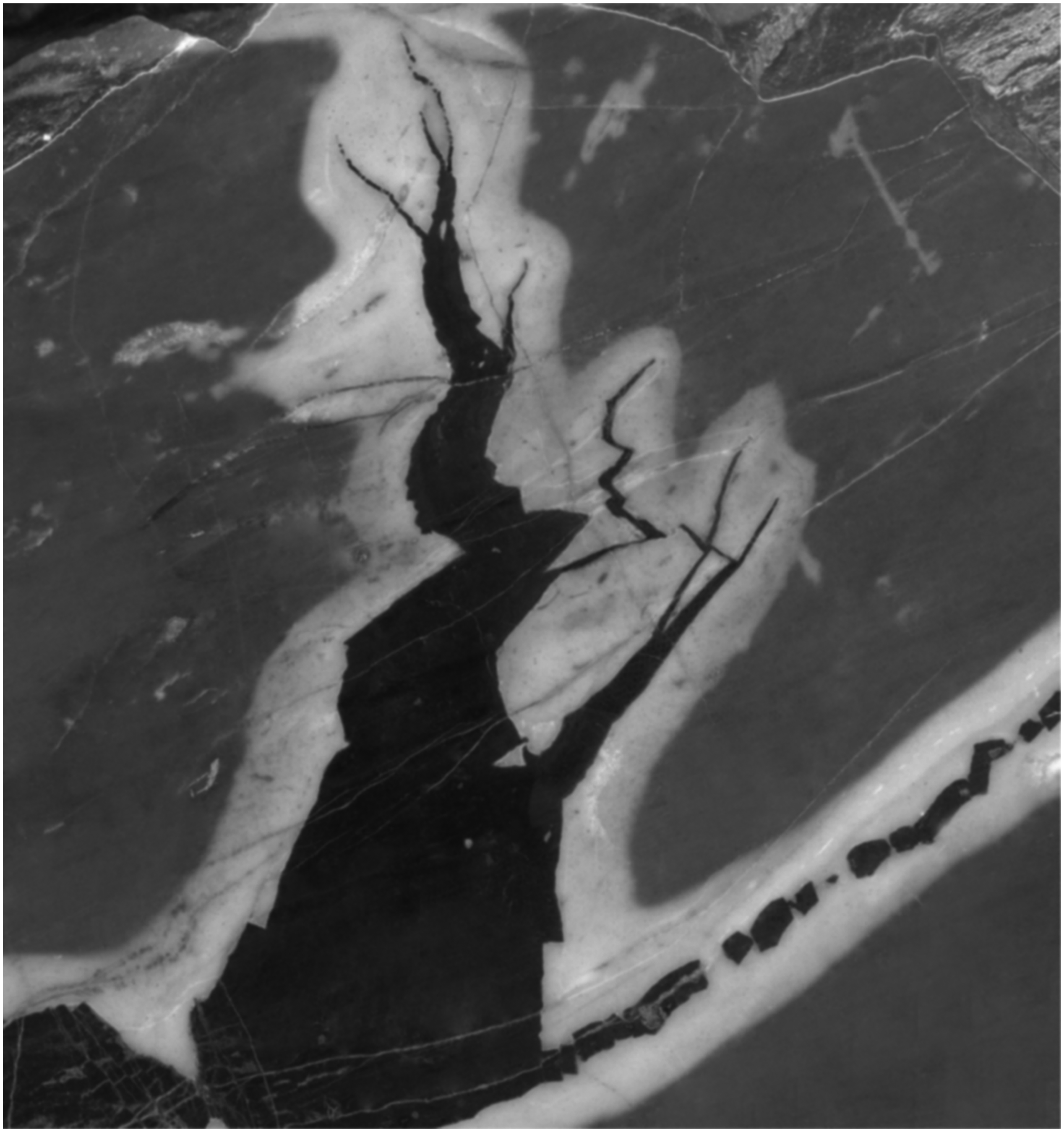
Principles of Igneous and Metamorphic Petrology
Second Edition

This textbook provides a basic understanding of the formative processes of igneous and metamorphic rocks through quantitative applications of simple physical and chemical principles. The book encourages a deeper comprehension of the subject by explaining the petrologic principles rather than simply presenting the student with petrologic facts and terminology. Assuming knowledge of only introductory college-level courses in physics, chemistry, and calculus, it lucidly outlines mathematical derivations fully and at an elementary level, and is ideal for intermediate and advanced courses in igneous and metamorphic petrology.

The end-of-chapter quantitative problem sets facilitate student learning by working through simple applications. They also introduce several widely used thermodynamic software programs for calculating igneous and metamorphic phase equilibria and image analysis software. With over 500 illustrations, this revised edition contains valuable new material on the structure of the Earth's mantle and core, the properties and behavior of magmas, recent results from satellite imaging, and more.

ANTHONY PHILPOTTS is a visiting fellow at Yale University and an adjunct professor at the University of Massachusetts, and has had over 40 years of teaching experience. He has worked on Precambrian massif type anorthosites, pseudotachylites, alkaline rocks, liquid immiscibility in Fe–Ti oxide systems and in tholeiitic magmas. He has been awarded the Peacock Memorial Prize of the Walker Mineralogical Club of Toronto and the Hawley Award of the Mineralogical Association of Canada. He has served as an editor for the *Canadian Mineralogist* and the *Journal of the Canadian Institute of Mining and Metallurgy.*

JAY AGUE is a Professor at Yale University and has been teaching for more than 20 years. He studies fluid flow, chemical reactions, mass transfer, and heat transfer in Earth's crust, focusing on the metamorphic and igneous rocks comprising the deep roots of mountain belts. He was the Senior Editor of the *American Journal of Science* from 1998 to 2008, and has served on the editorial board of the journal *Geology.* He is also Curator of Mineralogy at the Yale Peabody Museum of Natural History.



Gray Ordovician Trenton limestone intruded by aphanitic basalt at the margin of the Cretaceous alkaline gabbro intrusion in Montreal, Quebec, Canada. The intrusion of the network of small basaltic dikes caused brittle failure of the limestone, with fragments frozen in the act of detaching from the walls to form xenoliths in the basalt. Surrounding the igneous rock is a prominent contact metamorphic halo of white marble, where hydrocarbons in the gray limestone were converted to minute crystals of graphite (black specks) in recrystallized calcite. Width of field, 8 cm.

Principles of Igneous and Metamorphic Petrology

Second Edition

ANTHONY R. PHILPOTTS
Yale University

JAY J. AGUE
Yale University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521880060

First published by Prentice-Hall Inc. 1990
Second edition 2009

Printed in the United Kingdom at the University Press, Cambridge

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Philpotts, Anthony R. (Anthony Robert), 1938–
Principles of igneous and metamorphic petrology / Anthony R. Philpotts, Jay J. Ague.
p. cm.
ISBN 978-0-521-88006-0
1. Rocks, Igneous. 2. Rocks, Metamorphic. 3. Petrology. I. Ague, Jay J. II. Title.
QE461.P572 2009
552′.1–dc22
2008040943

ISBN 978-0-521-88006-0 hardback

Anthony Philpotts dedicates this book to his wife, Doreen,
who in the meantime learned to play the piano.

Jay Ague dedicates this book to his family for making everything possible.

Contents

Preface

The second edition of *Principles of Igneous and Metamorphic Petrology* follows the same general approach as the first edition. The book is designed to introduce igneous and metamorphic petrology to those who have completed introductory college-level courses in physics, chemistry, and calculus. Its emphasis is on principles and understanding rather than on facts and memorization. With this approach, it is hoped that students will not only gain a sound understanding of petrology but will develop skills that can be applied to the analysis of problems in many other fields of Earth Science.

Anthony Philpotts took many years to write the first edition of the book, and the thought of preparing a revision was daunting. He was therefore grateful when Jay Ague agreed to share the challenge of producing the revised edition. We both share the same approach to the teaching of petrology, and consequently the new edition retains the flavor of the original while benefiting from the dual authorship.

The first edition of the book was written during the 1980s (published 1990). Since then, the field of petrology has seen significant changes due to both increased knowledge of our planet and new research techniques. In preparing a book of this scope, one cannot help but reflect on the status of petrology as a field of scientific endeavor, especially in light of the trend at many universities to give only survey courses in petrology that are geared toward the environmental science student. From our perspective, the science is still growing and, indeed, the rate of growth of petrologic knowledge and new ideas appears to be increasing. In preparing the book, it was often difficult to decide where to draw the line on what new material to include and what to exclude. We are certain that many readers will find some favorite topic that we have omitted or short-changed. Our goal, however, was to cover the principles of petrology rather than to survey all of petrologic research.

How we study the Earth at both the macro and micro scales has changed dramatically since the 1980s. The whole Earth is now monitored almost continuously. Seismic networks track the motion of tectonic plates while tomography reveals the temperature distributions at depth. Satellites provide almost daily information about surface temperatures, composition of the atmosphere, and minor changes in elevation, which can be used, for example, to forecast volcanic eruptions, even in remote areas. At the other end of the scale, microanalytical techniques now provide isotopic and chemical analyses with micron or even submicron resolution. This has revolutionized the absolute dating of zoned crystals, and chemical gradients in crystals can now be used to investigate the kinetics of petrologic processes. Computers continue to change dramatically the way we investigate rocks. They have made it possible to access petrologic and thermodynamic databases and to use them to solve petrologic problems. The MELTS program, for example, allows one to appreciate an igneous rock in a way that was never previously possible. Similarly, programs such as THERMOCALC allow us to investigate the complexities of possible metamorphic reactions. No longer is a petrographic microscope the only way of examining rocks microscopically. Now, high-resolution images can be used with image analysis software to extract quantitative data about the textures of rocks.

An enormous amount of petrologic information is now available on the Web. At the end of this preface there is a list of web sites of general petrologic interest.Web sites referring to specific topics are given in the text.

As with the first edition, the book is arranged so that it can be used in a two-semester course, or a one-semester course on either igneous or metamorphic petrology. Although it covers both igneous and metamorphic petrology, the contents of chapters allow the text to be used for courses dealing with either of these groups of rocks separately. Such a division, however, is somewhat arbitrary because the principles involved in the formation of these rocks are so similar. Moreover, in the upper mantle, where many petrologic processes have their origins, igneous and metamorphic processes are so interdependent that one cannot be treated without the other. If the book is used for just one of these groups of rocks, cross-references will lead the reader to relevant material covered elsewhere in the text.

The order in which topics are presented in the new edition is essentially the same as in the first, except that diffusion is now treated in Chapter 5 along with heat flow and the chapter on isotopes has been moved from its penultimate position to the middle of the text. The new edition also contains many more illustrations of field relations and photomicrographs, all of which are available in color on the text's web site (www.cambridge.org/philpotts). However, no attempt was made to present a comprehensive treatment of petrography. A companion book, *Petrography of Igneous and Metamorphic Rocks* by A. R. Philpotts (www.waveland.com/Titles/Philpotts.htm), gives the optical properties of all the common rock-forming minerals and the textures of igneous and metamorphic rocks, illustrated in color on a CD-ROM accompanying that book.

As with the first edition, the end-of-chapter problems are an important part of the new edition. Considerable effort has gone into making the problems as instructive as possible. We believe that if users of the book learn as much about petrology in answering the problems as the authors did in creating them, they will be richly rewarded. The book, however, can be used without problems. We would recommend doing as many problems as possible. From our experience, students who have an aversion to mathematics may actually start to enjoy the subject when given the chance to apply it to geological problems. Deserving of special recognition are the students who, over the years, have struggled with these problems; their efforts have allowed us to clarify many of the questions.

Web sites of general petrologic interest

Early history (1565–1835) of ideas on volcanoes, basalt, and geologic time:
www.lhl.lib.mo.us/events_exhib/exhibit/exhibits/vulcan/index.shtml

NASA Earth Observatory: www.visibleearth.nasa.gov

World map of volcanoes, earthquakes, impact craters, and plate tectonics:
www.minerals.si.edu/tdpmap

Earthquakes, near-real time display of earthquakes worldwide: www.iris.edu/seismon
Heat flow from Earth: www.geo.lsa.umich.edu/IHFC

Mineral data

Mineralogical Society of America: www.minsocam.org

Material Properties at High Pressure: www.compres.us

Pegmatites: www.minsocam.org/msa/special/Pig

Volcanoes

Smithsonian Institution Global Volcanism Program: www.volcano.si.edu

Smithsonian's list of volcanoes as an overlay on Google Earth:
www.volcano.si.edu/world/globallists.cfm?listpage=googleearth

Volcano textbook by Robert I. Tilling: http://pubs.usgs.gov/gip/volc

Volcanoworld: http://volcano.und.edu

Volcano hazards

U.S. Geological Survey: www.usgs.gov/hazards/volcanoes
http://volcanoes.usgs.gov

Thermal map of Earth: http://modis.higp.hawaii.edu

SO map of Earth: http://aura.gsfc.nasa.gov

Image analysis

NIH Image: http://rsb.info.nih.gov/ij

USGS CIPW norm: http://volcanoes.usgs.gov/staff/jlowenstern/other/software_jbl.html

Thermodynamic calculations and phase equilibria

MELTS: http://melts.ofm-research.org

THERMOCALC: www.earthsci.unimelb.edu.au/tpg/thermocalc

winTWQ: http://gsc.nrcan.gc.ca/sw/twq_e.php

Phase equilibria: http://ees2.geo.rpi.edu/MetaPetaRen/Software/Software.html
http://serc.carlton.edu/research_education/equilibria/index.html
http://titan.minpet.unibas.ch/minpet/theriak/theruser.html

Acknowledgments

We would like to thank the many users of the first edition of the book who convinced us to prepare a second edition. Without their encouragement, it would not have happened. We would also like to express our gratitude to Cambridge University Press who suggested that they publish the second edition. It has been a pleasure working with a publisher who has not requested that the mathematics be removed from our presentation to make it more palatable.

We cannot possibly acknowledge all of the colleagues who have played important roles in developing our interests in petrology and who have contributed in so many ways to completing this textbook. Anthony Philpotts owes a special debt of gratitude to the following professors who influenced him early in his career: E. H. Kranck of McGill University, and C. E. Tilley, W. A. Deer, I. D. Muir, S. R. Nockolds, and S. O. Agrell of the University of Cambridge.

Jay Ague would like to thank his Ph.D. advisor G. H Brimhall (U.C. Berkeley), his M.S. thesis advisor A. P. Morris (Wayne State University), and his undergraduate advisor S. J. Birnbaum (also at Wayne), for their tireless and inspirational mentoring. Faculty colleagues at Yale have been a constant wellspring of scientific ideas and support, and Ague would especially like to acknowledge D. M. Rye, B. J. Skinner, M. T. Brandon, J. J. Park, and K. K. Turekian in this regard. Discussions and collaborations with research scientists J. O. Eckert and E. W. Bolton are also much appreciated. Numerous undergraduates, graduate students, and post docs have contributed in countless ways to the material presented in this book, including E. F. Baxter, C. M. Breeding, C. J. Carson, R. L. Masters, J. L. M. van Haren, and D. E. Wilbur. In addition, students C. E. Bucholz, X. Du, T. V. Lyubetskaya, and S. H. Vorhies read draft chapters and made numerous suggestions for improvement. R. W. White (Mainz) provided valuable help with THERMOCALC. Finally, Ague would like to gratefully acknowledge the National Science Foundation, the Department of Energy, and Yale University for research support.

We would both like to thank our families who have been totally neglected while we prepared this edition. Without their support, the second edition of *Principles of Igneous and Metamorphic Petrology* would never have been completed.

List of units

Basic Units

This text uses units of the *Système International* (SI). The basic units are

Length	meter (m)
Mass	kilogram (kg)
Time	seconds (s)
Temperature	kelvin (K)

Prefixes

For convenience these units can be preceded by the following prefixes:

pico (p)	10^{-12}
nano (n)	10^{-9}
micro (μ)	10^{-6}
milli (m)	10^{-3}
kilo (k)	10^{3}
mega (M)	10^{6}
giga (G)	10^{9}
tera (T)	10^{12}

Derived Units

Quantity	**Unit**	**Equivalent**
Force	newton (N)	kg m s^{-2}
Pressure	pascal (Pa)	N m^{-2}
Energy	joule (J)	N m
	joule (J)	Pa m^{3}
Power	watt (W)	J s^{-1}
Viscosity		Pa s
Kinematic viscosity		m^{2} s^{-1}

Other Common Units

Length	centimeter (cm)	10^{-2} m
	angstrom (Å)	10 nm
Mass	gram (g)	10^{-3} kg
Force	dyne	g cm s^{-2}
Heat Flow Unit	(HFU)	10^{-6} cal cm^2 s^{-1}
		41.84 mW m^{-2}
Heat Generation Unit	(HGU)	10^{-13} cal cm^{-3} s^{-1}
		0.4184 μW m^{-3}
Pressure	bar (b)	10^6dyne cm^{-2} (10^5 Pa)
	atmosphere (atm)	1.013 25 bar ($\sim 10^5$ Pa)
Energy	calorie (cal)	4.184 J
Parts per million (ppm)		kg/10^6 kg

Commonly Used Constants

Gas constant	(R)	8.3144J mol^{-1} K^{-1} (1.9872 cal mol^{-1} K^{-1})
Avogadro's number	(N_0)	6.022×10^{23} mol^{-1}
Stefan–Boltzmann constant	(σ)	5.67×10^{-8} W m^{-2} K^{-4}
Boltzmann's constant	(k)	1.3806×10^{-23} J K^{-1}
Acceleration of gravity at Earth's surface		$\sim$9.8 m s^{-2}

1 Introduction

1.1 PETROLOGY AND ITS SCOPE

Petrology is the science dealing with the description, classification, modes of occurrence, and theories of the origins of rocks. Its emphasis is commonly chemical and mineralogical, but it draws heavily on many disciplines, including the basic physical sciences, mathematics, geophysics, structural geology, and geochemistry. Its tools range from the simple hammer and hand lens, to sophisticated devices such as the electron microprobe or the laboratory equipment capable of reproducing conditions deep within the Earth. Its goal is to provide an understanding of the great diversity of rocks found on the surface of the Earth (and other planets), and to provide insight into the nature of those materials within the Earth that are not accessible to direct observation but play such important roles in the Earth's history.

Rocks can be divided into three main groups: igneous, sedimentary, and metamorphic. Those formed from the solidification of molten material are termed *igneous*, whereas those that originate from the deposition of material from water or air are termed *sedimentary*, and those formed from a previously existing rock by some process of change are termed *metamorphic*.

The study of igneous and metamorphic rocks, the subject of this book, is commonly treated separately from the study of sedimentary rocks, mainly because of the different approaches used. Sedimentary rocks are formed by processes that, for the most part, are observable on the surface of the Earth. Careful examination of present-day environments of deposition can, therefore, provide information on the origins of most sedimentary rocks. Igneous and metamorphic rocks, on the other hand, are formed largely by processes operating within the Earth and therefore not directly accessible to observation; the origin of these rocks must, consequently, be deduced through physical–chemical arguments. Also, at the higher temperatures existing within the Earth, reactions proceed more rapidly than they do on the surface, and thus principles of chemical equilibrium are more applicable to the study of igneous and metamorphic rocks than they are to most sedimentary ones.

Petrologic studies fall into two general categories: the identification and classification of rocks, and the interpretation of these data and the generation of theories on the origin of rocks. The early emphasis in petrology, as in other natural sciences, was on description and classification. There were, nonetheless, many lively discussions concerning the origins of rocks, such as that in the early nineteenth century between Werner and his student von Buch on whether basalt was a sedimentary or volcanic rock. An excellent summary of early ideas on the origin of basalt and other igneous rocks can be found at www.lhl.lib.mo.us/events_exhib/exhibit/exhibits/vulcan/index.shtml. This web site provides illustrations and brief descriptions from 66 rare books and journals, published between 1565 and 1835, held in the Linda Hall Library of Science, Engineering & Technology in Kansas City, MO. Most of the early work, however, involved the cataloging of the constituents of the Earth's crust. During the latter half of the nineteenth and early part of the twentieth centuries vast amounts of petrologic data were collected, from which came an enormous number of rock names and many different classifications. Despite the surfeit of names, generalizations concerning rock associations and mineral assemblages did emerge, which, in turn, allowed for simplifications in the classifications. Many different rocks that had previously been given separate names could be considered varieties of a single type, and this naturally led to theories explaining the associations. Fortunately, the rock-naming era is over, and modern petrology employs only a small number of rock names.

With recent investigations of the ocean floors, the inventory of rock types available for study in the Earth's crust is almost complete, and today most petrologists are concerned mainly with the genesis of rocks. This change in emphasis has been stimulated by development of experimental techniques that allow us to imitate, in a limited way, rock-forming conditions and processes within the Earth's crust, mantle, and even the core (see Consortium for Materials Properties Research in Earth Sciences www.compres.us). In fact, with such techniques it is possible to investigate the petrology of parts of the Earth that, while being of great importance as source regions for many igneous rocks, were virtually unavailable for examination and classification by the early petrologists.

Most processes involved in the formation of igneous and metamorphic rocks occur within the Earth and hence are not subject to direct observation (volcanic eruptions are obvious exceptions). Ideas on the origin of these rocks are, therefore, based on interpretations of field observations in the light of experimental and theoretical studies. Nature provides fragmentary evidence of events and processes that have formed rocks, and it is up to the petrologist to assemble this evidence into a coherent story. The task is similar to that faced by a

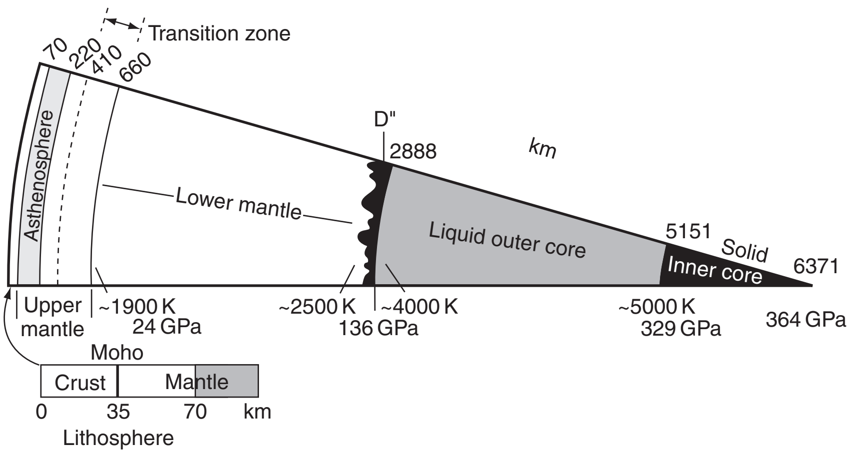


Fig. 1.1 Major divisions of the Earth based on seismic data (Masters and Shearer, 1995; Lay *et al*., 2004). Temperatures are from Boehler (2000). Note that the lithosphere, which is the layer in which almost all rocks that are exposed on the Earth's surface are formed, constitutes a very small fraction of the entire Earth.

detective who, on arriving at a crime scene, must deduce what happened from the limited evidence that remains. Unlike the legal case, however, which requires that a verdict be reached in a court of law, petrologic questions rarely have absolute verdicts. Indeed, as new data and new research techniques become available, interpretations frequently have to be revised, and this, in itself, makes petrology an active and interesting field for study.

Before beginning a discussion of igneous petrology, it will be useful to review briefly the major structural units of the Earth and the distributions of pressure and temperature within it. This will provide a framework for later discussions of rock-forming processes.

1.2 MAJOR STRUCTURAL UNITS OF THE EARTH

The division of the Earth into three major structural units, *crust*, *mantle*, and *core* (Fig. 1.1), was made early in the history of seismology on the basis of major discontinuities in both compressional (P) and shear (S) wave velocities (e.g. Masters and Shearer, 1995). The discontinuities are best explained as boundaries separating chemically and mineralogically distinct zones. Although their compositions cannot be determined with certainty, reasonable estimates can be made from seismic velocities, the mass and moment of inertia of the Earth, and solar and meteoritic abundances of elements. For a more detailed section through the Earth, refer to Figure 23.4.

The crust contains high concentrations of alkalis, calcium, aluminum, and silicon relative to solar abundances. In continental regions, where the crust is approximately 35 km thick, these elements are so abundant that when combined to form the common minerals quartz and feldspar, the rock approximates the composition of granodiorite. Many light elements and certain heavy ones, such as uranium, thorium, and zirconium, which have difficulty substituting into the structures of common minerals, are also concentrated in this zone. Because some of these elements are radioactive and generate heat, temperature measurements at the Earth's surface can be used to show that this heat-generating granitic zone must be limited to the upper 10 to 20 km of the continental crust (Section 1.4 and Problem 1.4). Increasing seismic velocities with depth in continents indicate the presence of rocks with higher concentrations of iron and magnesium at depth; these are probably basaltic in composition. Basaltic rocks constitute most of the oceanic crust which is about 6 km thick (see Fig. 1.2).

The Mohorovičić discontinuity (Moho or M discontinuity), at the base of the crust, marks a sharp increase in seismic velocities, which can be accounted for by the disappearance of feldspar, the most abundant mineral in crustal rocks. Below the Moho, the rock is thought to consist mainly of olivine and pyroxene with lesser amounts of spinel and garnet and thus is called a peridotite. Samples of this rock are brought to the Earth's surface by certain types of explosive volcanic eruption and confirm the seismological deductions as to its composition.

The disappearance of S waves at the core–mantle boundary indicates the presence of liquid below this. If the Earth were initially formed from accreting chondritic meteorites, which is the favored model, and with its known internal distribution of mass, the core is believed to be composed largely of iron and some nickel. However, high-pressure experiments with diamond anvils and shock waves indicate that the core's density must be decreased slightly by the presence of small amounts of lighter elements such as silicon, oxygen, sulfur, and hydrogen. The Earth's magnetic field is generated by convection within this liquid. With increasing depth and pressure, the liquid undergoes a phase change to form the solid inner core; that is, the inner–outer core boundary marks the floor of a giant magma chamber, which is slowly solidifying from the bottom up as the Earth loses heat. Seismic velocities in the inner core are found to be faster in a north–south direction than in an equatorial path, indicating that the inner core is crystallographically and/or structurally anisotropic. The inner core may also rotate faster than the mantle, which may play a role in generating the Earth's magnetic field.

In addition to these traditional structural units, modern seismologists detect many other discontinuities with which they further subdivide the Earth (Fig. 1.1). Moreover, by

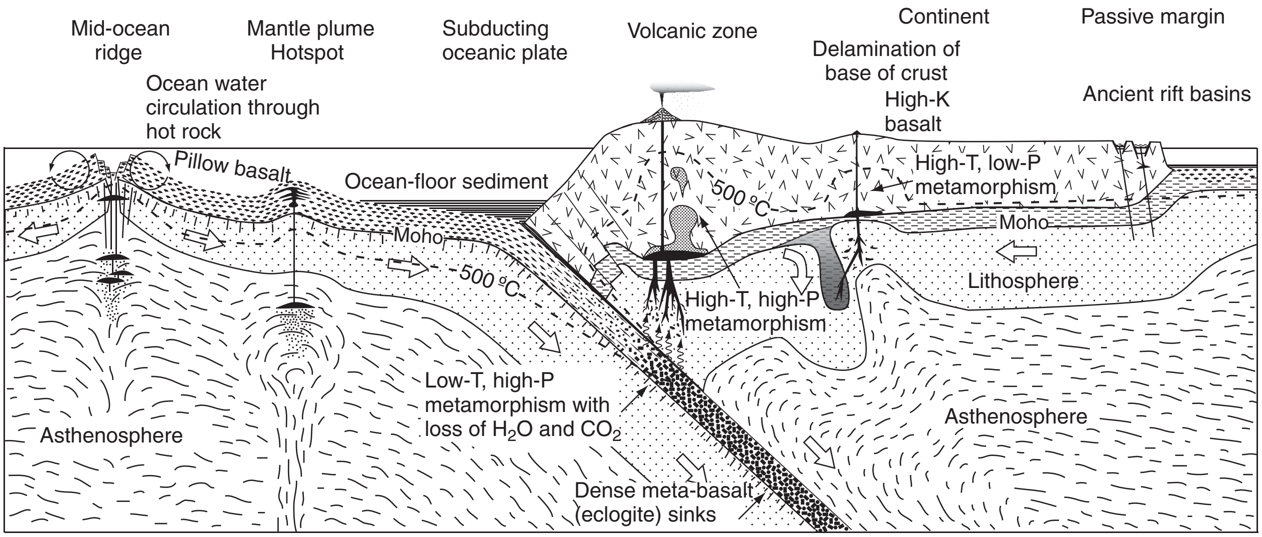


Fig. 1.2 Schematic, vertically exaggerated section through the lithosphere and asthenosphere showing an oceanic plate moving away from a spreading axis and being subducted beneath a continental plate. Most igneous rocks are formed either at spreading axes or above subduction zones, and most metamorphic rocks are formed in the vicinity of convergent plate boundaries. Slabs at the base of the continental lithosphere may be metamorphosed into rock that is dense enough to delaminate and sink into the mantle. As it sinks, the rising asthenosphere causes heating and generation of igneous and metamorphic rocks in the crust above. The range of temperatures and pressures rocks experience are determined largely by plate motions. Note how the depth (pressure) of the 500 °C isotherm varies across the section as a function of its position in a plate.

analyzing seismic recordings of earthquakes from arrays of stations around the planet, the three-dimensional distribution of seismic velocities can be determined. These tomographic studies reveal that the traditional zones within the Earth exhibit considerable complexity.

Of particular importance to petrology and plate tectonics is a zone extending from depths of approximately 70 to 250 km in which shear-wave velocities markedly decrease; this zone is known as the *low-velocity layer*. Because it has considerably less strength than overlying layers it is also referred to as the *asthenosphere* or weak zone. The more rigid overlying rocks, which include the upper part of the mantle and the crust, constitute the *lithosphere*. The reduced velocities and strengths of rocks in the asthenosphere are thought to result from small degrees of partial melting, possibly as much as 10% in regions of high heat flow. This partial melt is an important source of magma and a lubricant to ease the tectonic movements of the lithospheric plates.

Between 410 and 660 km, seismic velocities increase rapidly. Most of this change occurs at two major discontinuities. One, at 410 km, occurs at pressures where laboratory experiments indicate that the crystal structure of olivine changes to that of a distorted spinel, and the other, at 660 km, is at pressures where perovskite and simple oxides predominate. Indeed, below 660 km, most silicate minerals are converted to the dense perovskite structure and magnesiowüstite ([Mg,Fe]O). Between 660 km and the core–mantle boundary, velocities increase in a slower and steadier manner. This is known as the lower mantle. Little is known about its composition, yet it comprises 56% of the planet. Tomographic studies reveal considerable lateral variation in seismic velocities within the mantle, which may be due to changes in temperature or composition.

At the base of the lower mantle is a 200- to 350-km-thick heterogeneous thermochemical boundary layer separating the mantle from the core (Lay *et al*., 2004). This layer, designated D″ (*D Double Prime*), is marked by decreased gradients in seismic velocities with depth. Within this layer, at depths that range from 150 to 350 km above the core–mantle boundary, is a prominent discontinuity marked by an increase in seismic velocities, which may be caused by a still higher density form of the perovskite structure known as post-perovskite (Murakami *et al*., 2004). Shear waves become polarized in passing through D″ into faster horizontal – and slower vertical – traveling components. This anisotropy indicates the presence of preferred orientations either of crystals or of chemically and physically distinct structures, such as bodies of melt. At the base of D″ is a 5- to 40-km-thick ultralow-velocity zone, which is almost certainly partially melted by heat released from the rapidly convecting core beneath.

The D″ layer may play important roles in the dynamics and evolution of our planet. For example, some of the heterogeneity in D″ can be correlated with important features in the lithosphere. Seismic velocities in D″ are generally lower beneath surface hot spots, such as Hawaii, and higher in regions beneath subducted slabs, with lower and higher velocities possibly corresponding to hotter and cooler regions, respectively. Thus, major tectonic features in the lithosphere may owe their origin and location to the way in which heat is transferred out of the core through the D″ layer. Although the chemical makeup of D″ is uncertain, it has been proposed as a reservoir of ancient material, possibly dating back to the accretion of the planet (Tolstikhin *et al*., 2006). It may also be a “graveyard” for subducted slabs of lithosphere or even a chemical reaction zone between the core and mantle.

In later chapters when large numbers are used to describe the depths and pressures under which magmas are generated, it will be easy to lose sight of the fact that we will be dealing with only the extreme outer skin of the Earth. Almost all terrestrial rocks that geologists are able to examine are formed within the upper part of the lithosphere. We therefore can examine directly only a minute fraction of the whole Earth. For perspective, then, readers are advised to refer, from time to time, to Figure 1.1.

Despite the small relative volume of the lithosphere, it contains a great variety of rocks and structures. Much of its complexity has been elucidated by modern plate tectonic theories, which show the lithosphere to be a dynamic part of the Earth undergoing constant change as a result of heat transfer from the planet's interior to its surface. Most igneous and metamorphic rocks owe their origins directly to plate tectonic processes.

Figure 1.2 summarizes the principal components of the lithosphere in terms of their involvement in plate tectonics. The surface of the Earth is divided into ten major lithospheric plates and several smaller ones. Each plate moves as a rigid unit with respect to adjoining ones. Where plates move apart, new material must fill the intervening space. Thus, along mid-oceanic ridges, which are divergent plate boundaries, the rising asthenosphere decompresses and partially melts to form basaltic magma that buoyantly rises and solidifies to create new lithosphere. Elsewhere material must be consumed if the Earth's circumference is to remain constant. This occurs at convergent plate boundaries where cold lithospheric slabs are subducted into the mantle. Material at the base of the crust may also be converted into metamorphic minerals that are so dense that the lower part of the crust delaminates and sinks into the mantle. Plates may also slip past one another along transform faults, as happens on the San Andreas fault.

Lateral and vertical movements of lithospheric plates produce significant lateral variations in the temperature of rocks at equivalent depths in the lithosphere. In Figure 1.2, a line joining all points at 500 °C (the 500 °C isotherm) is traced through a typical section of lithosphere. Beneath a stable continental region this isotherm is at a depth of approximately 40 km, a position reflecting a balance between the rate at which heat rises into the lithosphere from the Earth's interior, the rate of heat production from radioactive elements within the lithosphere, and the rate at which heat is conducted to the Earth's surface (Section 1.6). At mid-ocean ridges, the emplacement of large volumes of basaltic magma, which has risen from the hotter asthenosphere, moves the isotherm much closer to the Earth's surface. As the new hot lithosphere moves away from the ridge, its temperature decreases due to conductive cooling and the circulation of cold ocean water. The circulating water also hydrothermally alters the rocks. The lithosphere thickens, becomes denser, and sinks deeper in the asthenosphere, thus causing oceans to deepen away from ridges (Stein and Stein, 1992). At subduction zones, the sinking of cool lithospheric plates into the asthenosphere depresses isotherms to considerable depths, and as a result, the ocean floor rocks are converted into low-temperature high-pressure metamorphic rocks. The metamorphism releases water and carbon dioxide, which rise into the overlying mantle wedge where they (especially water) behave as a flux that causes fusion and formation of basaltic and andesitic magma. Because these magmas are less dense than the surrounding mantle, they buoyantly rise at least to the base of the continental crust. Here, dense basaltic magma may intrude or underplate the crust. As it cools, it can melt the overlying rocks to form bodies of granitic magma that rise through the crust to form batholiths and possibly erupt on the surface as rhyolite. Andesitic magma may continue all the way to the Earth's surface to form large cone-shaped volcanoes. The rise of all this magma into the base of the continental crust above subduction zones deflects isotherms upward, which, in turn, causes high-temperature–high-pressure metamorphism and possibly melting of the crustal rocks (Marsh, 1979; Annen *et al*., 2006). If basaltic rocks at the base of the crust are metamorphosed into denser mineral assemblages, slabs of this rock are thought to be able to detach (delaminate) and sink into the mantle (Kay and Kay, 1993). The asthenosphere that rises to replace the sunken slab is hot, and consequently the isotherms above a zone of delamination are raised. Melting may occur in the ascending asthenosphere or in the sinking slab, which can produce potassium-rich basalts. The rise of the isotherms and the intrusion of magma can cause high-temperature–low-pressure metamorphism above zones of delamination. Well away from zones of subduction or delamination beneath the stable continent, the 500 °C isotherm returns to its undisturbed steady-state depth of approximately 40 km.

The variation in the depths of isotherms throughout the lithosphere gives rise to a wide range of possible pressure–temperature conditions. These, in turn, control the processes by which rocks are generated. We therefore turn our attention to the problems of how we determine pressure and temperature within the lithosphere.

1.3 PRESSURE DISTRIBUTION WITHIN THE EARTH

Rocks under high pressure do not have high shear strength, especially over long periods of time. Instead, they flow as if they were extremely viscous liquids. The depth at which this occurs is referred to as the *brittle–ductile transition*. In many parts of the Earth, this transition occurs at a depth of ~15 km, and this is why many earthquakes are restricted to depths shallower than this. Earthquakes do occur at greater depths, as for example in subduction zones, but here plate motion has taken relatively cold crustal rocks to depth, and before the rocks have time to heat, they can undergo brittle failure. Metamorphic reactions that liberate fluids can also build up fluid pressures rapidly enough to exceed the tensile strength of low permeability rocks and cause fracturing in the middle crust (Ague *et al*., 1998). Strain associated with the intrusion of magma can cause fracturing, which has been recorded seismically to depths of 60 km. Despite this fracturing, the

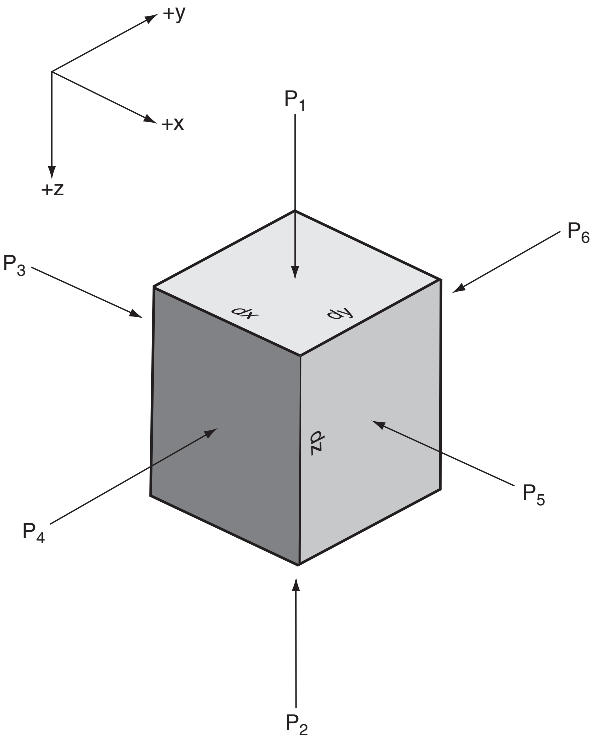


Fig. 1.3 Pressures exerted on the faces of a small volume (d*x* d*y* d*z*) of rock as a result of the load of the surrounding rocks. Note that the positive direction of *z* is downward (depth).

tendency of rocks to flow in response to long-term stresses allows us to calculate the pressure for any particular depth in the Earth in the same manner that hydrostatic pressure can be calculated for any depth in water. Indeed, the term *hydrostatic* is often used loosely instead of the correct term *lithostatic* to describe pressures in the Earth resulting from the load of overlying rock. At shallow depths, where open fractures exist or where the rock has high permeability, the pressure on the rock may actually be higher than that on the water in fractures or pores; that is, the lithostatic pressure may be greater than the hydrostatic pressure. In contrast, water exsolving from crystallizing magma or generated by metamorphic reactions may cause the hydrostatic pressure to exceed the lithostatic pressure and possibly cause explosions or fracturing and veining.

We will now derive the simple relation between depth and lithostatic pressure in the Earth, assuming that rocks are capable of flowing under high pressures. Although this derivation may appear trivial, it is included here to introduce a particular way of tackling a problem. This same approach will be encountered in later chapters dealing with more complicated problems involving fluid flow, heat transfer, and diffusion.

Consider the forces acting on a small volume of rock at some depth in the Earth. For convenience this volume is given the dimensions dx, dy, and dz (Fig. 1.3), where x and y are perpendicular directions in a horizontal plane, and z is vertical with positive values measured downward (i.e. z = depth). The surrounding rock exerts pressures P_1 to P_6 on each face of the volume. We further stipulate that the pressure at the center of this volume is P, and that there is a vertical pressure gradient, dP/dz.

If the volume does not move horizontally, the forces resulting from the pressures P_3, P_4, P_5, and P_6 must balance one another and need not be considered further. The forces acting in a vertical direction result from the pressures P_1 and P_2 and the acceleration of gravity acting on the mass of the small volume (remember that according to Newton's second law, a force is measured by the mass multiplied by the acceleration that the force produces; $F = ma$). Again, if the volume does not move, the upward and downward forces must balance, or as is commonly expressed mathematically, the forces acting in any particular direction must sum to zero. Recalling that the force resulting from a pressure is simply the pressure times the area on which it acts, we can sum the forces acting downward (positive direction of z) on the volume as follows:

$$P_1\ \mathrm{d}x\ \mathrm{d}y + (-P_2\ \mathrm{d}x\ \mathrm{d}y) + mg = 0$$

where m is the mass of the small volume and g is the acceleration of gravity at the depth considered. Note that the force resulting from P_2 is negative because it operates in an upward direction.

The mass of the small volume is obviously dependent on its size. We can, however, express mass in terms of the dimensions of the volume and its density (ρ), which allows the force of gravity to be expressed as follows:

$$mg = \rho\ \mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\ g$$

The two pressures P_1 and P_2 can be related to the pressure P at the center of the volume by the pressure gradient (dP/dz) within the volume. Because the distance from the center of the volume to the top or bottom is $^1/_2$dz, pressures P_1 and P_2 are given by

$$P_1 = P - \frac{1}{2}\mathrm{d}z\left(\frac{\mathrm{d}P}{\mathrm{d}z}\right) \quad \text{and} \quad P_2 = P + \frac{1}{2}\mathrm{d}z\left(\frac{\mathrm{d}P}{\mathrm{d}z}\right)$$

We can now write the sum of the forces acting downward on the volume as

$$\left[P - \frac{1}{2}\mathrm{d}z\left(\frac{\mathrm{d}P}{\mathrm{d}z}\right)\right]\mathrm{d}x\,\mathrm{d}y - \left[P + \frac{1}{2}\mathrm{d}z\left(\frac{\mathrm{d}P}{\mathrm{d}z}\right)\right]\mathrm{d}x\,\mathrm{d}y$$
$$+\rho\,\mathrm{d}x\ \mathrm{d}y\ \mathrm{d}z\ g = 0$$

which simplifies to

$$\frac{\mathrm{d}P}{\mathrm{d}z} = \rho g \tag{1.1}$$

This indicates that the pressure gradient at any particular depth is simply the product of the density of the material and the acceleration of gravity at that point. To obtain the pressure at that depth we have simply to integrate the expression

$$\int_{P_0}^{P_z} \mathrm{d}P = \int_0^z \rho g\,\mathrm{d}z$$

from the surface of the Earth to the depth, z, of interest. The pressure at the Earth's surface of 1 atm is so small compared with those at depth that it can be taken as zero. We can then write

$$P_z = \int_0^z \rho g \, \mathrm{d}z$$

To integrate the right-hand side of this equation, we must first determine whether ρ and g are constants or functions of depth (z). Most minerals and magmas are rather incompressible. Thus, no serious error is introduced by assuming that their densities remain constant, at least under the pressures encountered in the crust and upper mantle. However, there are large amounts of data on compressibilities of minerals and magmas (e.g. Bass, 1995), so density can be expressed as a function of pressure, if so desired (see Problem 3.1 in Chapter 3). Here, we will take the density to be constant.

Variation in the value of g with depth is more complex. At points above the surface of the Earth, its value varies inversely as the square of the distance from the center of the Earth. Within the Earth, however, only the underlying mass contributes any net gravitational force; those forces due to the overlying shell of rock sum to zero and hence produce no net gravitational attraction. For example, if the planet were hollow, as postulated by Abraham Gottlob Werner in his theory on the origin of Primary and Secondary rocks, you would have been weightless if you could have entered its interior. With increasing depth in the Earth, less and less mass remains to cause attraction. But this is partly offset by the increasing density of rocks with depth. The result is that although the value of g goes to zero at the Earth's center, throughout the crust and upper mantle it remains approximately constant, and will be taken as such for our purposes.

Taking the density of material and the acceleration of gravity as constants, we integrate the preceding equation to obtain

$$P = \rho g z \tag{1.2}$$

To illustrate this equation, let us determine the pressure at the base of a 35-km-thick granitic crust with an average density of 2800 kg m^{-3}. Substituting these values into Eq. (1.2) and being careful to list units, we obtain

$$P = 2800 \times 9.80 \times 35 \times 10^3 \quad \text{kg m}^{-3} \times \text{m s}^{-2} \times \text{m}$$

But a kg m s^{-2} is a unit of force, the *newton* (N). The units therefore reduce to N m^{-2}, which is the unit of pressure known as a *pascal* (Pa). The pressure at the base of a 35-km-thick crust is therefore 0.96×10^9 Pa or 0.96 GPa. In cgs units this would be 9.6 kilobars:

$$1 \text{ bar} = 10^6 \text{ dyne cm}^{-2} = 0.9869 \text{ atmosphere}$$

The pressure at the base of the 35-km-thick crust in this simple calculation is approximately 1 GPa. By taking into account details of the density distribution within the crust, a more accurate determination of pressure can, of course, be obtained (see Problem 1.1). Also, by making reasonable assumptions about the mineralogy of the deep Earth's interior and using ultrahigh-pressure experimental data on the compressibility of minerals (Bass, 1995) the density of the Earth's mantle and core can be deduced, from which we calculate the pressure (Fig. 1.1). At the base of the upper mantle (660 km), the pressure is 24 GPa (Masters and Shearer, 1995). At the core–mantle boundary it is 136 GPa, but at the inner–outer core boundary it has risen to 329 GPa and at the center of the Earth, it is 364 GPa.

1.4 TEMPERATURE GRADIENTS AND HEAT FLOW IN THE LITHOSPHERE

Temperatures in the Earth cannot be determined as easily as pressures. Deep drill holes probe only the top few kilometers of the crust, and extrapolating temperatures measured in these holes to lower parts of the crust and the upper mantle is fraught with difficulties. It is not surprising to find that the literature contains numerous, significantly different estimates of the geothermal gradient. Much interest in the internal temperatures of the Earth was piqued in the nineteenth century by attempts, such as that of Lord Kelvin (Thomson, 1863), to calculate the age of the Earth based on the length of time to cool from an initially molten state. The age was wrong, because his model did not take into account convection in the mantle (England *et al*., 2007). These early workers were unaware of a major source of heat in the Earth, that produced by radioactive decay. Today, radiogenic heat is certainly taken into account; nonetheless, markedly different gradients can still be calculated, depending on the assumed distribution of the heat-generating elements (Problem 1.5) and the amount of heat thought to be passing through the surface of the Earth. Fortunately, high-pressure experimental investigations of the melting of rocks provide rather tight constraints on the near-surface geothermal gradient, and for temperatures in the deep mantle and core, they provide the only constraints (Boehler, 2000).

Because the boundary between the inner and outer core is interpreted to be the transition between solid and liquid iron diluted with nickel and several lighter elements, high-pressure melting experiments can be used to set a temperature for this boundary. Achieving these temperatures at the pressure of 329 GPa, however, is difficult, and at present can be reached only in shock experiments where a projectile is fired in a sophisticated two-stage gas gun at the experimental charge (Brown and McQueen, 1986). Results of lower-pressure static experiments using diamond anvils can be extrapolated to the higher pressures. Considerable uncertainty surrounds all of this experimental work and its extrapolation to conditions at the boundary of the inner core. In addition, little is known of how the melting point of iron is affected by the addition of other constituents (see Boehler, 2000, for summary). With all these uncertainties, it is only possible to say that the temperature at the boundary of the inner core is around 5000 K (Fig. 1.1).

Temperature gradients through the liquid outer core must be controlled largely by convection, and in the limiting case, would follow what is known as an *adiabat*; that is, no heat would be lost or gained from the convecting liquid as it rose toward the core–mantle boundary, but its temperature would decrease as the liquid decompressed (see adiabat in Chapter 7.5 and Problem 7.7). The fact that the outer core is still liquid at the core–mantle boundary indicates that the temperature must be above the melting point of iron (plus other components) at this depth. High-pressure experiments indicate that at the 136 GPa pressure of this boundary, the temperature must be above ~4000 K (Boehler, 2000).

Seismic data indicate that the lower mantle is solid. Temperatures throughout the lower mantle must therefore be below the melting point, except perhaps at the base of D″ where the ultralow-velocity zone indicates that partial melting probably occurs near the core–mantle boundary. Experiments and theoretical calculations show that the extrapolated melting point of pure $MgSiO_3$ with a perovskite structure, one of the major components of the lower mantle, is 5400 ± 600 K at 136 GPa, the pressure at the core–mantle-boundary (Stixrude and Karki, 2005). The addition of other components would lower this melting point by as much as 1300 K. Thus, the temperature at the base of the mantle must be below (or near, if melt is present) 3500 to 5200 K, which is consistent with the ~4000 K estimated for molten iron in the outer core (Boehler, 2000).

Although the mantle is solid, calculations show that it must convect, and therefore it will have a nearly adiabatic temperature gradient of between 0.25 and 0.3 K km^{-1} (see Problem 7.7 and 7.8). This gradient can be pinned by experimental work to 1900 ± 100 K at a depth of 660 km, where the olivine composition switches from a spinel to a perovskite structure. An adiabatic gradient of 0.3 K km^{-1} would therefore give a temperature of ~2500 K at a depth of 2600 km. Below this depth, the D″ zone is a thermal boundary layer through which the temperature must rise rapidly by ~1500 K to reach the 4000 K of the outer core.

At shallower depths in the mantle, the geothermal gradient can be extrapolated from measurements near the surface of the Earth and from knowledge of the experimentally determined melting behavior of rocks that are known to exist at these depths. Measurements in deep mines and drill holes indicate that the near-surface geothermal gradient, dT/dz, ranges from 10 to 60 °C km^{-1}, with a typical value in nonorogenic regions being near 25 °C km^{-1} (Fig. 1.9). If this gradient continued to depth, the temperature would be 625 °C at 25 km (0.65 GPa). Experiments indicate that in the presence of water, crustal rocks melt to form granitic magma under these conditions. The transmission of seismic shear waves through this part of the crust indicates that melting at this depth is neither a common nor a widespread phenomenon. Similarly, at a depth of 40 km (1.2 GPa), a geothermal gradient of 25 °C km^{-1} would reach a temperature of 1000 °C, and peridotite, the rock constituting the mantle at this depth, would begin to melt in the presence of excess water. By 52 km (1.6 GPa) the temperature would be 1300 °C, which exceeds the beginning of melting of even dry peridotite. Again, seismic data do not indicate large-scale melting at this depth. Clearly, the near-surface geothermal gradient must decrease with depth. The question, then, is why should the gradient decrease, and can we calculate or predict its change?

A combination of petrological and geophysical observations allows us to place limits on the temperature of several points along the geotherm. Seismic velocities decrease abruptly by about 10% at a depth of 60 to 100 km, the upper boundary of the low-velocity zone. This is probably the contact between solid peridotite above and peridotite that has undergone a small percentage of melting below. A thin film of melt along grain boundaries can greatly reduce the rigidity of a rock and strongly attenuate seismic energy. Experiments show that water-saturated peridotite begins to melt at about 1000 °C at a pressure of 2 GPa, which corresponds to a depth of 70 km (Fig. 1.9). The top of the low-velocity zone is the boundary between the lithosphere and the asthenosphere. The base of the low-velocity zone, at 250 km, is marked by a small increase in the rate of increase of velocity with depth. This is interpreted as the depth at which the *P*–*T* curve for the beginning of melting of mantle peridotite recrosses the geotherm, so that at depths greater than 250 km, the mantle is solid. This places an upper limit of about 1500 °C at 250 km.

Common experience teaches us that heat flows from regions of high temperature to ones of low temperature, and that it is transferred in several ways depending on the nature of the medium through which it is transmitted. For example, in a vacuum heat can be transferred only by radiation; in a gas, liquid, or even plastic solid, such as the mantle, it may be transferred by convection; and in rigid, opaque solids, it can be transferred only by conduction. These mechanisms are discussed in detail in Chapter 5. Heat transfer through the lithosphere, however, especially in old, stable continental and oceanic regions, is almost entirely by conduction.

Experience also shows that the amount of heat transferred by conduction is proportional to the negative temperature gradient ($-dT/dx$); that is, the greater the temperature decrease in a particular direction x, the greater will be the amount of heat transferred in that direction. This can be expressed mathematically by introducing the quantity known as *heat flux*, J_{Qx}, which is the quantity of thermal energy passing in the x direction through a unit cross-sectional area in a unit of time. We can write this as

$$J_{Qx} = -K\left(\frac{dT}{dx}\right) \tag{1.3}$$

where K, the constant of proportionality, is known as the *coefficient of thermal conductivity*. In the case of the Earth, heat flows out because of the geothermal gradient. This gradient (dT/dz) is normally recorded as a positive value with z increasing downward. In the upward direction, however, this gradient is negative, so that a positive quantity of heat is transferred upward and out of the Earth.

Direct measurement of heat flux in the field is not practical, but temperature gradients can be measured easily in boreholes sunk into sediments on the ocean floor and in deep drill holes on continents. These gradients can then be used, along with laboratory measurements of the thermal conductivity of the penetrated material, to calculate the heat flux according to Eq. (1.3). For example, with a typical thermal conductivity coefficient of 2.0 J m^{-1} s^{-1} °C^{-1}, a gradient of 25 °C km^{-1} results in a heat flux of 50.0×10^{-3} J m^{-2} s^{-1}; but a joule per second is a watt, so the heat flux is 50.0 mW m^{-2}. For comparison, a 60 watt light bulb emits 60 joules per second from a surface area of approximately 11×10^{-3} m^2. Its energy flux is therefore 5.5 kW m^{-2}, that is, one-hundred-thousand times more than the heat flux from the Earth. The Earth's heat flux is therefore a very small quantity, and even where the geothermal gradient is 60 °C km^{-1}, which can be found near mid-ocean ridges, as for example in Iceland where it powers geothermal power plants (Fig. 1.4), the heat flux is still only 120 mW m^{-2}. Over very active spreading axes, such as the East Pacific Rise, the flux may be as high as ~300 mW m^{-2} (Fig. 1.5). In cgs units, heat flow is measured in μcal cm^{-2} s^{-1}, which is commonly referred to as 1 *heat flow unit* or 1 HFU; 1 HFU = 41.84 mW m^{-2}.

Fig. 1.4 Geothermal power plant near the Hengill volcano in southern Iceland. The valley in which the plant is situated is part of the Mid-Atlantic Ridge system passing through Iceland. This plant, which supplies Iceland's capital, Reykjavik, with much of its heating and electrical power, taps a high-temperature geothermal field where temperatures exceed 200 °C at a depth of 1 km. Pipes from many geothermal wells can be seen leading to the power plant. Geothermal power plants supply Iceland with ~17% of its electrical power, and ~90% of its central heating.

The International Heat Flow Commission's synthesis of worldwide heat flow data is available on the Web (www.geo.lsa.umich.edu/IHFC) and is reproduced here as Figure 1.5 (Pollack *et al*., 1993). This synthesis reveals that the new ocean floor created at divergent plate boundaries plays an important role in the loss of heat from the planet. Indeed, it is estimated that 50% of all heat lost from the Earth is liberated from these regions, which constitute only 30% of the surface of the planet. The heat flux from ancient ocean floor has a relatively constant value of 38 mW m^{-2}. As the ocean floor gets younger toward mid-ocean ridges, the heat flux steadily increases, reaching a mean value of 250 mW m^{-2} in ocean floor younger than 4 Ma, but there is considerable scatter in the younger data. The scatter has two main causes. Old ocean floor has had time to accumulate thick layers of sediment, which seals the underlying fractured igneous rocks from circulating ocean water and provides a layer into which drilling probes can be sunk easily to obtain a conductive geothermal gradient. On new ocean floor, the lack of sediment makes it difficult to drill holes in which to measure the

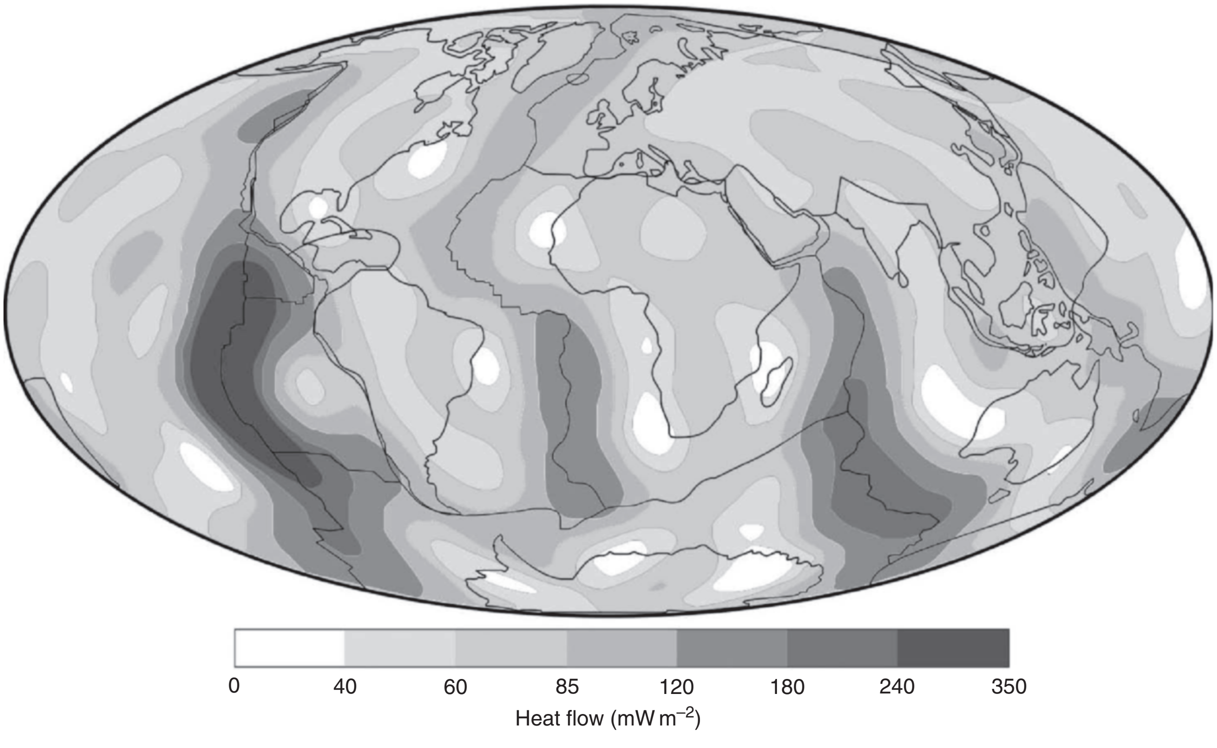


Fig. 1.5 Global heat flow. Data are shown as a degree 12 spherical harmonic representation (after Pollack *et al*., 1993).

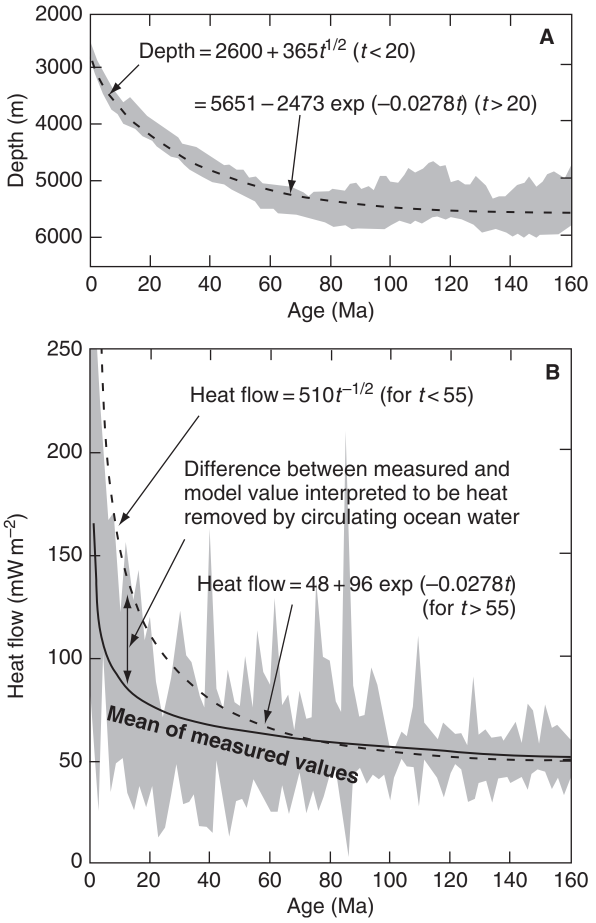


Fig. 1.6 (**A**) Shaded area shows variation in mean ocean depth (± 1 standard deviation) of the North Pacific and Northwest Atlantic as a function of ocean floor age. The dashed line shows mathematical fits to the data (see text for discussion). (**B**) Shaded area shows variation in mean heat flow (± 1 standard deviation) from the same ocean floor as in (A). The solid line is the mean of the measured data, whereas the dashed line is a model based on the cooling of the lithosphere as it moves away from an ocean spreading axis. The difference between the model and measured values is interpreted to be heat removed by circulating ocean water. (Drawn from figures presented in Stein and Stein, 1992, and Hofmeister and Criss, 2005.)

temperature gradient. In addition, without sediment sealing fractures, ocean water is free to circulate through the igneous rocks and remove heat that would otherwise have to be conductively transferred through the rock. Vents have been found along many oceanic ridges emitting high temperature water from which sulfides commonly precipitate to form what are called "black smokers" (Von Damm, 1990). Clearly, a considerable amount of heat is removed from the Earth by these circulating waters, which is in addition to the heat that is transferred conductively through the rock. To understand the Earth's heat budget, which essentially controls the rate of many processes on the dynamic planet, we need to know the magnitude of this hydrothermal component. To see how scientists have tried to answer this question, we must first examine one of the most remarkable features of the planet, the relation between the age and depth of the ocean floor.

One of the most important findings to come from the exploration of the ocean floor is the discovery that the ocean's depth is proportional to the age of the ocean floor; that is, the older the ocean floor, the greater is its depth (Fig. 1.6(A)). Indeed, for ocean floor younger than 20 Ma, its depth is proportional to the square root of its age, with depth being expressed accurately by:

$$\text{Depth} = 2600 + 365 \times t^{1/2} \tag{1.4}$$

where depth is in meters and time in Ma BP (million years before present; Stein and Stein, 1992; Parsons and Sclater, 1977). Even for ocean floor as old as 70 Ma, the relation still holds quite well, but older ocean floor sinks more slowly. The square root of time relation has been interpreted in terms of the cooling of the lithosphere as it moves away from the ocean ridges; the crust becomes denser as it cools, and isostatic readjustment causes the ocean floor to sink deeper with increasing age. Numerous attempts have been made to model the cooling and sinking of lithospheric plates (see Oxburgh, 1980). We will discuss these models in more detail in Chapter 23, but for the moment it is necessary to recognize only that when an object conductively cools by transferring heat across a plane boundary, the heat flux across that boundary typically decreases in proportion to the square root of time (see Eq. (5.13)). Although we may not have been aware of the square root term, we are all familiar with this effect from the cooling of a cup of coffee. Initially, when the coffee is hot, heat is lost rapidly and the coffee cools rapidly, but with time, as the coffee cools, it loses less and less heat and cools more slowly.

If cooling is the correct explanation for the increased depth of the ocean floor with age, then the heat flux through the ocean floor would be expected to be inversely proportional to the square root of the age of the ocean floor (Fig. 1.6(B)). This, in fact, is found to be the case where the ocean floor is covered with a significant thickness of sediment. Stein and Stein (1992) have shown that in these regions the heat flux is well described by

$$J_Q = 510 \times t^{-1/2} \tag{1.5}$$

where J_Q is in mW m^{-2} and t is in Ma BP, as long as t is less than 55 Ma. However, where the ocean floor is young and has little sediment covering it, the heat flux measurements fall farther and farther below the predicted curve as the ocean ridge is approached (Fig. 1.6(B)). This shortfall is interpreted to represent heat that is removed by the circulating ocean water. Stein and Stein (1994) estimate that 34% of the total global oceanic heat flux occurs by hydrothermal circulation (see also papers in Davis and Elderfield, 2004). In the synthesis of worldwide heat-flow data shown in Figure 1.5, the heat-flux values for young ocean floor are not measured values but are extrapolations from measurements on older ocean floor using Eq. (1.5).

Despite the common acceptance of calculating young ocean-floor heat fluxes in this way, other models have been proposed (Hamilton, 2003). Hofmeister and Criss (2005) argue that the heat flux in these regions is actually much

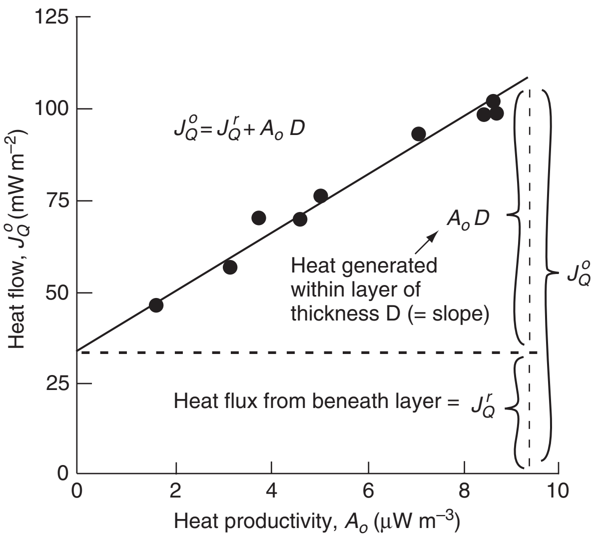


Fig. 1.7 Heat flow plotted as a function of the radiogenic heat production of surface rocks in New England (after Birch *et al.*, 1968). See text for discussion.

closer to the measured values than it is to the extrapolated values and that far less heat is being removed by hydrothermal circulation than predicted by the square root of time cooling model (Fig. 1.6(B)). Depending on which model is used for the heat flux in young ocean floor, huge differences result in the calculated total heat budget for the Earth. For example, according to Pollack *et al*. (1993), the total heat flux from the Earth is 44 TW, whereas Hofmeister and Criss (2005) estimate it to be only 31 TW. The lower estimate agrees well with the amount of heat that would be expected to be generated from radioactive decay if the planet were formed from chondritic meteorites (see Chapter 6). The higher heat flux estimates require additional heat sources, which could be left over from an earlier age, such as planetary accretion and large impacts, short-lived radioactive isotopes, core formation, and crystallization of the outer core.

Heat-flux measurements on continents do not share the problems of those on new ocean floor. As first shown by Polyak and Smirnov (1968), and substantiated by other studies (Sclater *et al*., 1980), heat flux on continents decreases with increasing age, at least back to about 800 Ma. The rate of decrease, however, is much less than that found in the ocean floor and therefore cannot be modeled as simple cooling of newly formed crust. The relation is more complex, possibly involving the removal, by erosion, of heat-generating radioactive elements that are concentrated in the upper parts of young crust. In young orogenic belts, heat fluxes may be as high as 150 mW m^{-2}, but in crust older than 800 Ma, they tend to be about 40 mW m^{-2}, which is similar to the value in ancient oceanic crust.

One of the most important findings to come from the study of heat flow is the relation between the surface heat flux and the concentration of heat-generating radioactive elements in the local rocks (Birch *et al*., 1968; Lachenbruch, 1968; Lachenbruch and Sass, 1977). The heat flux from old eroded plutonic bodies of igneous rock is very nearly linearly related to the local concentration of heat-generating elements, as shown in Figure 1.7. This linear relation can be expressed mathematically as

$$J_Q^0 = J_Q^r + DA_0 \tag{1.6}$$

where J_Q^0 is the surface heat flow (depth = 0), J_Q^r the intercept at $A_0 = 0$, A_0 the local radiogenic heat productivity, and D the slope of the line.

The simplest interpretation of this relation is that the slope of the line indicates the thickness of the heat-generating layer. Thus the heat flux at any locality can be interpreted as consisting of two parts; a constant flux of heat, which is given by the intercept J_Q^r, comes from below the layer; to this is added the heat flux generated by radioactive decay in the layer. This type of relation is known as an *energy balance* or *conservation equation*, for it states that the energy coming out of the top of the layer must equal the energy entering the layer from below plus the energy created within the layer, that is,

$$J_Q\Big|_{z=0}^{\text{out}} = J_Q\Big|_{z=D}^{\text{in}} + \int_{z=0}^{z=D} A_z \, \mathrm{d}z \tag{1.7}$$

Values of J_Q^r, which are known as the *reduced heat flow*, are characteristic of a given geological province. For example, its value in the eastern United States is ~33 mW m^{-2}, in the Basin and Range Province it is ~59 mW m^{-2}, and in both the Baltic and Canadian Precambrian Shields it is ~22 mW m^{-2}. Values of D, the thickness of the layer containing the radioactive elements, in these same provinces are 7.5, 9.4, 8.5 and 12.4 km respectively (Oxburgh, 1980). The fact that the surface concentration of radiogenic elements cannot extend to depths much greater than 10 km is surprising considering that seismic data indicate that "granitic" rocks at the surface extend to depths of approximately 25 km under most continental areas. Clearly, the radioactive elements must be concentrated near the surface of the Earth. This being so, our model of a layer with a constant concentration, A_0, of radioactive elements throughout its thickness D is unlikely to be valid. The concentration is more likely to decrease downward (see Problem 1.6).

1.5 HEAT SOURCES IN THE EARTH

Heat flows into the base of the crust from the mantle below in response to temperature gradients that were set up early in the Earth's history. This heat has two sources, the release of gravitational potential energy and the decay of naturally occurring radioactive nuclides. Conversion of kinetic to thermal energy during the Earth's accretion, which was largely complete by 4.567 Ga (Jacobsen, 2003), release of gravitational energy on separation of the core very early in Earth's history (Solomon, 1979), energy from a Moon-forming mega impact after the core had formed but before 4.45 Ga (Zhang, 2002), and generation of heat by decay of short-lived radioactive isotopes, such as ^{26}Al, have all contributed to the

Table 1.1 *Typical concentration of radiogenic elements in common rock types and their heat production*

	Concentrations (wt ppm)					
	U	Th	K	Density (kg m^{-3})	Radiogenic heat production A (μW m^{-3})	Thermal conductivity K (W m^{-1} K^{-1})
Granite	4.7	20.0	36 000	2650	2.95	2.93
Basalt	0.9	2.2	15 000	2800	0.56	2.09
Average crust	1.55	5.75	15 000	2800	1.00	2.51
Peridotite	0.019	0.05	59	3150	0.01	3.35

Decay energies in mW kg^{-1}: U = 9.66×10^{-2}, Th = 2.65×10^{-2}, K = 3.58×10^{-6}.

thermal energy of the early Earth (Hamilton, 2003). This early heat would have been sufficient to melt the planet and create a magma ocean. Convection in this ocean would have been vigorous, so that much of the early heat would have been released early in the Earth's history. Since then, most of the heat flux from the Earth has come from the decay of long-lived radioactive isotopes. These isotopes tend to be concentrated in the crust, especially the continental crust. If the Earth were formed from the accretion of chondritic meteorites, which is the favored hypothesis, the Earth appears to be deficient in potassium. Because potassium is a moderately volatile element, it has been suggested that it may have been lost from the planet during the magma ocean stage. However, recent experiments indicate that potassium is soluble in molten iron at high pressure, in which case it may have been incorporated into the Earth's core (Lee and Jeanloz, 2003). The potassium isotope ^{40}K is a long-lived heat-producing radioactive isotope whose presence in the core could provide the heat necessary to maintain the Earth's geodynamo that creates the magnetic field and drives mantle convection.

Heat is generated in the Earth by the decay of the long-lived radioactive isotopes ^{235}U, ^{238}U, ^{232}Th, and ^{40}K. As discussed above, these elements tend to be concentrated in the crust, although ^{40}K may also occur in the core. Because these elements are very minor constituents of the oceanic crust they contribute no more than 10% to 25% of the heat flux in the ocean floor, but in continental crust, they are much more abundant and are a major source of thermal energy.

The kinetic energy of particles emitted from the nucleus of radioactive nuclides is dissipated and transformed into heat by collisions with surrounding atoms. Alpha (α) particles ($^4_2$He) emitted by U and Th isotopes have a mass of 4 units, whereas beta (β) particles (electrons emitted from the nucleus) emitted by ^{40}K during radioactive decay to ^{40}Ca have a mass of only 1/1833 units. ^{40}K also decays to ^{40}Ar by the process of electron capture by the nucleus (see Section 13.2), but this does not contribute to heat generation. Because of the difference in mass, a disintegration involving an alpha particle gives off about 7300 times more heat than one involving a beta particle. The concentration of K in crustal rocks, however, is about 10 000 times that of U or Th, so that the contribution to the total radiogenic heat productivity of each of these elements is about the same (see Table 1.1 and Problem 1.3).

The rate of radiogenic heat production per unit volume of rock, A, is the sum of the products of the decay energy of each of the isotopes present, e_i, and the isotope's concentration in the rock, c_i. This can be written as

$$A = \rho \sum_i e_i c_i \quad \frac{\text{kg}}{\text{m}^3}\,\frac{\text{J}}{\text{kg s}}\,\frac{\text{kg}}{10^6\text{kg}} \quad \text{or} \quad \frac{\mu\text{W}}{\text{m}^3} \tag{1.8}$$

Density appears as a factor in Eq. (1.8) to convert units of power per unit mass to power per unit volume. The concentrations are given in parts per million by weight (ppm = kg/10^6 kg). Typical values of radiogenic heat production for the major rock types of the lithosphere are given in Table 1.1 along with data needed for their computation with Eq. (1.8). Values of A lie in the range 0.008 to 8.0 μW m^{-3}. In the cgs system of units, the quantity 1×10^{-13} cal cm^{-3} s^{-1} is known as one *heat generation unit* or 1 HGU; 1 HGU = 0.4184 μW m^{-3}.

1.6 TEMPERATURES IN THE LITHOSPHERE: THE STEADY-STATE GEOTHERM

We have seen in Section 1.4 that the downward extrapolation of near-surface geothermal gradients leads to excessively high temperatures at depth. The gradient must therefore decrease with depth. Several factors contribute to the flattening of the gradient. The concentration of heat-producing radioactive elements in near-surface rocks produces a heat source that diminishes with depth and thus causes flattening of the geothermal gradient. Convection of the solid (but plastic) mantle beneath the lithosphere maintains the temperature gradient near a relatively flat adiabat of about 0.6 °C km^{-1}.

Because simple linear extrapolation of the near-surface geothermal gradient leads to erroneous temperatures at depth, it is desirable to develop a more reliable extrapolation that accounts for the change in gradient with depth. Although this is difficult to do for lithosphere near tectonic plate boundaries, it can be done easily for old continental and oceanic areas, where the geothermal gradient has attained a steady state; that is, it does not vary significantly with time.

If the geotherm is in a steady state, temperature can vary only as a function of depth; that is,

$$T = T(z) \tag{1.9}$$

A useful means of approximating a function beyond the region in which its value is known is with a Taylor series expansion (Turcotte and Schubert, 1982b). This expansion can be used if *T(z)* is a function with continuous derivatives of all orders. Physically, this means that the function must vary gradually – there can be no abrupt changes in its value. Experience tells us that this is true of the geotherm. If an abrupt step in the temperature gradient could be produced, heat would transfer from the region of high temperature to that of low temperature until a smooth gradient was produced. Because we are considering only steady-state geotherms, *T(z)* will be a function with continuous derivatives of all orders. Given temperature $T_{z=0}$ at the Earth's surface, we expand T downward with a Taylor series in z from $z = z_0$ to $z = z$. Temperature at depth, z, is given by

$$T_z = T_{z_0} + \frac{dT}{dz}(z - z_0) + \frac{1}{2!}\frac{d^2T}{dz^2}(z - z_0)^2 + \ldots + \frac{1}{n!}\frac{d^nT}{dz^n}(z - z_0)^n \tag{1.10}$$

In any complex equation it is worth having a qualitative idea of the meanings of each group of terms. In Eq. (1.10), the first term is the temperature at the Earth's surface. The second group of terms includes the thermal gradient, which is the change of temperature with depth (first derivative). The third group of terms includes the change of gradient with depth (second derivative). Note that the exclamation mark indicates factorial; that is, $n! = 1 \times 2 \times 3 \times \cdots \times n$. The series includes higher order derivatives, but a good approximation can be obtained with just the first three groups of terms.

To solve Eq. (1.10), we must evaluate the derivatives. From Eq. (1.3) we have

$$\frac{dT}{d(-z)} = -\frac{J_Q^z}{K} \quad \text{or} \quad \frac{dT}{dz} = \frac{J_Q^z}{K} \tag{1.11}$$

The ($-z$) is required in the first term to indicate that the gradient used in Eq. (1.3) is in the direction of heat flow; that is, upward ($-z$). The geothermal gradient, dT/dz, is then a positive quantity. We can then write the second derivative as

$$\frac{d^2T}{dz^2} = \frac{1}{K}\frac{dJ_Q^z}{dz} \tag{1.12}$$

If we assume that the concentration of heat-producing elements is constant with depth over the interval $z = 0$ to $z = D$, the conservation equation (1.7) becomes

$$J_Q^z = J_Q^D + A_0(D - z) \tag{1.13}$$

which, on differentiating with respect to z, gives

$$\frac{dJ_Q^z}{dz} = -A_0 \tag{1.14}$$

Substituting this value for the derivative into Eq. (1.12) gives

$$\frac{d^2T}{dz^2} = -\frac{A_0}{K} \tag{1.15}$$

Table 1.2 *Parameters used in calculating the geothermal gradients in Figure 1.9*

Depth (km)	Ocean geotherms		Continental geotherms	
	O_1	O_2	C_1	C_2
0_	$J_Q^0 = 34$	$J_Q^0 = 42$	$J_Q^0 = 46$	$J_Q^0 = 46$
10_	$A = 0.84$ $K = 2.51$	$A = 0.42$ $K = 2.51$	$A = 2.09$ $K = 2.51$	
20_	$J_Q^r = 25.1$ $A = 0$ $K = 3.35$	$J_Q^r = 37.7$ $A = 0$ $K = 3.35$	$A = 0.261$ $K = 2.51$	$A = 0.31$ $K = 2.51$
30_				
40_				
50_			$J_Q^r = 17$ $A = 0$ $K = 3.35$	$J_Q^r = 34$ $A = 0$ $K = 3.35$

Units are as follows: J_Q (mW m^{-2}); A (μW m^{-3}); K (W m^{-1} K^{-1}). After Sclater *et al.* (1980).

and the Taylor series approximation to temperature becomes

$$T_z = T_{z_0} + \frac{J_Q^{z_0}}{K}(z - z_0) - \frac{A_0}{2K}(z - z_0)^2 \tag{1.16}$$

Note that if the temperature at depth z has units of °C, the units of each group of terms on the right-hand side of Eq. (1.6) must also be °C. Noting this can help avoid errors in doing calculations.

The inclusion of successive terms in this series provides ever-closer approximations to the true value of T (Fig. 1.8). The first term underestimates T, because it simply equates the temperature at any depth z with the temperature on the surface of the Earth (T_{z_0}). The first two terms overestimate T, because they linearly extrapolate the near-surface gradient to depth. The first three terms take into account possible distributions of radioactive heat-generating elements, and the equation converges closely with the true value of *T(z)*. Incorporation of higher terms in the series would improve the fit still more, but these cannot be evaluated without additional information on the variation of A with z.

Heat flow through oceanic lithosphere approaches a steady state after 180 Ma, with a surface heat flow of 38 ± 4 mW m^{-2} and a reduced heat flow into the base of the crust from the mantle of 25 to 38 mW m^{-2} (Sclater *et al.*, 1980). These limiting values can be used along with other data in Table 1.2 to calculate steady state geotherms with Eq. (1.16). The results are shown in Figure 1.9. Note that a relatively small difference in near-surface geothermal gradient (13 °C km^{-1} for ocean geotherm 1, O_1, and 17 °C km^{-1} for O_2) leads to significantly different temperatures at depth.

Heat flow through stable continental shields approaches a steady state after about 800 Ma, with a surface heat flow of 46 ± 17 mW m^{-2}, and a best estimate of the reduced heat flux from the mantle being 21 to 34 mW m^{-2} (Sclater *et al.*, 1980). Geotherms computed with Eq. (1.16) and data from Table 1.2 are shown in Figure 1.9. The models for continental

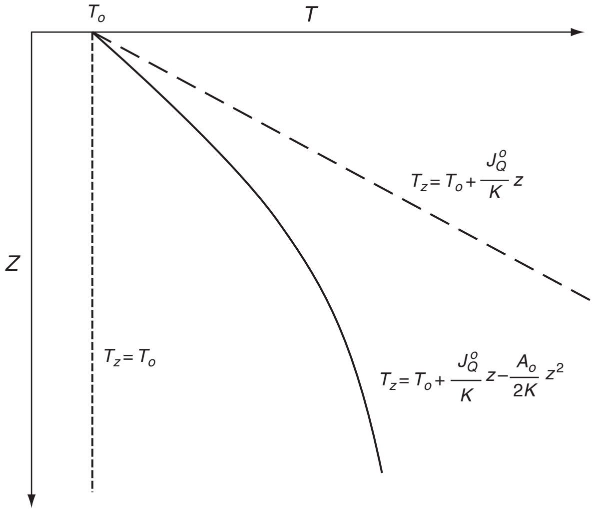


Fig. 1.8 Plots showing how inclusion of additional terms in the Taylor series expansion of temperature as a function of depth gives closer approximations to the actual geothermal gradient. The short-dashed line includes only the first term of the series; the long-dashed line includes the first and second terms; and the solid line includes the first three terms.

lithosphere differ in the relative contribution of heat generated in the crust and heat transported from the mantle. Concentration of heat-producing elements in the crust leads to a significant contribution to surface heat flow, thus diminishing the flux from the mantle needed to produce the observed flux at the surface. The net result is lower temperatures at depth than would be produced for a given surface heat flow through a crust depleted in U, Th, and K (see Problem 1.5).

Two important facts are evident from the calculated geotherms. First, despite differences in surface heat flow, crustal heat productivity, and heat flow into the base of the crust, the range of temperatures at depth beneath old ocean and old continents overlap. Second, temperatures in the oceanic and in the continental crust are too low to produce magmas or metamorphic rocks. The eroded cores of orogenic belts contain rocks that have been formed at temperatures well above the calculated geotherms. For instance, large granite batholiths are formed at temperatures between 700 and 900 °C, and metamorphic rocks in the core of these belts commonly exhibit evidence of partial melting, which also requires temperatures in excess of 700 °C. Even where metamorphic rocks have not been melted, mineral assemblages provide evidence of high temperatures.

The Al_2SiO_5 polymorphs, andalusite, kyanite, and sillimanite (A, K, and S in Fig. 1.9), which are common minerals in rocks formed from the metamorphism of shales, are particularly useful indicators of metamorphic temperature and pressure. Kyanite is the stable phase at high pressures and low to moderate temperatures; andalusite is stable at low pressures and moderate temperatures; and sillimanite is the high-temperature phase. All three polymorphs can coexist at a "*triple point*" at about 500 °C and 0.375 GPa (Fig. 1.9). In metamorphic terranes, these polymorphs are commonly found to change with increasing metamorphic intensity from andalusite to kyanite, or from andalusite to sillimanite, or from kyanite to sillimanite, and rocks with all three polymorphs have been found. Examination of Figure 1.9 reveals that the calculated continental geotherms lie entirely in the stability field of kyanite. Indeed, geotherms could pass through the andalusite or sillimanite fields only if the heat flux into the crust from the mantle were substantially increased. We will see that this condition is met above subduction zones (Fig. 1.2) and results from the rise of mantle-derived magmas into the crust. Elevated temperatures also exist in regions of lithospheric extension due to the rise of hot mantle from the asthenosphere and where continental crust has been thickened in collisional orogenic belts. It is these transient effects that are responsible for the formation of rocks. They are consequently of paramount importance to petrology but cannot be dealt with in our simple steady state model for geothermal gradients. We will, however, return to their consideration in Chapter 23.

It was suggested above that the top of the low-velocity zone marks the intersection of the geotherm with the pressure–temperature curve for the beginning of melting of mantle peridotite. Such an interpretation is consistent with the calculated geotherms (Fig. 1.9), which intersect the H_2O-saturated beginning of melting curve (solidus) for peridotite between 75 km (C_2) and 150 km (O_1).

Because the geotherms intersect the solidus at a small angle, slight differences in heat flux correspond to large differences in lithospheric thickness. For example, the difference in surface heat flow of 8.4 mW m^{-2} between the two calculated oceanic geotherms leads to a 70-km difference in lithospheric thickness. In general, high surface heat flow causes the geotherm to intersect the peridotite solidus at a shallower depth, whereas low surface heat flow causes it to intersect at greater depth. If worldwide values of surface heat flow are converted into depths of intersection with the peridotite solidus, the lithosphere is found to be up to 300 km thick beneath Precambrian shields but thins to about 50 km beneath oceanic ridges (Pollack and Chapman, 1977). The continents, therefore, resemble ships with very large lithospheric keels descending into the asthenosphere.

1.7 GENERAL OUTLINE OF THE ORIGIN OF IGNEOUS AND METAMORPHIC ROCKS

In the following chapters, we will discuss a wide range of petrologic topics that will be synthesized in the final chapter (Chapter 23) into a general discussion of the origin of igneous and metamorphic rocks. To help show how the various topics relate to the broader topic of the origin of rocks, we will close this introductory chapter with an overview of the main points that will be found discussed in detail in Chapter 23.

From our preliminary discussion of the pressures and temperatures in the Earth, it is clear that the conditions

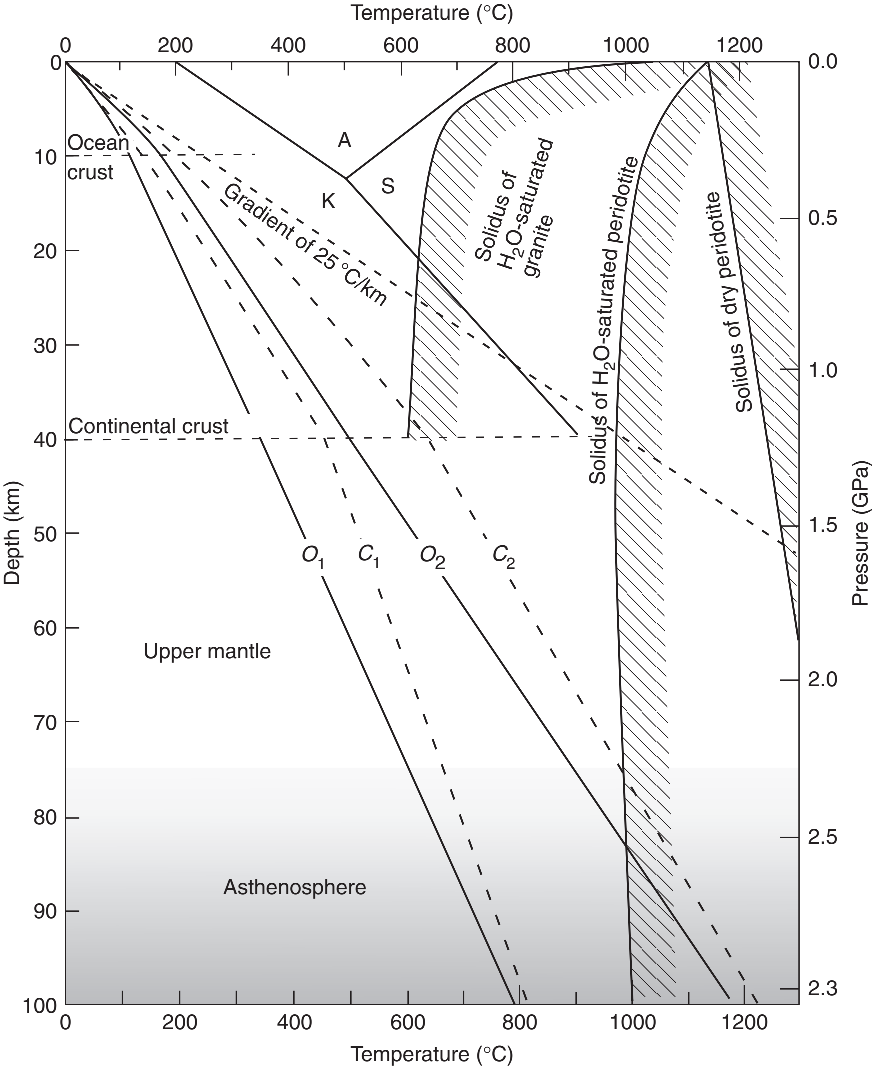


Fig.1.9 Possible temperature distributions in a stable lithosphere, far removed from spreading or convergent plate boundaries, calculated using Eq. (1.16) and data in Table 1.2 (after Sclater *et al.*, 1980). Two oceanic (O_1 and O_2) and two continental (C_1 and C_2) geotherms are shown to cover the range of possible values. Also included are stability fields of the Al_2SiO_5 polymorphs: kyanite, K; andalusite, A; and sillimanite, S (after Holdaway, 1971); the beginning of melting curve (solidus) for water-saturated granite (after Huang and Wyllie, 1973); and the solidus of peridotite, both dry and water saturated (after Kushiro *et al.*, 1968). The geotherms are seen to intersect the solidus of water-saturated peridotite below 75 km.

necessary for the formation of rocks are not those that normally exist in what we refer to as a steady state Earth. The formation of rocks is a relatively rare event that requires major perturbations of the geothermal gradient. The reader can test this by asking themselves when igneous and metamorphic rocks were last formed in the area where they live. For example, in southern New England where the authors reside, rocks were last formed ~200 Ma BP, and prior to that, ~400 Ma BP. Although the steady-state geotherm is continuously transferring heat to the Earth's surface, only during periods of rock formation does major cooling of the Earth occur. Rock formation is the major way in which Earth rids itself of heat. People perspire to get rid of excess heat – the Earth forms rocks.

The perturbations of the geothermal gradient needed for rock formation are caused by tectonic plate motion – large meteorite impacts would be a rare exception to this general rule. Most igneous rocks are formed near plate margins, but some are formed over hot spots, which can be within plates. Most metamorphic rocks are formed on a regional scale on either side of convergent plate boundaries or beneath zones of extension, but they can be developed locally at the contact of igneous bodies wherever these occur.

The type of plate margin determines the nature of the geothermal perturbation. For example, near divergent plate boundaries, the rise of the asthenosphere into the thinned lithosphere steepens the geotherm, which causes melting and high-temperature metamorphism. At convergent boundaries, subduction of cool lithosphere into the mantle makes the geotherm in the subducting slab shallower, which produces low-temperature, high-pressure metamorphic rocks. Water released during these metamorphic reactions rises into the mantle wedge above the subducting slab and causes melting; as magma rises, it steepens the geotherm and, along with deformation accompanying plate convergence, causes metamorphism on a regional scale. At transform boundaries, different geotherms can be juxtaposed, and mechanical deformation can add heat to the geotherm. Finally, over hot spots, the geotherm is obviously steepened, with resulting magmatism and thermal metamorphism.

Plate tectonics is the surface expression of the mantle convection that transfers heat from the Earth's deep interior to the near surface where it can be conducted and radiated into space. Although it has yet to be proved conclusively, seismic evidence strongly points to whole-mantle convection rather than separate upper- and lower-mantle convection cells. This convection, while providing the perturbations in the lithosphere necessary for rock formation, also provides access to a wide range of chemically distinct reservoirs within the Earth. These range from the D″ layer on the core–mantle boundary, which may contain material dating back to the accretion of the planet, to sediments on the surface, whose chemical composition may reflect a long and complex history involving melt extraction from the mantle and magmatic differentiation, followed by chemical weathering, transport, and deposition.

The rise of magma is the most rapid way in which heat is transferred out of the Earth to reestablish steady-state conductive geotherms. But if hot magma rises toward the surface, there must be a corresponding return flow of cool material into the Earth. The bulk of this return flow is in the form of subducting lithospheric plates. New igneous rocks formed at oceanic spreading axes are cooled and altered by circulating ocean water. By the time these igneous rocks are subducted at convergent plate boundaries, they have been changed into low-temperature metamorphic rocks characterized by an abundance of hydrous and carbonate minerals (~2.7% H_2O, ~2.9% CO_2; Staudigel *et al*., 1989). As the plate subducts, these rocks, along with their covering of ocean-floor sediments, lose volatiles as they are metamorphosed into progressively denser mineral assemblages, which eventually form a rock consisting largely of pyroxene and garnet (eclogite). The high density of this rock probably provides much of the force that drives convection and may cause the material to sink as deep as the D″ layer on the core–mantle boundary.

The release of metamorphic volatiles during subduction is one of the most important processes to have helped shape planet Earth (Kerrick and Connolly, 2001b). Water released from subducting plates provides the flux that causes magmatism above subduction zones (Ulmer, 2001). The resulting igneous rocks have steadily created the continental crust throughout Earth history. Much of the fluxing water is eventually released back into the atmosphere during volcanic eruptions, which often occur with explosive violence. These eruptions constitute one of the major hazards faced by people living in active volcanic regions. Devolatilization of the subducting slab has also been proposed as the triggering mechanism for deep earthquakes in these same regions. The release of volatiles raises pore pressures that can decrease frictional resistance and lead to shear failure, especially in rock that is embrittled by the change from hydrous to anhydrous mineral assemblages. Although CO_2 is released by volcanism along ocean spreading axes, the formation of carbonate alteration minerals in ocean-floor basalt and the deposition of carbonate-bearing sediments on the ocean floor results in a net loss of $1.5\ \ 2.4 \times 10^{12}$ mol of carbon per year from the ocean atmosphere system (Alt and Teagle, 1999). The carbon is trapped in this sink until the oceanic lithosphere is subducted at convergent plate boundaries, where it may be released in hydrothermal solutions or magmatic gases, but it may also be subducted back into the deep mantle if the geotherm is shallow. The rate and nature of the loss of volatiles from the subducting slab can therefore affect the atmosphere's CO_2 content, which being a greenhouse gas can affect global climates.

One of the main ways in which a perturbed geotherm returns to the steady state is by producing magma, which is normally less dense than its surroundings and so buoyantly rises through the crust where it may solidify and release heat to cause metamorphism or continue to the surface to erupt and form volcanic rocks.

It is natural to think that magma is the result of heating, but most melting in the Earth occurs with little or no heat addition. Instead, it is caused by decompression or addition of a flux that lowers melting points. The melting point of relatively dry rock is lowered by decreasing pressure. Thus, as the asthenosphere rises beneath divergent plate boundaries, it partially melts to form basaltic magma. The melting point of rock is also lowered by the addition of water. Consequently, when hydrothermally altered ocean floor is subducted into the mantle, the increased pressure causes metamorphic reactions to take place that release the water, which rises into the overlying mantle wedge and causes melting to form andesitic magma. Even beneath hot spots, the main cause of melting to form basaltic magma is the decompression that accompanies the rise of a mantle plume. Heating does play a role where mantle-derived basaltic magma intrudes at the base of the continental crust (underplates). As this magma cools, it releases heat that causes partial melting in the overlying crust to form bodies of granitic magma. These processes are summarized in Table 1.3.

Mixtures of minerals melt at lower temperatures than do individual minerals, so melting always takes place at multiple-mineral-grain junctions. The liquids formed between these minerals have rather restricted compositions (eutectic, peritectic), which correspond to the compositions of the main types of igneous rock (Table 1.3). Once the melt is formed at the multiple-mineral-grain junctions, its distribution in the rock depends on whether or not the liquid wets the minerals with which it is in contact, which is determined by surface tension. Experiments indicate that silicate liquids do wet silicate minerals, and as a result, as soon as liquid is formed, it is dispersed throughout the rock along all grain edges. Partially melted rock can be thought of as behaving much the same as a paper towel does when it mops up liquid. This means that the melt can flow through partially melted rock by porous flow, but it also means that the rock will retain a certain percentage of liquid to satisfy surface tension requirements. By analogy, no matter how much you squeeze a wet paper towel, the towel remains damp. Removal of liquid from rock undergoing partial melting will be impossible until the fraction of liquid exceeds that required to wet all of the grain boundaries (3 vol%).

Table 1.3 *Important properties of common igneous rocks and their mode of origin*

Rock type								
Volcanic	Plutonic	Minerals	Composition (% SiO_2)	Viscosity (Pa s)	Source	Cause of melting	Plate tectonic location	Volcanic feature
Basalt	Gabbro	Ol, Pyx, Plag	50	10^2	Mantle	$-\Delta P$, $+\Delta T$	MOR, rift valley, hot spot	Flood basalt, shield volcano
Andesite	Diorite	Ol, Pyx, Plag	60	10^3	Mantle	$+H_2O$	Above subduction zone	Strato/composite volcano
Rhyolite	Granite	Alkali feld, Q	70	$10^{6–9}$	Crust	$+H_2O$, $+\Delta T$	Above subduction zone, continental rift and hot spot	Dome, ash flow, ash fall

Abbreviations used: Ol, olivine; Pyx, pyroxene; Plag, plagioclase; feld, feldspar; Q, quartz; $-\Delta P$, decompression; $+\Delta T$, heating; $+H_2O$, addition of water; MOR, mid-ocean ridge.

The initial separation of liquid from partially melted rock occurs because of density contrasts between the liquid and solid. In the upper mantle and crust, liquids are normally less dense than the solids from which they form, and thus they tend to rise. In the lower mantle, however, liquids may be denser than the solids and would thus sink. If the solids are denser than the liquid, they will tend to compact, and the liquid will be expelled upward. For melt to rise at significant rates it must segregate from the solids. Flow is initially slow as the liquid moves through pores, but it increases significantly with the formation of channels, which are promoted by deformation of the crystal mush. Gradually, regions of crystal mush develop that contain such high proportions of liquid that the bulk properties change from that of a crystal mush to that of a liquid suspension. At this point a body of magma is formed.

Basaltic and andesitic magmas are formed in the upper mantle, whereas most granitic magma is formed near the base of the continental crust. In both regions, magmas are formed in response to perturbations of the geotherm. Once melting starts, the amount of heat required to change solid into liquid (latent heat of fusion) is so great that only partial melting occurs; rarely is there enough heat to cause total melting. Consequently, although the liquid fraction may tend to segregate in zones of partial melting, magmas typically contain crystals, and most igneous rocks have a porphyritic texture, that is, larger crystals in a finer grained matrix. This important feature was correctly interpreted by Bowen (1928) to indicate that the Earth does not have superheat. Again, using the human analogy, our bodies do not overheat because excess heat is removed by perspiration that evaporates from the surface of the body (latent heat of vaporization). Likewise, the Earth never overheats, because excess heat causes melting (latent heat of fusion) and formation of bodies of magma that rise and transfer heat rapidly to the surface.

Most bodies of magma contain large percentages of crystals. Even beneath extremely active oceanic spreading axes, such as the East Pacific Rise where magma must be present, geophysical evidence has failed to identify large bodies of liquid. Instead, the evidence points to bodies of crystal mush, toward the top of which may occur thin lenses of liquid (Sinton and Detrick, 1992). At depth in the Earth, magma is likely to contain still greater percentages of crystals than it does at shallower depths, because melting results from the decompression of the magma as it rises.

Bodies of magma in the mantle and lower crust most likely rise as large inverted teardrop-shaped bodies (diapirs) that must force aside the surrounding rock. On reaching the brittle–ductile transition (~15 km), fracturing of the lithosphere becomes important in providing magma conduits. Although the buoyant force on magma can cause fracturing, tectonic stresses play important roles in determining the orientation and amount of opening of fractures.

The velocity with which magma rises through these fractures is strongly dependent on the magma's viscosity, which depends largely on its silica content (Table 1.3). Basaltic magma, which contains about 50 wt% silica, has one of the lowest viscosities of common magmas; its viscosity (10^2 Pa s) is about the same as that of ketchup. Andesitic magma, with about 60% silica, has a viscosity similar to that of smooth peanut butter (10^3 Pa s). Rhyolitic (granitic) magma, with 70% silica, has a viscosity of at least 10^6 Pa s, which means it behaves like very thick tar. The viscosity of each of these magmas is increased substantially by the presence of crystals, with the viscosity becoming essentially infinite when the percentage of crystals approaches 60% (by analogy, the viscosity of chunky peanut butter is essentially infinite).

The viscosity of magma plays the dominant role in determining the shape of both intrusive and extrusive bodies of igneous rock (Table 1.3). Low-viscosity basalt is able to intrude rapidly into thin fractures (see cover of book) and, upon extrusion, can spread considerable distances before cooling and solidifying. In contrast, slowly moving highly viscous magma cannot travel far in thin bodies before cooling and solidifying. Consequently, these magmas must form thicker bodies if they are to move and, upon extrusion, commonly form dome-like bodies that inflate beneath a solidifying crust. As magma nears the surface of the Earth, decreasing pressure allows dissolved gases to escape, just as they do when the cap is removed from a bottle of carbonated beverage. Here again, the magma's viscosity plays an important role in determining the consequences of this exsolution. In low-viscosity magma, bubbles are able to move readily and escape from the magma.

However, in more viscous magma, the bubbles cannot escape so easily, and as they continue to grow, they eventually touch one another, at which point the magma can be disrupted and explosively ejected to form volcanic ash.

The composition of magma is determined largely by the mineralogy of the source region in which partial melting occurs, but some modification can occur during transit of the magma toward the Earth's surface. The source region for many bodies of granitic rocks is clearly exposed in the deeply eroded roots of mountain chains, where the minerals in metamorphic rocks testify to elevated temperatures that are sufficient to bring about partial melting. Water that is liberated by a number of metamorphic reactions that take place at successively higher temperatures acts as a flux to cause melting. The minerals taking part in these reactions also affect the composition of the granitic magma that is formed. At the lowest temperature, the breakdown of the aluminous mineral muscovite produces aluminous granites. Next the breakdown of biotite produces granites that do not contain excess alumina. At the highest temperatures, the breakdown of hornblende produces relatively silica poor granitic magma. The heat required to raise the geotherm to these melting temperatures is most likely brought into the base of the crust by the intrusion of basaltic magma from the mantle. Testament to this heat source are the almost omnipresent blebs of basaltic rock found in most granite.

Basaltic magmas are formed in the upper mantle, so their source region is not normally exposed, except in some convergent plate boundaries where the lithosphere has been thrust onto the continent instead of being subducted. Some basalt contains fragments of rock that have been brought up from the source region. This evidence indicates that basaltic magma is formed by the partial melting of rock composed primarily of olivine and pyroxene (*peridotite*). In addition to these main minerals, another important mineral is always present, but the nature of that mineral depends on the pressure. The second most abundant oxide in basaltic magmas is Al_2O_3. In crustal rocks, this oxide is found mainly in feldspar (e.g. $CaAl_2Si_2O_8$). However, on crossing the Moho, feldspar disappears and spinel ($MgAl_2O_4$) becomes the main aluminous mineral, and at still greater depth, spinel is replaced by garnet (e.g. $Mg_3Al_2Si_3O_{12}$). Where the reactions involving the change from one aluminous phase to another intersect the beginning-of-melting curve (solidus), low-temperature cusps are formed in the solidus. Consequently, when the geotherm is perturbed, partial melting preferentially occurs at depths corresponding to the intersections of these reactions with the solidus. For example, mid-ocean ridge basalt (MORB) is formed at the cusp marking the change from plagioclase to spinel. These basaltic magmas have sufficient silica so that when they crystallize they may contain a small amount of free quartz. In contrast, melting at higher pressures where garnet is stable produces basaltic magma that has lower silica content, and when these crystallize they may contain the silica-poor mineral nepheline ($NaAlSiO_4$).

The release of water from subducting slabs causes melting in both the overlying mantle wedge and continental crust. Experiments show that although melting of mantle peridotite under dry conditions produces basaltic magma, it produces andesitic magmas when wet. Consequently, as the oceanic slab is subducted and the water is released, andesitic magmas are formed. However, as the slab subducts to greater depths and the amount of water left to be expelled diminishes, the magma can change to basaltic composition. The rise of all these magmas and fluids into the crust steepens the geothermal gradient and causes metamorphism on a regional scale. This metamorphism liberates other fluids that transport heat still higher in the crust.

Without a doubt, the mid-ocean ridge system is the largest "rock factory" on the planet. However, very few of these rocks become part of the permanent geological record, because most ocean floor is subducted back into the mantle before it is more than 200 Ma old. Nonetheless, the rapid (geologically speaking) production and destruction of ocean floor is ultimately responsible for the production of rocks at convergent boundaries that do become part of the permanent geologic record. The igneous and metamorphic rocks formed near convergent plate boundaries owe their existence to the rapid cycling of H_2O and CO_2 from the ocean into alteration minerals in the ocean floor rocks and eventual release from the subducting slab. These processes are all driven by the Earth trying to eliminate geothermal perturbations created by the convection of the mantle as expressed in the lithosphere by plate tectonics.

1.8 PROBLEMS

1.1 Calculate and graph the pressure–depth relation in the outer part of the Earth where it consists of an upper 25-km-thick granitic layer of density 2.75 Mg m^{-3}, underlain by a 10-km-thick lower basaltic crust of density 3.0 Mg m^{-3}, which in turn is underlain by an upper mantle of density 3.3 Mg m^{-3}; this density continues to a depth of 400 km, where it increases to 3.6 Mg m^{-3}.

1.2 If the continental lithosphere described in Problem 1.1 is juxtaposed with an oceanic one consisting of 5 km of ocean water (ρ = 1.025 Mg m^{-3}) and 5 km of basaltic rock (ρ = 3.0 Mg m^{-3}), and both sections are underlain by mantle with density of 3.3 Mg m^{-3}, calculate the height of the continent above sea level, assuming isostatic equilibrium.

1.3 Calculate the radiogenic heat productivity of basalt in μW m^{-3} given the following data:

	e_i (mW kg^{-1})	c_i (ppm)
U	9.66×10^{-2}	0.9
Th	2.65×10^{-2}	2.2
K	3.58×10^{-6}	15 000

The density of basalt is 2.8 Mg m^{-3}. Compare individual contributions of U, Th, and K to the total heat productivity.

1.4 If the surface heat flow is 46 mW m^{-2}, compute a steady-state geotherm to 100 km for continental lithosphere having the following properties (take surface temperature to be 0 °C):

Layer	Thickness (km)	A (μW m^{-3})	K (W m^{-1} K^{-1})
1	10	2.1	2.51
2	30	0.26	2.51
3	60	0.0	3.35

1.5 To examine the effect of radioactive heat generation in the crust, consider a section of crust, 40 km thick, with heat productivity of 2.1 μW m^{-3} and a surface heat flow of 46 mW m^{-2}. K is taken to be 2.51 W m^{-1} K^{-1}. Take the surface temperature to be 0 °C. Compute the geotherm within the crust. Comment on the origin of the temperature maximum at a depth of 22 km.

1.6 Compute a steady-state geotherm for a 40-km-thick lithosphere in which the concentration of heat-generating radioactive components decreases exponentially with depth, that is, $A = A_0 e^{-z/D}$, where $A_0 = 2.1$ μW m^{-3}, $D =$ 40 km and the surface heat flow is 46 mW m^{-2}; K is taken to be 2.51 W m^{-1} K^{-1}. Take the surface temperature to be 0 °C. Compare the geotherm with that calculated in Problem 1.4.

1.7 The surface heat flow in the Basin and Range province of the western United States is unusually high – 87.86 mW m^{-2}. The heat flow–heat productivity relation for this region is modeled with a 9.4-km-thick crustal layer in which radioactive heat-producing elements are concentrated. For a radiogenic heat productivity for this layer of 2.1 μW m^{-3}, compute the reduced heat flow into the base of the 9.4-km-thick layer from the mantle below. Comment on the value of the reduced heat flow.

1.8 Taking the pressure at the surface of the Earth to be P_0, use a Taylor series to expand pressure (P_z) as a function of depth (z). Determine and discuss the physical significance of the first and second derivatives in this series expansion. To answer this question completely, it is necessary to use Eq. (3.4).

2 Physical properties of magma

2.1 INTRODUCTION

Igneous rocks are formed from molten material known as magma, which usually consists of a solution of the Earth's most abundant elements: oxygen and silicon with smaller amounts of aluminum, calcium, magnesium, iron, sodium, and potassium. Robert Bunsen (1851), of "Bunsen burner" fame, was the first to recognize that magmas were actually "solutions" no different from those of salts in water, except that they were hotter. Most magmas are silicate melts, but rarer ones contain little or no silica and, instead, are composed essentially of calcium carbonate, sulfide, or iron oxide. The major elements of common magmas combine, on cooling, to form the so-called *rock-forming minerals*: quartz, feldspars, feldspathoids, pyroxenes, olivine, and, when water is present, amphiboles, and mica. In addition, common *accessory minerals*, such as iron–titanium oxides, apatite, zircon, and sulfides, form from minor magmatic constituents (Fe^{3+}–Ti, P, Zr, and S, respectively) that do not readily enter the structures of the major rock-forming minerals.

The history of an igneous rock begins with the formation of magma at some depth in the Earth. The composition of the magma is determined by the chemical and mineralogical composition of the rock in the source region and by the process of melting. When sufficient melt has formed and coalesced, buoyancy causes it to rise. Further chemical modifications may occur during this period of transport. The magma may rise to the surface and extrude as lava or, if it contains volatiles, may explode onto the surface to be widely distributed as volcanic ash. Most magma, however, solidifies beneath the Earth's surface, where slow cooling and crystallization allow for further modification of the initial composition. During this final period, a rock develops its characteristic appearance, which we refer to as *texture*.

A discussion of igneous rocks could begin with a treatment of the initial melting processes in the source region, but this is a theoretical topic and is better left to a later chapter (Chapter 23). Instead, we shall begin with a more tangible subject, the physical properties of magma. It is important to have an accurate conception of the physical nature of magma before considering its emplacement and crystallization. Fortunately, active volcanoes provide opportunities to witness first hand the eruption of magma onto the Earth's surface. Here, such important properties as magmatic temperature, density, and viscosity can be measured. These will be discussed first. Other physical properties, such as heat capacity, thermal conductivity, and compressibility, can be determined only through sophisticated laboratory measurements. These properties will be touched upon in later chapters.

2.2 MAGMATIC TEMPERATURES

During periods of volcanic eruption, magmatic temperatures can be measured directly with optical pyrometers and thermocouples. Many eruptions, however, are too violent for this to be done safely. In these cases, magmatic temperatures can be determined by heating samples of the erupted rock in the laboratory and measuring their melting points. Such experiments were first done in 1790 by the Scottish geologist Sir James Hall (Fig. 2.1). He withheld publication of his results (Hall, 1805) until after the death of his distinguished friend James Hutton in 1797, who, while teaching that nature could be discovered only through observation, considered it rash "to judge of the great operations of the mineral kingdom from having kindled a fire and looked into the bottom of a little crucible" (www.lhl.lib.mo.us/events_exhib/exhibit/exhibits/vulcan/56.shtml). Although such distrust of experimental petrology did not die with Hutton, this approach is accepted today. Modern techniques permit accurate measurements to be made of melting temperatures of rocks under compositionally controlled atmospheres and pressures. Many such determinations have been made on a wide variety of rocks, and most agree well with direct measurements of magmatic temperatures in the field where such measurements can be made.

In discussing magmatic temperatures, it is important to keep in mind that a rock, being a mixture of minerals, does not have a single melting temperature. Instead, it melts over a temperature range that is commonly several hundred degrees. In addition, this melting range gives only the minimum temperature necessary to have liquid present; magmas could have temperatures well above this. However, one line of evidence indicates that magmas rarely exceed this temperature range. Magma that cools rapidly by coming in contact with cold rock or by being extruded on the Earth's surface crystallizes to a fine-grained aggregate of minerals or is quenched to a glass. Suspended in this material are almost always larger crystals known as *phenocrysts* (Fig. 2.2(B)). Because these crystals were growing prior to the quenching of the magma, their presence indicates that

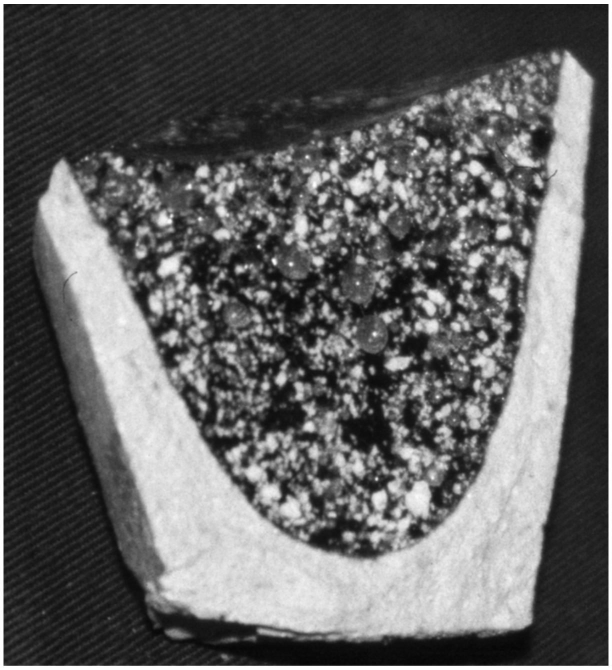


Fig. 2.1 One of James Hall's early experiments from which he tried "to judge of the great operations of the mineral kingdom from having kindled a fire and looked into the bottom of a little crucible." This section shows melted basalt.

most magmas have temperatures between the beginning and end of the melting interval. This observation is extremely important, as emphasized by Bowen (1928), for it indicates that magmas are rarely superheated. Most magma is a mixture of crystals and liquid. If the proportion of crystals is small, we describe the magma as having the properties of a suspension, which are predominantly those of a liquid. If the proportion of crystals is large enough that the crystals touch one another and form a three-dimensional network, we describe the magma as a mush, which has properties closer to those of a solid than to a liquid yet is still capable of moving. All magmas pass through the mush stage during cooling.

Extrusion temperatures for basalt, the type of lava commonly erupted from the Hawaiian volcanoes, commonly range from 1100 to 1200 °C. Higher temperatures do occur, but this results from the oxidation of magmatic gases. Most rhyolitic lava, the volcanic equivalent of granite, has temperatures between 800 and 1000 °C. Increased pressure causes melting temperatures to rise, but if water is present, the increased pressure allows water to dissolve in the magma, which in turn significantly lowers the melting temperatures. Basalt and rhyolite can both have their beginning of melting temperatures lowered by several hundred degrees by the addition of water.

Radiation cooling of the surface of lava flows causes solid crusts to form rapidly. The rate of conductive heat transfer through magma and solid crust is, however, so slow that the interior of a flow may remain hot for long periods. For example, in 1959, the eruption of Kilauea Iki on the Big Island of Hawaii produced a 100-m-deep lava lake on which a crust developed almost immediately (within hours). Over the next 20 years, the U.S. Geological Survey repeatedly drilled into the lava lake to study its solidification (Fig. 2.2(A)). The crust thickened with time (Fig. 2.3), but three years later, the base of the crust, which was at a temperature of 1065 °C, had grown to a depth of only 14 m. Even 19 years later, some liquid still remained amongst the olivine, plagioclase, and pyroxene crystals at depths of 75 to 85 m in the lake (Fig. 2.2(B)). Conductive cooling of magma is obviously a slow process in the lava lake, and in bodies of magma cooling at depth in the Earth, it would be still slower.

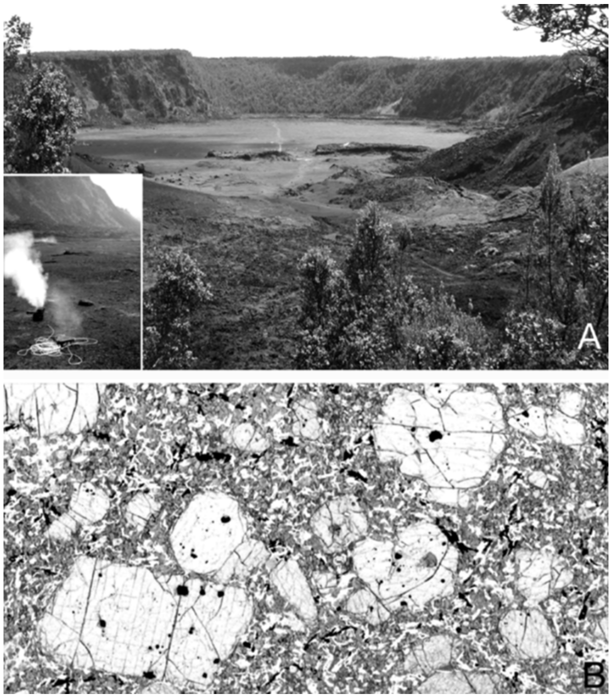


Fig. 2.2 The Kilauea Iki crater, Hawaii, was filled with a 100-m deep lava lake by an eruption in 1959. A crust formed on the lava lake within hours. As soon as the crust was strong enough to support the weight of a helicopter and a drill rig, the U.S. Geological Survey started a drilling program to monitor the solidification of the lava lake (see Fig. 2.3). (**A**) Panorama of crater in 2004. The vent from which the eruption occurred is in the right foreground. Inset shows steam rising from one of the drill holes in the lava lake in 1985. A thermocouple is attached to the white cable to measure the temperature in the lava lake. (**B**) Photomicrograph of a 1988 sample of drill core from 77 m below the surface of the lava lake showing large clear phenocrysts of olivine, in a groundmass of intermediate colored pyroxene, opaque plates of ilmenite, clear laths of plagioclase, and interstitial liquid that was quenched to a brown glass during the 1988 drilling. Width of field is 6.7 mm. Plane polarized light.

Several factors determine the cooling rate of magma, the most important of which are thermal gradients, rate of heat diffusion through magma and rock, heat capacity of magma and surrounding rocks, and the shape of magmatic bodies. A discussion of the relations between these factors is deferred to Chapter 5 to allow us first to consider the forms and mechanisms of emplacement of igneous bodies.

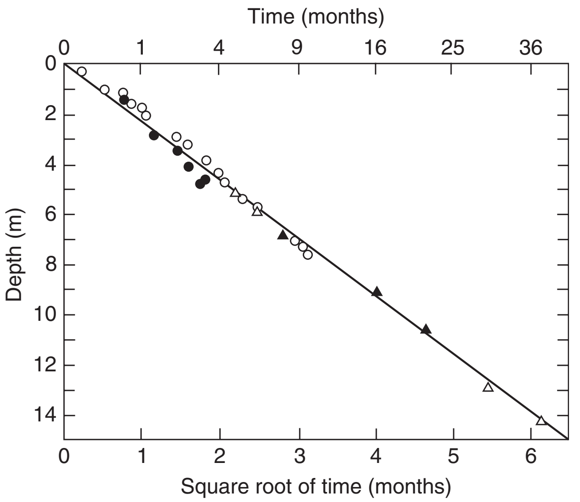


Fig. 2.3 Thickness of crust in the Alae and Kilauea Iki lava lakes, Hawaii. The base of the crust was defined as that depth at which the drill sank into the liquid. Solid circles represent the base of the crust, open circles the 1067 °C isotherm at Alae Crater and solid triangles the base of the crust, and open triangles the 1065 °C isotherm at Kilauea Iki Crater. The base of the crust was at approximately 85 °C below the initial extrusion temperature. The linear relation between the rate of crustal thickening and the square root of the length of time of cooling is discussed in Chapter 5 (after Peck *et al.*, 1964).

Table 2.1 *Partial molar volumes and coefficients of thermal expansion of common oxide components of silicate liquids containing 40 to 80 mol% SiO_2 at 1400 °C*

Oxide	Molar volume $10^6 \times V_i$ ($m^3\ mol^{-1}$)	Coefficient of expansion $10^5 \times \alpha$ (°C^{-1})^a
SiO_2	26.75	0.1
TiO_2	22.45	37.1
Al_2O_3	37.80	2.6
Fe_2O_3	44.40	32.1
FeO	13.94	34.7
MgO	12.32	12.2
CaO	16.59	16.7
Na_2O	29.03	25.9
K_2O	46.30	35.9

[a] Values in this column have been multiplied by 10^5. For example, the coefficient of expansion of SiO_2 is 0.1×10^{-5} °C^{-1} and its molar volume is $26.75 \times 10^{-6}\ m^3\ mol^{-1}$.
From Bottinga and Weill (1970) and Bottinga *et al.* (1982).

2.3 MAGMA DENSITIES

Many important petrologic processes are controlled by the density of magma. For example, buoyancy largely determines the rise of magma from a source region. Indeed, magmas that ascend through the lithosphere may have a limited compositional range that is determined by magma density; the lithosphere may act effectively as a density filter. Crystals may float or sink in a magma, and by so doing they can change the composition of the remaining magma (differentiation); this process is clearly determined by contrasts in crystal and liquid densities. Where magma chambers are periodically replenished with surges of fresh magma, as for example beneath oceanic ridges, mixing of magma may occur if magma densities are closely matched, but if densities are different, the magmas remain separate. Magma mixing, which plays an important role in generating certain rock types, some of which are of economic significance, is thus controlled, in part, by magma densities.

The density of magma can be determined in several ways. First, it can be measured directly at high temperatures. This is technically difficult; so, instead, density measurements are often made at room temperature on glasses formed from the rapid quenching of these liquids. Because glass is a supercooled liquid, the high-temperature density can be calculated from the room temperature measurements if the coefficient of thermal expansion is known. If V_1 is a volume at T_1, the temperature at which the density measurement is made, the volume, after heating to temperature T, is given by

$$V_T = V_1[1 + \alpha(T - T_1)] \tag{2.1}$$

where α, the coefficient of thermal expansion at constant pressure, is defined as

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P \tag{2.2}$$

The *coefficient of thermal expansion* is the relative increase in volume *($\partial V/V$)* per unit increase in temperature at constant pressure (Table 2.1). The partial derivative $(\partial V/\partial T)_P$ is used instead of the ordinary derivative to show that this relation is restricted to the change at constant pressure.

Because density is mass per unit volume, Eq. (2.1) can be rewritten as

$$\rho_T = \frac{\rho_1}{1 + \alpha(T - T_1)}$$

where ρ_1 and ρ_T are the densities at temperatures T_1 and T, respectively. Multiplying the top and bottom of the right-hand side of this equation by $[1 - \alpha\,(T - T_1)]$ gives

$$\rho_T = \frac{\rho_1[1 - \alpha(T - T_1)]}{1 - [\alpha(T - T_1)]^2}$$

Because α is very small (about 10^{-5} °C^{-1}), the squared term is negligible, and the denominator is essentially one. Hence, the density at any temperature T is related to the density at a temperature T_1, by

$$\rho_T = \rho_1[1 - \alpha(T - T_1)] \tag{2.3}$$

Another method of determining magma density uses a chemical analysis of the magma, and determines how much each constituent contributes to the total density by using what are known as the *partial molar volumes* of each constituent. These were determined by Bottinga and Weill (1970) and Bottinga and others (1982) from the analysis of density measurements of silicate liquids (Problem 2.1).

In any solution, such as a magma, the partial molar volume of a component i, denoted by the symbol $\bar{V}_i$, is defined as the change in volume resulting from a change in the number of moles of component i (n_i) at constant temperature, pressure, and numbers of moles of all other components. This is expressed mathematically as

$$\bar{V}_i \equiv \left(\frac{\partial V}{\partial n_i}\right)_{T,P,n_{j\neq i}} \tag{2.4}$$

The partial derivative is used to show that this expression is restricted to the volume change at constant temperature, pressure, and number of moles of all constituents except component i. The partial molar volumes of the common oxide constituents of magma are given in Table 2.1 for 1400 °C. To obtain volumes at other temperatures, the coefficient of thermal expansion is used along with Eq. (2.2).

The volume of a magma consisting of many components is given by

$$V = \bar{V}_a n_a + \bar{V}_b n_b + \cdots + \bar{V}_i n_i = \sum_i \bar{V}_i n_i \tag{2.5}$$

Thus, the partial molar volume of a component is that quantity which, when multiplied by the number of moles of the component, gives the contribution of the component to the total volume.

Compositions are often more conveniently expressed in terms of mole fractions than as numbers of moles of constituents. The *mole fraction* of a component i, denoted by X_i, is simply $n_i/(n_a + n_b + \cdots + n_i)$. Consequently, the number of moles of each component in Eq. (2.5) can be converted to mole fractions by dividing both sides of the equation by $(n_a + n_b + \cdots + n_i)$, giving

$$\frac{V}{n_a + n_b + \cdots + n_i} = \bar{V}_a X_a + \bar{V}_b X_b + \cdots + \bar{V}_i X_i$$

The volume of magma per total number of moles of all components is referred to as the *molar volume* and is denoted by $\bar{V}$. Hence,

$$\bar{V} = \sum_i \bar{V}_i X_i \tag{2.6}$$

The density of a magma is obtained by dividing this molar volume into the molecular weight of the magma, which is given by $\sum_i M_i X_i$, where M_i is the molecular weight of component i. Thus

$$\rho = \frac{\sum_i M_i X_i}{\sum_i \bar{V}_i X_i} \tag{2.7}$$

Another way of calculating the density of magma from a chemical analysis is by using the computer programs MELTS (Ghiorso and Sack, 1995; Asimow and Ghiorso, 1998) for low pressure, and pMELTS (Ghiorso *et al.* 2002) for higher pressures between 1 and 3 GPa. These programs, which can be downloaded from http://melts.ofm-research.org, use an internally consistent thermodynamic data set that was calibrated with a large number of careful experimental results for a wide compositional range of liquids coexisting with common rock-forming minerals. These programs will be discussed in more detail in Chapter 8, but for the moment, these data provide the molar volume of the liquid from which the density is calculated. After a liquid's composition is entered as oxides and the desired temperature, pressure, and oxygen fugacity (approximately the partial pressure of oxygen) selected, the programs return the properties of the liquid, including density, under the selected conditions (Problem 2.4).

A final method of obtaining the approximate density of magma is based on the fact that the volume expansion on melting for most rocks is ~10%. The density of most magmas, therefore, is 90% of the density of the equivalent solid rock.

The densities of common magmas near the surface of the Earth vary from 2.3 to 3.0 Mg m^{-3}, with the more iron-rich varieties having the higher values. Most basaltic magmas have densities of 2.60 – 2.65 Mg m^{-3}, which is similar to that of plagioclase feldspar (2.63 – 2.76 Mg m^{-3}) but less than that of pyroxene and olivine (3.0 – 3.7 Mg m^{-3}). Most granitic magmas have a density of ~2.4 Mg m^{-3} and are therefore less dense than any silicates (Problem 2.3).

Because most magmas are mixtures of liquid and crystals, their bulk density will be greater than if they consisted entirely of liquid. To determine the bulk density of a magma it is necessary to know the density and proportions of the liquid fraction and of each of the solid phases present. The MELTS program provides you with all of these values and the calculated bulk density.

Under the higher pressures at depth in the Earth, magmas will be compressed and hence have a higher density. To determine the magnitude of this effect we make use of the *isothermal coefficient of compressibility*, which is denoted by β. This coefficient gives the relative decrease in volume $(-\partial V/V)_T$ per unit increase in pressure at constant temperature. This is expressed mathematically as

$$\beta \equiv -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T \tag{2.8}$$

The inverse of this coefficient is known as the *bulk modulus*, which has units of Pa. Typical β values for magmas are of the order of 5×10^{-11} Pa^{-1} (Bass, 1995), which translates to a density change of only ~0.13 Mg m^{-3} per GPa. For example, experiments show that the density of a Kilauea basaltic magma decreases from 2.743 to 2.610 Mg m^{-3} in rising from a depth of 35 km (Kushiro, 1980). If increased pressures result in higher dissolved water contents, the resulting change in composition results in significantly lower magma densities (Bottinga and Weill, 1970).

Although magmas have low values of β, they are more compressible than solids, and under the extremely high pressures in the mantle their densities increase and may even become greater than that of the solids from which they are formed. Experiments indicate that at pressures greater than 13.5 GPa (equivalent to a depth of 400 km) melts of "basaltic" composition formed from the partial melting of peridotite have greater densities than do the olivine crystals in the peridotite (Agee and Walker, 1993; Suzuki and Ohtani, 2003). At shallower depths, basaltic magma would rise relative to olivine, but below this depth it would sink, which raises the interesting possibility that magma might sink to the base of the lower mantle and become part of the D″ layer (Section 1.2; Ohtani and Maeda, 2001).

2.4 MAGMA VISCOSITIES

Some lavas are able to flow rapidly over great distances, whereas others barely move, even when erupted on steep slopes. Viscosity is the physical property that describes this resistance to flow (Fig. 2.4). Viscosity is extremely important in determining the rates of emplacement of magma and the shapes of igneous bodies, and in determining whether sinking or floating crystals will separate fast enough from their parent magma to change the magma's bulk composition (Dingwell, 2006).

Like any fluid, magma deforms continuously under the action of a shear stress, but the deformation may be very slow in many magmas. When a fluid is at rest, there should be no shear stresses (see yield strength below). Fluids react to a shear stress by flowing, but some do this more slowly than do others. The property of a fluid that defines the rate at which deformation takes place when a shear stress is applied is known as *viscosity*:

$$\text{viscosity} \equiv \frac{\text{shear stress}}{\text{rate of shear strain}}$$

If a large shear stress applied to a liquid results in slow deformation, the liquid is said to have a high viscosity. Tar and honey are familiar examples of highly viscous liquids, whereas water and gasoline have low viscosities (Table 2.2).

As the magnitude of a shear stress applied to a liquid increases, the rate of strain also increases. For many liquids, a linear relation exists between the applied shear stress and the strain rate (Fig. 2.5). These liquids are referred to as being *Newtonian*. From the definition of viscosity, the slope in Figure 2.5 is clearly the viscosity. *Non-Newtonian* liquids exhibit a nonlinear relation between applied shear stress and strain rate; their viscosities (slopes) vary with shear stress. Most magmas, at least when totally liquid and deformed at a low strain rate, show near Newtonian behavior. As the percentage of crystals in a magma increases, the bulk viscosity of the mixture also increases. If the fraction of crystals present in the magma is F, the effective bulk viscosity of the magma, η_B (eta), is related to the viscosity of the liquid, η_L, by

$$\eta_B = \eta_L(1 - 1.65F)^{-2.5} \tag{2.9}$$

It will be seen that as F approaches 60%, the bulk viscosity approaches infinity. This is known as the *Einstein limit* in recognition of his work in this field. As a magma approaches this limit, it would behave essentially as a solid, despite the presence of 40% liquid.

Not only do crystals, or bubbles, in a magma increase viscosity, they also commonly cause the magma to exhibit non-Newtonian behavior of the type exhibited by *Bingham liquids* (Shaw, 1969; Petford, 2003). These liquids resist deformation until some minimum shear stress (*yield strength*) is applied, after which strain rates increase linearly with shear stress (Fig. 2.5). Thus, at low shear stresses, they appear to have infinite viscosity, but above the yield strength, they behave as Newtonian liquids. A common example of such a liquid is household paint. When one applies a thin coat of

Table 2.2 *Viscosities of magmas and common substances*

Material	Viscosity (Pa s)	Wt% SiO_2	Temp. (°C)
Water	1.002×10^{-3}	–	20
ASE 30 motor oil	2×10^{-1}	–	20
Kimberlite	10^{-1}–1	30–35	~1000
Komatiite	10^{-1}–10	40–45	1400
Ketchup	~5 × 10	–	20
Basalt	10–10^2	45–52	1200
Peanut butter	~2.5×10^2	–	20
Crisco® shortening	2×10^3	–	20
Andesite	~3.5×10^3	~58–62	1200
Silly Putty®	~10^4		
Tonalite 6% H_2O	~10^4	65	950
Rhyolite	~10^5	~73–77	1200
Granite 6% H_2O	~10^5	75	750
Rhyolite	~10^8	~73–77	800
Average mantle	10^{21}	–	–

Magma viscosities from Dingwell (1995) and references therein. Granite and tonalite viscosities from Petford (2003). Mantle viscosity is from King (1995).

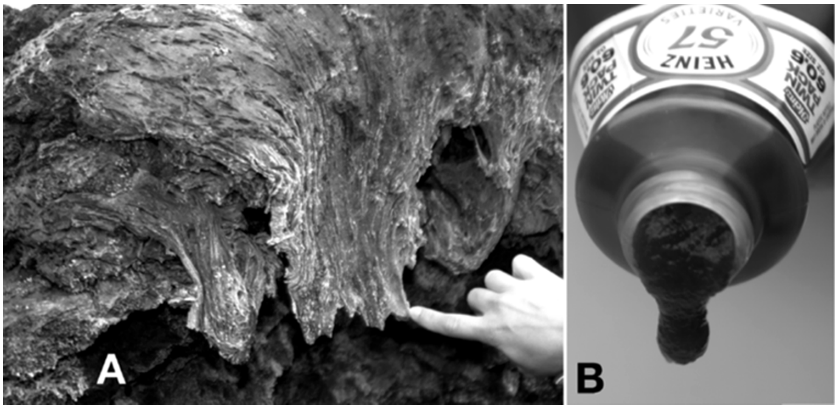


Fig. 2.4 (**A**) Basaltic spatter frozen while flowing back into the 1984 fissure vent at the Krafla volcano, Iceland. Basaltic magma is the most fluid of the common magmas, yet from the shape of these dribbles, it can be seen to be a viscous liquid. (**B**) For comparison, ketchup, which has approximately the same viscosity as basalt (Table 2.2), is shown dripping from a bottle.

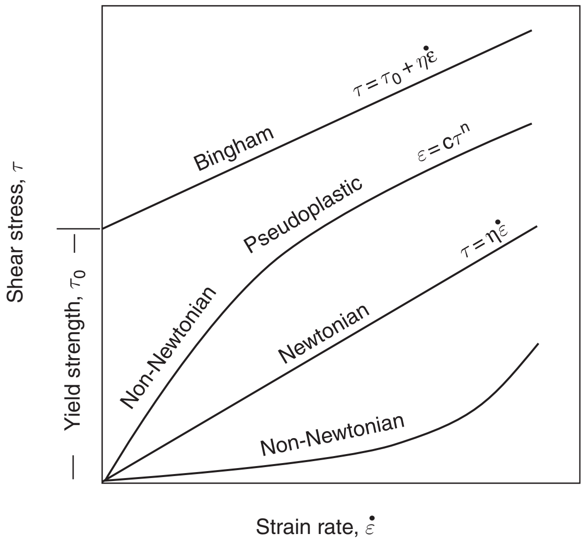


Fig. 2.5 Shear stress versus rate of shear strain for Newtonian and non-Newtonian liquids. The slope of these lines gives the viscosity of the liquid, which for a Newtonian liquid remains constant regardless of shear stress. The viscosity of non-Newtonian liquids varies with the applied shear stress.

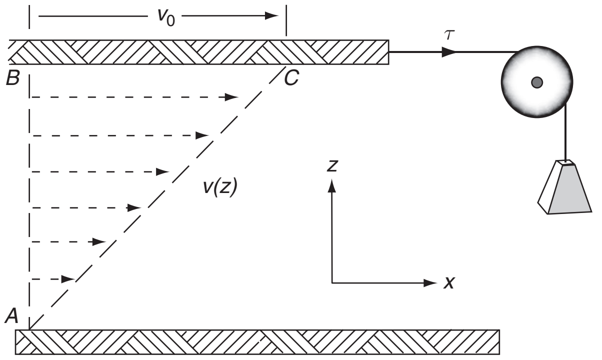


Fig. 2.6 Laminar flow of liquid between two parallel plates resulting from the upper plate moving in the *x* direction with a velocity v_0 relative to the stationary lower plate when a shear stress, v, is applied. Note that the positive direction of *z* is upward (height).

paint to a vertical wall, the paint does not dribble (hopefully), because the weight of the paint is insufficient to overcome the paint's yield strength, which is commonly due to the presence of a fine suspension of mica crystals. However, if too thick a coat is applied, the yield strength is exceeded, and the paint begins to flow and dribble. A good paint that gives "no drips, no runs, and no errors" has a significant yield strength. Bingham behavior has profound effects on velocity profiles across bodies of flowing magma (Section 3.7) and on the ability of crystals to move buoyantly through magma (Section 14.3). Even when magmas do not contain crystals, Newtonian behavior may occur only at low strain rates. At higher strain rates, many magmas exhibit what is known as *shear thinning*; that is, they experience a decrease in viscosity with increasing shear rate (Dingwell, 1995). In this respect, they tend to show pseudoplastic behavior (Fig. 2.5).

Another type of non-Newtonian liquid is one that has low viscosity at low shear stresses (shallow slope in Fig. 2.5) but a high viscosity at high shear stresses. Silly Putty® is a common example of such a liquid. If you gently pull on a piece of Silly Putty, it stretches with almost no resistance, but if you suddenly pull forcefully, it resists and will probably fracture as if it were a solid that has an extremely high viscosity.

Viscosity is better appreciated by considering the following example. A shear stress τ (tau) is applied to the liquid between two parallel plates (Fig. 2.6) by moving the upper one steadily at a velocity v_0 relative to the lower one. Liquid in contact with the upper plate is dragged along at a velocity v_0, whereas that in contact with the lower plate remains stationary. The intervening liquid can be thought of as consisting of a series of thin parallel plates or lamellae, each of which slides over the one beneath. Such movement is referred to as *laminar flow*, in contrast to *turbulent flow*, where both the direction and rate of movement of the liquid experience wide fluctuations. If the length of the arrows in the figure indicates the velocity in the *x* direction at various heights along the *z* axis, a marker line *AB* would move to the position *AC* after a unit interval of time. The strain rate must therefore be dv/dz, and the viscosity, denoted by η, is given by

$$\eta = \frac{\tau}{dv/dz} \quad \frac{\mathrm{N}}{\mathrm{m}^2}\frac{\mathrm{s}}{\mathrm{m}}\mathrm{m} = \mathrm{Pa\ s} \tag{2.10}$$

In SI units, where shear stress is measured in newtons m^{-2}, the velocity in m s^{-1}, and distance in meters, the units of viscosity are Pa s. The cgs unit of viscosity is the *poise* (P), which is a dyne s cm^{-2} (1 P = 0.1 Pa s). In many fluid mechanics calculations, the viscosity is divided by the density of the fluid. This ratio is referred to as the *kinematic viscosity* and is defined as $v = \eta/\rho$ with units of $\mathrm{m}^2\ \mathrm{s}^{-1}$.

It is now possible to make a quantitative statement about the viscosity of magmas (Table 2.2). Measurements of viscosity have been made in the laboratory at both atmospheric and higher pressures on a variety of liquids formed by melting common igneous rocks. In addition, viscosities have been calculated from the actual measured rates of flow of lava issuing from active volcanoes. Common magmas cover a wide range of viscosities (Dingwell, 1995). At a temperature of 1200 °C and atmospheric pressure, basaltic magmas (gabbroic) have viscosities in the range 10 to 100 Pa s. This is about the same as tomato ketchup (Fig. 2.4), but the pulp in ketchup tends to make it behave as a Bingham liquid, which is why we often need to give an extra tap when pouring it to overcome the yield strength. The volatile-rich silica-poor magma kimberlite, which brings diamonds to the surface from great depth, has a viscosity of slightly less than 1 Pa s at temperatures of approximately 1000 °C. The magnesium rich, silica-poor lava komatiite which erupted only in the Archean had a viscosity of ~1 Pa s at its eruption temperature of ~1400 °C. Andesitic magmas (dioritic) at 1200 °C average about 10^3 Pa s, which is slightly greater than smooth peanut butter and about the same as Crisco® shortening. Most rhyolitic magmas (granitic) have viscosities above 10^5 Pa s and can be as high as 10^8 Pa s at 800 °C, which is a more normal magmatic temperature for rhyolite.

The viscosity of a magma is determined largely by its chemical composition, which controls its degree of polymerization. Just as silicate minerals show different degrees of polymerization, so do silicate melts. Most of the oxygen atoms in silica-rich melts are tetrahedrally (T) coordinated around silicon atoms. These silica-oxygen bonds are strong and give the melt a high viscosity. As the silica content decreases and the abundance of Fe, Mg, Ca, etc. increases so do the number of nonbridging oxygens (NBO). These are weaker bonds than the tetrahedral ones, and they therefore lower the viscosity. The degree of polymerization of a silicate melt is commonly given by the ratio of the average number of nonbridging oxygens to the number of tetrahedrally coordinated cations (NBO/T). Thus, in a melt of albite ($NaAlSi_3O_8$) composition, NBO/T = 0, whereas in a melt of diopside composition ($CaMgSi_2O_6$), NBO/T = 2 (draw a silica chain and check this ratio).

The viscosity of a silicate melt is controlled largely by the abundance of the polymerizing, tetrahedrally coordinated groups, SiO_2, $KAlO_2$, and $NaAlO_2$. The increase in viscosity from komatiite through basalt and andesite to rhyolite is attributable to this chemical variation, with komatiites containing approximately 40–45 wt% SiO_2, basalts ~50 wt% SiO_2, andesites ~60 wt%, and rhyolites ~73 wt%. This chemical control makes it possible to calculate reasonable viscosities for magmas of known composition, in much the same way that densities were calculated from partial molar volumes (Bottinga and Weill, 1972). The MELTS program also calculates the viscosity of melts based on their chemical composition (Problem 2.6).

Temperature has an enormous effect on viscosity, as anyone knows who has tried to start a car on a cold winter's day. The oil in the engine becomes extremely viscous at low temperatures. By heating the engine only slightly, for example by hanging a light bulb under the hood, the viscosity drops remarkably and starting becomes much easier. Viscosity is an exponential function of temperature. The viscosity of rhyolitic magma, for example, is shown in Table 2.2 as changing from 10^5 Pa s at 1200 °C to $>10^8$ Pa s at 800 °C (Problem 2.6c).

The range of viscosities of magmas becomes even greater when account is taken of the lower temperatures at which rhyolites, for example, crystallize compared with those at which basalt crystallizes. Rhyolitic magma is typically several hundred degrees cooler than basaltic magma and consequently has very much higher viscosity than does basalt. This single fact is largely responsible for the morphological differences between volcanoes and other igneous bodies formed from these different magmas.

The dramatic change in viscosity of magma with temperature generally obeys an Arrhenius relation of the form

$$\eta = \eta_0 e^{E/RT} \tag{2.11}$$

where η_0 is a constant, E the activation energy, R the gas constant, and T the absolute temperature. Such a relation, which describes the temperature dependency of many reaction rates, can be thought of as consisting of two parts; the preexponential factor η_0 expresses the frequency of a particular event involved in the reaction (in this case, perhaps the formation of a hole or the breaking of a bond in the liquid structure), and the exponential term gives the fraction of these events that have sufficient energy to permit the reaction to proceed.

Converting Eq. (2.11) to logarithmic form gives

$$\ln \eta = \ln \eta_0 + \frac{E}{RT}$$

which is the equation of a straight line. A plot of $\ln \eta$ versus $1/T$ for viscosities determined at several temperatures yields the activation energy from the slope of the line and the frequency factor from the intercept (see Problem 2.7). Once the values of η_0 and E are known, the viscosity can be calculated for any temperature.

Pressure has a very much smaller effect on magma viscosity than has temperature. The viscosity of a basaltic magma at 1400 °C, for example, decreases from 3.5 Pa s at atmospheric pressure to 1.5 Pa s at 2.0 GPa (Kushiro, 1980). Indirectly, however, pressure can modify viscosities more dramatically. Higher pressure permits more water to be dissolved in magmas, which depolymerizes the melt and lowers the viscosity (Shaw, 1965). The effect of water (independent of pressure) is demonstrated, for example, in experiments carried out on a melt of albitic composition at 1550 °C and a pressure of 2.5 GPa (Poe *et al.*, 2006). This melt, when dry, has a viscosity of 114 Pa s, but with 2.8 wt% H_2O, its viscosity decreases to 16.4 Pa s. Granitic melts containing ~6 wt% H_2O at a pressure of 0.8 GPa and 750 °C have a viscosity of ~10^5 Pa s (Petford, 2003). For more references to the literature on viscosity see Dingwell (2006).

Unless one is familiar with the numerical values for the viscosity of liquids, numbers such as 10^2 or 10^8 Pa s have little significance. They can, however, be made more meaningful by considering the effect that viscosity has on the flow rate of lava. Let us consider a lava flow (Fig. 2.7) of thickness h composed of an incompressible Newtonian magma with viscosity η and density ρ that flows with a steady rate down an unconfined plane surface with a slope of θ degrees. For convenience we shall choose a set of orthogonal axes, x, y, and z, such that x parallels the flow direction and z is normal to the plane surface. The lava flows in a laminar manner and for the moment we will neglect any effects due to cooling of the lava. Gravitational attraction causes the lava to flow, and if no other forces are present, the lava accelerates to greater and greater velocities. The viscous force, of course, prevents this from happening.

To analyze this situation we consider the forces acting on a very small volume of lava, dx dy dz. The gravitational force acting in the direction of flow is simply the mass times the acceleration, which is given by ρ dx dy dz $g \sin \theta$, where g is the acceleration of gravity. However, the shear stress in the liquid retards the flow. Because the shear stress varies with height in the flow, there is a difference in shear stress between the top and bottom of the volume, which we will designate dτ. Because this stress (force/unit area) acts on an area d d , the

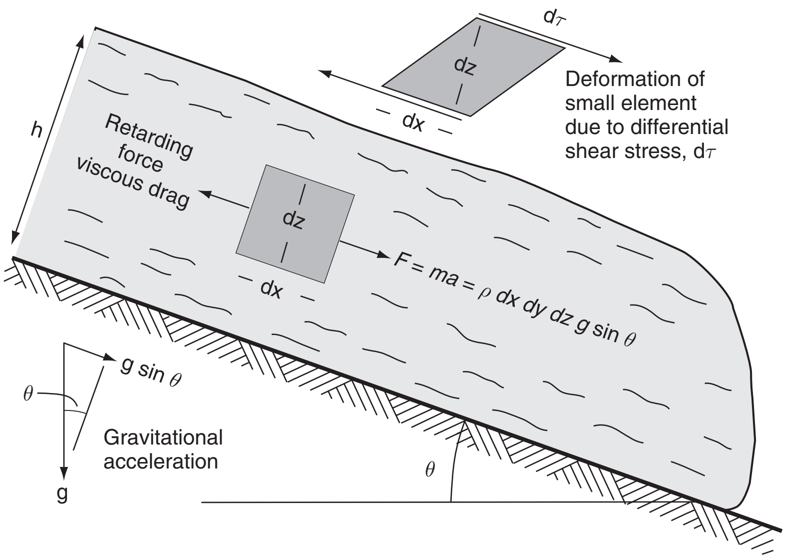


Fig. 2.7 Forces acting on a small element (d*x* d*y* d*z*) of lava within a flow of thickness *h*, descending a plane surface with a slope of *θ* degrees.

force retarding the flow is dτ d*x* d*y*. Because the flow rate of the lava is steady, there is no acceleration. Consequently, the sum of the viscous and gravitational forces acting in the plane of flow must be zero; that is,

$$d\tau\, dx\, dy + \rho\, dx\, dy\, dz\, g\, \sin\theta = 0$$

Dividing both sides of the equation by *dx* and *dy* and rearranging gives

$$d\tau = -\rho g\, \sin\theta\, dz$$

which, upon integration, becomes

$$\tau = -\rho g \sin\theta z + \text{constant}$$

We can determine the value of the constant of integration because we know that the shear stress is zero ($\tau = 0$) on the surface of the flow ($z = h$). The constant therefore equals $\rho g \sin\theta\, h$. Hence the shear stress at any point in the flow is given by

$$\tau = \rho g \sin\theta(h - z)$$

Because the lava is a Newtonian liquid in laminar flow, $\eta dv/dz$ can be substituted for τ (Eq. (2.10)), giving

$$\frac{dv}{dz} = \frac{\rho g \sin\theta}{\eta}(h - z) \tag{2.12}$$

This gives the velocity profile through a lava flow under the specified conditions. Integration of this equation gives

$$v = \frac{\rho g \sin\theta}{\eta}\left(hz - \frac{z^2}{2}\right) + \text{constant}$$

This integration constant is zero, because $v = 0$ where $z = 0$. Hence, the velocity at any height *z* is given by

$$v = \frac{\rho g \sin\theta}{\eta}\left(hz - \frac{z^2}{2}\right) \tag{2.13}$$

It is now possible, therefore, to use the viscosities of magmas to calculate flow rates for some common lavas. This is left for the reader to do in Problem 2.9 to 2.11.

When integrating Eq. (2.12), the viscosity was taken to be a constant. But, in reality, the magma would cool along the lower and upper surfaces of the flow, and this would cause a significant increase in the viscosity, which, in turn, would decrease flow velocities. To improve the calculated velocities, it would be necessary, before integrating Eq. (2.12), to express η as a function of height within the flow. This can be done, because a reasonable estimate of the temperature gradient within the flow can be made (see Chapter 5), and if the activation energy in Eq. (2.11) is known, the viscosity can be expressed as a function of *z* (see Problem 2.11).

2.5 PROBLEMS

2.1 Two lime-silica liquids at 1700 °C, one with 30 and the other with 60 mole % CaO, have densities of 2466 and 2665 kg m^{-3}, respectively. Determine the partial molar volumes of SiO_2 and CaO in these liquids at this temperature. (Note: Your answers may differ slightly from the values determined from Table 2.1, because the table is based on average values determined from many calculations such as those done in this problem.)

2.2 The molar volumes of SiO_2 and CaO at 1200 °C are 26.744×10^{-6} and 16.05×10^{-6} m^3 mol^{-1}, respectively, whereas at 1600 °C they are 26.755×10^{-6} and 17.15×10^{-6}m^3 mol^{-1}, respectively. Calculate the coefficients of expansion, α, for SiO_2 and CaO over this temperature range and give an explanation, in terms of the possible structures of the melts, for the large difference between these two values. (Reminder: $\int dV/V = \ell n V + \text{constant}$.)

2.3 Using the partial molar volumes in Table 2.1, calculate the densities of typical rhyolitic (granitic) and basaltic (gabbroic) magmas at 1200 °C having the compositions in Table 2.3. (Note that the compositions are expressed in weight percent of the oxides, and these must first be converted to mole fractions. The iron in these analyses is expressed as FeO for simplicity. Natural magmas contain both ferrous and ferric iron, the proportions depending on the degree of oxidation of the magma.)

Table 2.3 *For Problem 2.3*

	Weight % oxides	
Oxide	Rhyolite	Basalt
SiO_2	74.0	48.7
TiO_2	0.2	2.8
Al_2O_3	13.5	15.0
FeO	2.0	12.0
MgO	0.3	7.5
CaO	1.2	10.5
Na_2O	3.4	2.5
K_2O	5.4	1.0
Total	100.0	100.0

2.4 Using the MELTS program, calculate the density of the rhyolitic and basaltic magmas given in Problem 2.3 at 1200 °C and a pressure of 1 bar (~1 atmosphere). Set the oxygen fugacity to the quartz-fayalite-magnetite (Q-Fa-Mt) oxygen buffer (this is a common value for many magmas). Compare your results with those obtained in Problem 2.3 using Bottinga and Weill's partial molar volumes.

2.5 A sample of granite with density 2670 kg m^{-3} is fused and quenched to a glass that has a density of 2450 kg m^{-3} at 25 °C. The coefficient of expansion, α, for the glass is $2 \times 10^{-5}\,°C^{-1}$. Determine the density of the magma at 1025 °C, and compare this with the value obtained by simply taking 90% of the density of the rock.

2.6 **(a)** Using the MELTS program, calculate the viscosities of the rhyolitic and basaltic magmas given in Problem 2.3 at 1200 °C, a pressure of 1 bar (~1 atmosphere), and oxygen fugacity of the quartz-fayalite-magnetite (Q-Fa-Mt) oxygen buffer. Note that the MELTS program gives viscosity as $\log_{10}$ in poise. You will have to convert to Pa s.

(b) Why are the viscosities of the two melts so different?

(c) Calculate the viscosity of the rhyolite for the same conditions except at a temperature of 800 °C.

(d) Assuming the variation in viscosity with temperature obtained in parts (a) and (c) obeys an Arrhenius relation, calculate the rhyolite's activation energy for viscous flow. Be sure to convert temperatures to kelvin.

(e) Although pressure has little effect on viscosity, higher pressures allow magmas to dissolve more water, which does lower the viscosity. At 1 GPa, granitic magma can dissolve up to ~10 wt% H_2O. Recalculate the granite analysis in part (a) to include 10 wt% H_2O, and then calculate the viscosity at 1200 °C, 10 kb, under the Q-Fa-Mt oxygen buffer. By comparing your answer in part (a), comment on the effect of water on the viscosity of this magma.

2.7 The viscosity of a particular lava is found by experiment to vary with temperature in the following manner:

T (°C)	1100	1200	1300	1400
η (Pa s)	700	100	20	5

By plotting ln η versus $1/T$ and obtaining a least-squares best fit to the data, determine the value of the activation energy and the preexponential factor in the Arrhenius expression for the temperature dependency of viscosity. (Temperature must be expressed in kelvin.)

2.8 From the equation for the velocity distribution through a lava flow that is in steady-state laminar flow, derive an expression for the maximum velocity and show that this occurs at the upper surface of the flow. (Hint: Check your calculus book for maximum–minimum problems.)

2.9 Plot a graph of the velocity profile through a 2-m-thick basaltic lava flow with a viscosity of 300 Pa s and density 2750 kg m^{-3} descending a 2° slope. What is the shape of this profile?

2.10 Calculate the maximum velocity within the flow given in Problem 2.9. Do the same for a similar flow of rhyolite with a viscosity of 10^6 Pa s and density 2500 kg m^{-3}, and a lunar basalt from the Sea of Tranquility having a viscosity of 9 Pa s and a density of 3000 kg m^{-3}. Express your own maximum rate of walking (approximately 4 miles/hour) in the same units of velocity. Why would the flow velocities be too high? (Note that the acceleration of gravity on the Moon is only 1.62 m s^{-2}.)

2.11 Calculate the velocity profile through the same lava flow as given in Problem 2.9, but because of cooling toward the upper and lower surfaces, the viscosity increases away from the center of the flow. The viscosity at any height z, measured in meters, above the base of the flow is given by $\eta = \eta_c + f(z - h/2)^2$, where η_c, the viscosity at the center of the flow, is 300 Pa s, f is a constant with the value of 10^4 N s m^{-4}, and h is the thickness of the flow. Useful integrals are:

$$\int \frac{dz}{a + bz + cz^2} = \frac{2}{(4ac - b^2)^{1/2}} \tan^{-1} \frac{2cz + b}{(4ac - b^2)^{1/2}} + \text{constant} \quad \text{when } 4ac - b^2 > c$$

$$\int \frac{z\,dz}{a + bz + cz^2} = \frac{1}{2c} \ln\left(a + bz + cz^2\right) - \frac{b}{2c} \int \frac{dz}{a + bz + cz^2} + \text{constant}$$

3 Intrusion of magma

3.1 INTRODUCTION

Most field evidence indicates that igneous rocks have formed from upward-moving bodies of either magma, mixtures of magma and crystals, magma and gas bubbles, or even solid rock. Basaltic lava flows cover most of the ocean floor and large areas on continents. Rhyolitic lava flows and volcanic ash also cover large parts of continents. Clearly, magma of a wide range of composition is able to rise to the Earth's surface.

Exposures of deeply eroded parts of the crust reveal that basaltic magmas most commonly rise through fractures and, on cooling, form steeply dipping sheet-like intrusions known as *dikes*. Rhyolitic magmas, in contrast, tend to rise in large dome-like bodies known as *diapirs*, which, on solidifying beneath the surface, form granite batholiths. Still other magmas with high volatile contents such as kimberlite, which bring diamonds up from depths of approximately 200 km, rise through pipe-like bodies known as *diatremes* and erupt explosively onto the Earth's surface.

Regardless of the mode by which magma ascends, buoyancy is the main driving force. Magmas are simply less dense than most rocks and so tend to float toward the Earth's surface. When the upper mantle partially melts, a basaltic liquid is formed that is less dense than the refractory residue, so the liquid begins to rise. It does so slowly at first as an interstitial liquid moving through the pores of the deformable crystal mush (McKenzie, 1984). The flow then becomes channelized (Kelemen *et al*., 1995), and the pressure causing the magma to rise is sufficient to propagate fractures through the lithosphere. As the magma rises in these fractures, cracking at the fracture tip generates a characteristic seismic signal known as *harmonic tremor*, which can be used to track the advance of the magma (McNutt, 2005; Roman and Cashman, 2006; Burlini *et al*., 2007). Velocities of up to several meters per second have been recorded in this manner (Sigurdsson and Sparks, 1978; Aki and Koyanagi, 1981; Wright *et al*., 2006). Velocities have also been estimated from the width of dehydration rims on olivine crystals brought up from the mantle (Demouchy *et al*., 2006). Although olivine does not normally contain water, under high mantle pressures small amounts of hydrogen can enter the mineral and bond with oxygen to form hydroxyl. As magma rises toward the surface, the pressure decreases and the hydrogen starts diffusing out of the crystal. Knowing the rate of this diffusion allows the rate of magma ascent to be calculated from the shape of the dehydraion profile on the rim of the crystals (see Problem 5.17). Based on these profiles Demouchy *et al*. estimate that a basaltic magma in Pali-Aike, Chile, rose at a rate of ~6 m s^{-1}. Volatile-rich magma can rise even more rapidly. Based on the transport of large fragments of country rock (*xenoliths*) in kimberlites, velocities of up to 20 m s^{-1} have been calculated (Sparks *et al*., 2006). In volcanic explosions, such as the May 18, 1980, eruption of Mount St. Helens (Kieffer, 1981), magma velocities can even become supersonic (>344 m s^{-1}). In contrast, the rate at which granite diapirs rise is on the order of meters per year (Marsh and Kantha, 1978).

In considering the flow of magma, it is important to bear in mind that magmas, like other fluids, flow only in response to pressure gradients. Although these gradients are due largely to buoyancy, pressures can also result from volume increases on melting in the source region, liberation of gas from vapor-supersaturated magma, and from tectonic forces. We will consider each of these in turn.

3.2 BUOYANT RISE OF MAGMA

Within a body of magma surrounded by denser country rock, there is an excess pressure gradient (dP_{ex}/dz) due to the load of the surrounding rock, and it is this gradient that provides the buoyant force on magma. The density of the magma generates a pressure gradient in the magma itself, which we can refer to as the *magmastatic* pressure gradient ($dP_m/dz = \rho_m g$) to distinguish it from the surrounding *lithostatic* pressure gradient ($dP_r/dz = \rho_r g$), where ρ_m and ρ_r are the densities of magma and country rock respectively. Visualize two juxtaposed columns, one consisting of dense rock, and the other of less dense magma. At any given depth z, the pressure at the base of these two columns would be $P_r = \rho_r gz$ and $P_m = \rho_m gz$ respectively. Because $\rho_r > \rho_m$, the pressure at the base of the rock column would exceed that at the base of the magma column by an amount $(\rho_r - \rho_m)gz$. If the rock is able to flow, which we have argued is normally the case except very near the surface of the Earth (Section 1.3), the higher pressure at the base of the rock column relative to that in the adjoining magma column would cause the rock to flow into the magma and force it upward. If $\Delta\rho \equiv \rho_r - \rho_m$, then the excess pressure causing buoyant rise of the magma is $\Delta\rho gz$, and the excess pressure gradient in the magma due to the loading of the adjoining country rock is

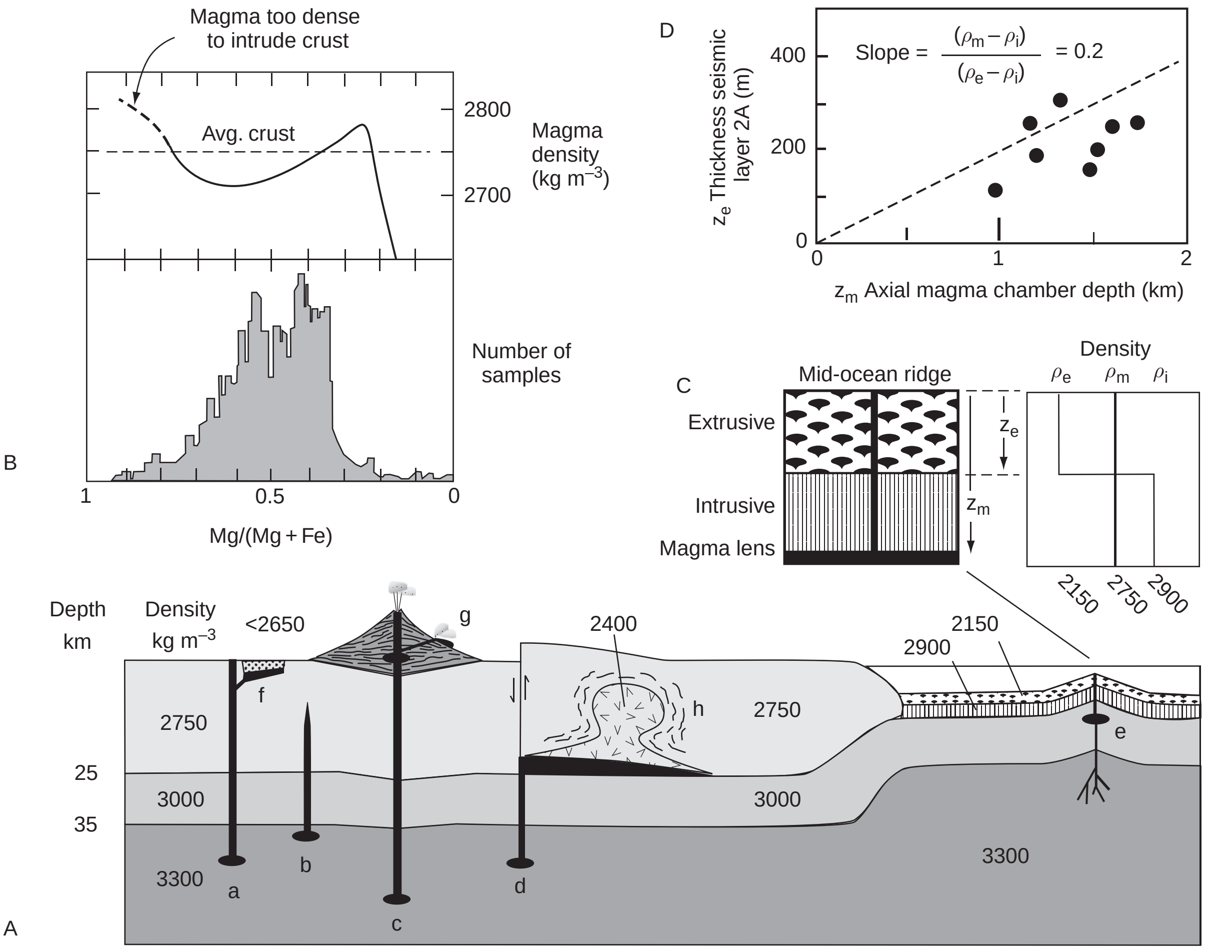


Fig. 3.1 (**A**) Section through continental and oceanic crust showing the general density distribution. Various magma source regions are shown as ***a*** through ***h***. Magma from source ***a*** has just sufficient buoyant force to rise to the surface. Similar density magma rising from a shallower source such as ***b*** would be able to rise only part way through the crust. For magma to rise to the top of a large volcano, its source would have to be deeper (***c***). Dense magma may intrude along the base of the upper crust (***d***), thus underplating the continental crust. Beneath oceanic ridges (***e***) magma may rise through less dense extrusive rocks if forced up by the pressure of dense intrusive rocks. On the other hand, magma encountering less dense material such as poorly cemented sediments (***f***) or porous volcanic rocks (***g***) may pond and spread laterally as a sill or dike. Granitic magma formed by partial melting of crustal rocks has low density and rises as large diapirs (***h***). (**B**) Stolper and Walker (1980) have shown that the density of melts derived by partial crystallization of basaltic magmas correlates with the molecular ratio of Mg/(Mg + Fe) in the melt, and that the abundance of continental basaltic rocks drops off sharply where the density of the melts exceeds that of the upper crust. (**C**) Buck *et al.* (1997) have shown that basaltic magma can rise through less dense extrusive rocks on mid-ocean ridges if it first rises through a sufficient thickness of denser intrusive rocks. (**D**) Consequently, the thickness of extrusive rocks shows a positive correlation with the depth of the axial magma chamber (see text for discussion).

$$\frac{dP_{ex}}{dz} = \Delta\rho g \tag{3.1}$$

Crustal rocks are generally denser than most magmas and consequently produce a buoyant force on magmas (Walker, 1989; Ryan, 1994). Typical continental crust consists of an upper 25 km with a density of ~2.75 Mg m^{-3} and a lower 10 km with a density of ~3.0 Mg m^{-3} (Fig. 3.1(A)). In sedimentary basins, unconsolidated and uncemented sediments can have densities significantly less than 2.65 Mg m^{-3}. The oceanic crust consists of an upper layer of extrusive basaltic rocks with densities between 2.0 and 2.8 Mg m^{-3}, depending on their degree of fracturing, vesiculation, and porosity. This layer is believed to correspond to layer 2A in seismic profiles of the oceanic crust. Beneath this layer are intrusive rocks with densities of 2.8 to 3.0 Mg m^{-3}. The density of the mantle immediately beneath the oceanic and continental crust is ~3.3 Mg m^{-3}. In contrast, most basaltic magmas have densities between 2.6 and 2.75 Mg m^{-3}, and rhyolitic magmas have densities of ~2.4 Mg m^{-3}.

Most basaltic magmas are able to rise buoyantly through the entire continental crust, except perhaps in sedimentary basins where the magma can spread laterally along the base of a low-density sedimentary fill (*f* in Fig. 3.1(A)). This appears to have occurred during the intrusion of the first basaltic magmas into the Mesozoic sedimentary basins of eastern North America, which formed during the initial opening of the Atlantic Ocean. The Palisades sill, for example, which forms the cliffs along the western bank of the Hudson River facing New York City, was formed by basaltic magma that intruded along, or near, the base of the sediments filling the Newark Basin. In the oceanic crust, basaltic magma can rise buoyantly through the dense lower part, but when it reaches the less dense extrusive part, it may pond or spread laterally (Ryan, 1993).

Buoyancy may be such an important driving force for the rise of magma that only magmas with densities less than that of the crust ever make it to the surface; denser magmas would presumably intrude downward. Stolper and Walker (1980) have shown that the densities of basaltic magmas vary systematically as a function of their Mg/(Mg + Fe) value, with a pronounced density minimum occurring between values of 0.8 and 0.3 (Fig. 3.1(B)). Samples having compositions in this low-density range are far more abundant than are samples falling outside the range. This led Stolper and Walker to conclude that the crust acts as a density filter that prevents dense magma from rising into the crust.

Dense magma could rise through a lower density crust if the average density of the entire rock column were greater than the density of the magma column. This appears to take place beneath mid-ocean ridges, where bright seismic reflectors are interpreted to image magma chambers that periodically are tapped to form extrusions of basaltic lava on the ocean floor (***e*** in Fig. 3.1(A)). However, this magma must rise through a crust that in its upper part consists of low-density extrusive rocks. Buck *et al*. (1997) have argued that the magma rises buoyantly because it is forced up by a column of rock in which the low-density extrusive rocks are compensated for by a sufficient thickness of higher density intrusive rocks beneath (Fig. 3.1(C)). For example, if a column of magma with density ρ_m were able to rise just to the ocean floor through dense intrusive rocks with density ρ_i and overlying extrusive rocks with density ρ_e, where $\rho_i > \rho_m > \rho_e$, the pressures at the base of the magma (P_m) and rock (P_r) columns must be the same. If we let the thickness of the extrusive layer be z_e and the depth of the magma source beneath the ocean floor be z_m (Fig. 3.1(C)), and ignoring the pressure of water, which would be the same on both columns, then

$$P_r = \rho_e g\, z_e + \rho_i g(z_m - z_e) = P_m = \rho_m g\, z_m$$

from which it follows that

$$z_e = z_m(\rho_m - \rho_i)/(\rho_e - \rho_i) \qquad (3.2)$$

The thickness of the extrusive layer is proportional to the depth of the magma source, with the proportionality constant being the ratio of the density differences. Buck *et al*. (1997) give typical values for ρ_i, ρ_m, and ρ_e of 2.9, 2.75, and 2.15 Mg m^{-3} respectively, in which case

$$z_e = 0.2 z_m$$

They show that in seismic profiles, the thickness of layer 2A, which is interpreted to be a layer of extrusive rocks overlying intrusive rocks, does vary as a function of the depth to the axial magma chamber, and that the slope on this linear relation is close to 0.2 (Fig. 3.1(D)). However, the thickness of most extrusive layers is slightly less than predicted from this relation. This may be due to the extrusive rocks increasing their density at depth due to closing or filling of fractures with hydrothermal minerals. Inspection of Eq. (3.2) shows that increasing the value of ρ_e would decrease the thickness of the extrusive rocks (Problem 3.1).

We could, in a similar manner, argue that dense magmas could rise through less dense continental crust if the source were deep enough in the dense mantle to provide the pressure necessary for the magma to rise through the crust. Let us assume that a fracture is filled with magma that can rise buoyantly and unhindered from its source depth (s km) in the mantle just to the Earth's surface (***a*** in Fig. 3.1(A)). This is an oversimplification, as will be discussed below, but it does allow us to appreciate how the magma source depth can affect the intrusion of magma. If the continental crust has the density structure shown in Figure 3.1(A), and the density of the magma is 2.9 Mg m^{-3} (an exceptionally dense basaltic magma), then the pressures at the base of the rock and magma columns ($\rho g z$) must be equal; that is,

$$2750 \times g \times 25\,000 + 3000 \times g \times 10\,000 + 3300 \times g(s - 35) \times 10^3 = \int_{z=0}^{z=s} \rho_m g \mathrm{d}z$$

If we assume the magma is incompressible, and g remains constant over the depth considered, then $s = 41.9$ km; that is, a magma starting at this depth would derive sufficient pressure from the rock column to intrude through the low-density crustal rocks.

Rocks are not significantly compressible ($\beta \approx 10^{-12}$), but magmas, if exposed to sufficient pressure do compress ($\beta \approx 5 \times 10^{-11}$ Pa^{-1}) and hence become denser. If we are to take into account the compressibility of magma in calculating buoyancy effects, we can integrate Eq. (2.8) to give

$$\beta(P_z - P_0) = \ln \frac{V_z}{V_0} \qquad (3.3)$$

P_0 is negligible compared with P_z, so Eq. (3.3) can be rewritten as

$$\frac{V_z}{V_0} = \exp(-\beta P_z)$$

and since $V_z/V_0 = \rho_0/\rho_z$, where ρ_0 and ρ_z are the densities of the magma at the surface and depth z, respectively, we obtain

$$\rho_z = \rho_0 \exp(\beta P_z) \qquad (3.4)$$

But the pressure at depth z is obtained by integrating Eq. (1.1),

$$\int \mathrm{d}P = \int \rho_z g \mathrm{d}z$$

The expression for ρ_z from Eq. (3.4) can be substituted in the right-hand side of this equation, to give, after separating variables

$$\int \exp(-\beta P)\mathrm{d}P = \rho_0 g \int \mathrm{d}z$$

which, upon integration between the limits of the surface of the Earth, where $P = 1$ atm, and a depth s, where $P = P_s$, yields

$$\exp(-\beta P_s) = \exp(-\beta) - \beta\rho_0 gs$$

Thus, the pressure at depth s in a column of magma with a coefficient of compressibility of β and density at the Earth's surface of ρ_0 is given by

$$P_s = -\frac{1}{\beta}\ln(\mathrm{e}^{-\beta} - \beta\rho_0 gs)$$

Because β is very small, $\mathrm{e}^{-\beta}$ is nearly 1, and thus

$$P_s = -\frac{1}{\beta}\ln(1 - \beta\rho_0 gs) \tag{3.5}$$

This pressure can then be equated with the pressure at the base of the adjoining rock column, thus allowing us to solve for s, the depth to the magma source region. The effect of compressibility is to increase the average density of the magma column, which in turn makes the source region deeper than it would be if compressibility were ignored (see Problem 3.2).

Let us return to considering the depth of the source of magmas shown in Figure 3.1(A) that rise in response to buoyancy. If the magma that rose from source ***a*** was able just to reach the surface of the Earth, a similar density magma rising from a shallower depth in the mantle, for example reservoir ***b***, would have insufficient pressure to rise to the surface. Conversely, a magma that was able to build a large volcano on the Earth's surface would have to have a deeper source (***c*** in Fig. 3.1(A)). If magma rising from the mantle is denser than the upper crust, it may intrude along the base of the upper crust, essentially floating the crust on the dense magma (***d*** in Fig. 3.1(A)). This process of intrusion, which is commonly referred to as *underplating* (Herzberg *et al*., 1983), may have been responsible for developing the lower part of the continental crust throughout geologic time. The cooling and crystallization of the underplating magma may supply the heat necessary to fuse the base of the upper crust and form granitic magmas and cause metamorphism. Granitic magmas have much lower densities (~2.4 Mg m^{-3}) than do high-grade metamorphic rocks (~2.8 Mg m^{-3}) and as a result are buoyant. However, because they also have extremely high viscosities (Section 2.4, Table 2.2) and the surrounding rocks behave plastically, they typically rise as diapirs rather than as dikes (***h*** in Fig. 3.1(A); see also Section 3.11).

Although the balancing of pressures at the base of juxtaposed columns of magma and rock provides a simple way of visualizing the buoyant forces acting on magma, the model is not realistic. Its main deficiency is that it is static. For magma to rise into the crust it must flow, and flow involves forces other than those experienced by a static column of magma (Lister and Kerr, 1991). First, the magma must have sufficient excess pressure to propagate a fracture through the lithosphere. This involves creating new fracture surface at the fracture's tip (ΔP_{f}), deforming the surrounding rock elastically (ΔP), and overcoming any tectonic stress (ΔP). In addition, the flow of magma is resisted by the viscous drag of the magma against the walls of the dike thus generating a pressure drop (ΔP_{v}). The buoyant pressure ($\Delta P_{\mathrm{h}} = \Delta\rho g$) must therefore balance or exceed the sum of these other pressures if a magma is going to rise.

The relation between these various pressures and their effect on the intrusion and shape of dikes will be discussed in Section 4.7. Lister and Kerr (1991) argue that for fractures of significant length, the forces around the tip of the fracture and the elastic deformation of the host rock become of secondary importance, and the rise of magma is controlled dominantly by a balance between the buoyancy forces and the viscous pressure drop.

In addition, they point out that the walls of a dike are not rigid and therefore the buoyant force on a magma is caused not by the density difference between the magma and the entire rock column but only by the local density difference. In calculating pressures in the Earth (Section 1.3), we argued that at depth, rocks have little strength and therefore lithostatic pressures can be calculated the same way that we calculate hydrostatic pressures in water. Thus, if pressure differences exist across a dike wall, they would either force the wall outward or have it crack inward. The importance of this local pressure balance is that magmas rise in the Earth until they reach a *level of neutral buoyancy* (LNB). On reaching the LNB, magmas tend to spread laterally, either as vertical dikes or horizontal sills, depending on the local stress conditions (Lister and Kerr, 1991; Ryan, 1994; Rubin, 1995). In Figure 3.1(A) several examples of a LNB are given. Magma rising from source ***a***, on intersecting the base of the sedimentary basin filled with low density rocks, spreads laterally to form a sill, such as the Palisades sill in the Newark Basin of New Jersey. Similarly, the magma rising from source ***d*** became neutrally buoyant at the base of the upper crust, which it underplated. In large volcanoes (***g*** in Fig. 3.1(A)), magma chambers may form at the LNB caused by a capping of lower density extrusive rocks. Magma in these chambers may spread laterally into rift zones rather than rising and erupting from the top of the volcano (see Fig. 4.3).

The buoyant rise of basaltic magmas presents a serious problem on the Moon. Mare volcanism has covered the large impact craters on the Earth-facing side of the Moon with basaltic rocks that had liquid densities of between 2900 and 3100 kg m^{-3}. The crust of the Moon, however, is made of a rock consisting largely of anorthite-rich plagioclase feldspar (anorthosite) with a density of 2890 kg m^{-3}, which is less than that of the basaltic magmas. Although a denser lower crust could have added the pressure necessary to bring magmas to the surface, as long as the source were deep enough (Wieczorek *et al*., 2001), the lunar rocks would have behaved much like those on Earth and magma is likely to have risen to a level of neutral buoyancy. The LNB on the Moon, which is at a depth of approximately 60 km, is where most basaltic magmas probably spread laterally (Head and Wilson, 1992). Wieczorek *et al*. (2001), however, believe that most basaltic melts on the Moon had densities that were slightly less than a lower crust, which in addition to the plagioclase, also contains orthorhombic pyroxene to form a rock known as norite, which has a density of ~3100 kg m^{-1}. They suggest that basaltic liquids were able to rise through the crust where the upper low-density crust had been removed by giant impacts.

In conclusion, because buoyancy on a local scale is the main force causing magmas to rise toward the surface, the long-term consequence of magmatic activity has been to create density-stratified planets.

3.3 VOLUME EXPANSION ON MELTING

Although buoyancy is the main force causing magmas to rise, in the region of magma generation, the increase in volume that accompanies the transformation from solid to liquid is capable of producing pressure gradients (Roberts, 1970). Most silicates expand 4 to 15% on melting, with 10% being a common value for rocks (see Section 2.3). The formation of the first small amounts of melt in a source region, consequently, results in a volume expansion. But the strength of surrounding rock opposes this expansion and causes pressures on the liquid to rise. This, in turn, raises the melting point, and no further melting occurs until the temperature increases. When this does occur, the formation of new liquid results in further pressure increases. Of course, if the process were slow, the surrounding rock would yield by flow, and the excess pressure would be relieved. At faster rates, however, the pressure within the liquid builds until the tensile strength of the surrounding rock is exceeded, at which point brittle failure occurs. The liquid expands into the fracture and, by so doing, decreases its own pressure. This, in turn, allows for more melting and expansion, which causes further fracturing and intrusion, the process continuing as long as the source region remains above the melting temperature.

Two questions are immediately raised by this process. First, how high a pressure could be generated within an embryonic volume of liquid, and second, how fast would melting have to occur to produce these pressures? The answer to the first question is not simple, because rocks under high confining pressures behave as strong elastic solids when exposed to short-duration stresses, but as rather weak, highly viscous materials when exposed to long-duration stresses. Consequently, the amount of pressure that might develop within a volume of rock during melting would be rate dependent. If the mechanism is to be capable of providing pressures for intrusion over extended periods, rocks must have sufficiently high long-term strength.

Energy released from deep-focus earthquakes and the magnitude of regional gravity anomalies both indicate that the long-term strength of rocks in the upper mantle is about 0.05 GPa. Anything greater than this results in fracture or flow. Because the excess pressure that can be sustained within a spherical volume of liquid by surrounding solid rock under lithostatic pressure is two-thirds of the tensile strength of the rock (Roberts, 1970), values greater than 0.03 GPa in the liquid would be sufficient to cause brittle failure of the surroundings.

Because magmas are rather incompressible, only small amounts of partial melting are necessary to cause significant increases in pressure. For example, if the volume increase on melting is 10%, only 0.24% of a rock need melt to produce the excess pressure of 0.03 GPa necessary to cause brittle failure. This fracturing due to the excess fluid pressure is commonly referred to as *hydraulic fracturing*, by analogy with the common engineering practice of fracturing rock in water and oil reservoirs by injecting water into the reservoirs under high pressure. Dikes in mantle peridotite have been attributed to hydraulic fracturing resulting from partial melting (Nicolas and Jackson, 1982; Maaløe, 2003). Partial melting is expected to occur in upwelling mantle due to the decrease in lithostatic pressure. If this is accompanied by hydraulic fracturing, a feedback process may be initiated that would promote melting. The fracturing releases pressure, which causes more melting, which decreases the bulk density of the rock and promotes buoyant rise; this in turn promotes melting and so on. The process would be self-limiting only because of the large amount of heat that is required to fuse rock (the latent heat of fusion of rock is large; see Chapter 5). Because the excess pressure is due to the volume expansion on melting, these pressures are not likely to be transmitted far from their source, and the liquid must remain connected to that source.

One example where the volume expansion on melting has definitely resulted in the flow of magma is at the contacts of basaltic intrusions where heating of wall rock has resulted in the generation of granitic liquids that intrude back into the basaltic intrusion. For example, the Higganum dike is a 50-m-wide basaltic dike that traverses all of New England from Long Island Sound in the south to New Brunswick, Canada, in the north (see Fig. 4.49). The wall rocks along this dike are commonly partially melted out to a distance of 1 m from the contact (Philpotts and Asher, 1993). Zones of granitic liquid developed between grains of quartz and feldspar in the wall rock, because these minerals provided the lowest melting mixtures, as will be discussed in Section 10.2. On crystallizing, this liquid formed a characteristic texture known as a granophyric intergrowth (Fig. 3.2(B)). As the melt developed at the quartz–feldspar grain boundary, it expanded and fractured the surrounding grains. As melting continued, the melt migrated into these fractures and caused them to propagate (Fig. 3.2(C)). At the same time, the margins of the basaltic dike were cooling and crystallizing, which caused a volume decrease in the dike. The expansion in the wall rock and the shrinkage in the dike resulted in a pressure differential that caused the granitic liquid to intrude into the dike's margins (Fig. 3.2(A)). The contact of the dike is dipping into the rock face at 45°. The veins of granite are therefore intruding down from the wall rock into the denser basalt, which would be difficult to explain if buoyancy were the force causing the granite to intrude. If the volume expansion on melting is the driving force for intrusion, gravity would have played a minor role.

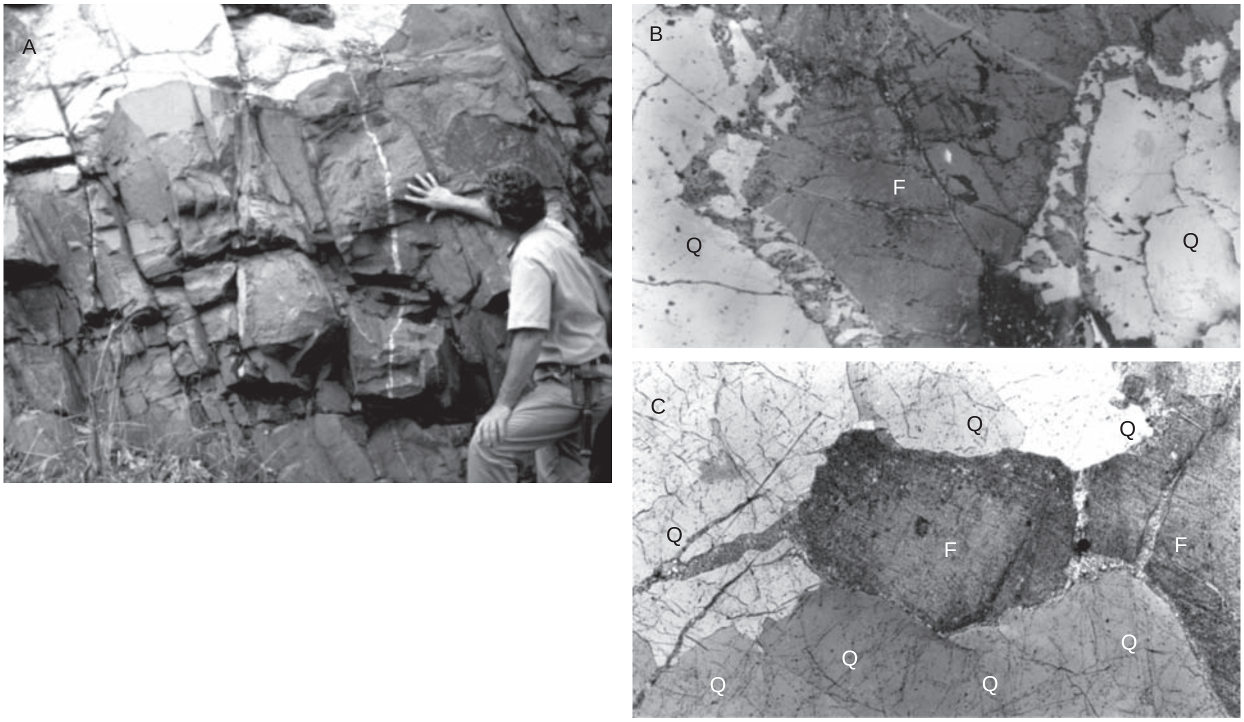


Fig. 3.2 (**A**) Contact of the 50-m-wide Higganum diabase dike on Route 9 Connecticut showing small dikes of granite that were formed by partial melting of the dike's wall rocks (upper left of photograph) intruding down into the diabase. (**B**) Photomicrograph of a polished thin section under reflected light showing zones of granophyric intergrowth of quartz and feldspar formed by partial melting between grains of quartz (Q) and andesine feldspar (F) in the wall rock near the contact with dike shown in (A). The polished surface was etched with HF fumes for 2 seconds to turn feldspar brown. Width of field 1 mm. (**C**) Photomicrograph of pegmatite in the wall rock near the contact of the diabase dike shown in (A) taken under partly crossed polarizers. The volume expansion on melting at the grain boundaries between quartz (Q) and orthoclase feldspar (F) fractured the surrounding grains. A small vein of granitic melt can be seen extending out from the central feldspar grain into the quartz at the left side of photograph, and a similar vein extends into the feldspar grain at the right. Width of field is 7 mm. (From Philpotts and Asher, 1993.)

3.4 VESICULATION

Explosive volcanic eruptions pose one of the most serious natural hazards. The supersonic speed with which huge volumes of magma and rock can be ejected from a volcano attests to the pressures that can be generated by volatiles escaping from magma (Sparks, 1978). Historic catastrophies document the power of these volcanic eruptions. Perhaps one of the most famous of these was the AD 79 eruption of Vesuvius, which buried the nearby Roman cities of Pompeii and Herculaneum. In 1815, the eruption of Tambora on the island of Sumbawa, Indonesia, killed 88 000 people. In 1875, ash and gas that erupted from Askja devastated the Icelandic economy and resulted in 20% of its inhabitants emigrating. In 1902, the eruption of Mount Pelée in Martinique killed all but a few of the 29 000 residents of St. Pierre. But the magnitude of these historic eruptions pales by comparison with eruptions in the geologic record. For example, 2 million years ago the first of three large eruptions in Yellowstone, Wyoming, spewed out more than 250 km^3 of magma to form the "yellow stone" volcanic ash. This compares with 3 km^3 erupted in 1815 from Tambora.

Prior to eruption, magma may contain considerable quantities of gas held in solution by the confining pressure, but once pressure is lowered, the gas separates and forms bubbles (*vesicles*), just as it does when the cap is removed from a bottle of carbonated beverage. The vesicular magma is much more buoyant than the original and can rise rapidly to greater heights.

Experimental studies indicate that the maximum amount of gas that can be held in solution by a magma depends on the composition of the magma and, more important, on the confining pressure (see Chapter 11). For example, at atmospheric pressure the amount of water that can be dissolved in rhyolitic and basaltic magmas at their typical magmatic temperatures is negligible. At 0.5 GPa, however, rhyolite can contain 10 wt% H_2O and basalt 8 wt% H_2O. This, of course, does not mean that natural magmas must contain this amount under these conditions; all degrees of saturation are possible, depending on the availability of water. But with the pressure decrease that accompanies the rise of magma toward the Earth's surface, even undersaturated magmas must eventually become saturated.

Only a small degree of supersaturation (<1 MPa) is necessary to cause bubbles to nucleate (Fig. 3.3(A)). These bubbles grow as the pressure decreases on the rising magma, or new bubbles nucleate if diffusion rates are too low to transport gas to existing bubbles. During bubble growth the bulk density of the magma (liquid plus gas) decreases, making the magma more buoyant. If the density of the liquid fraction is and

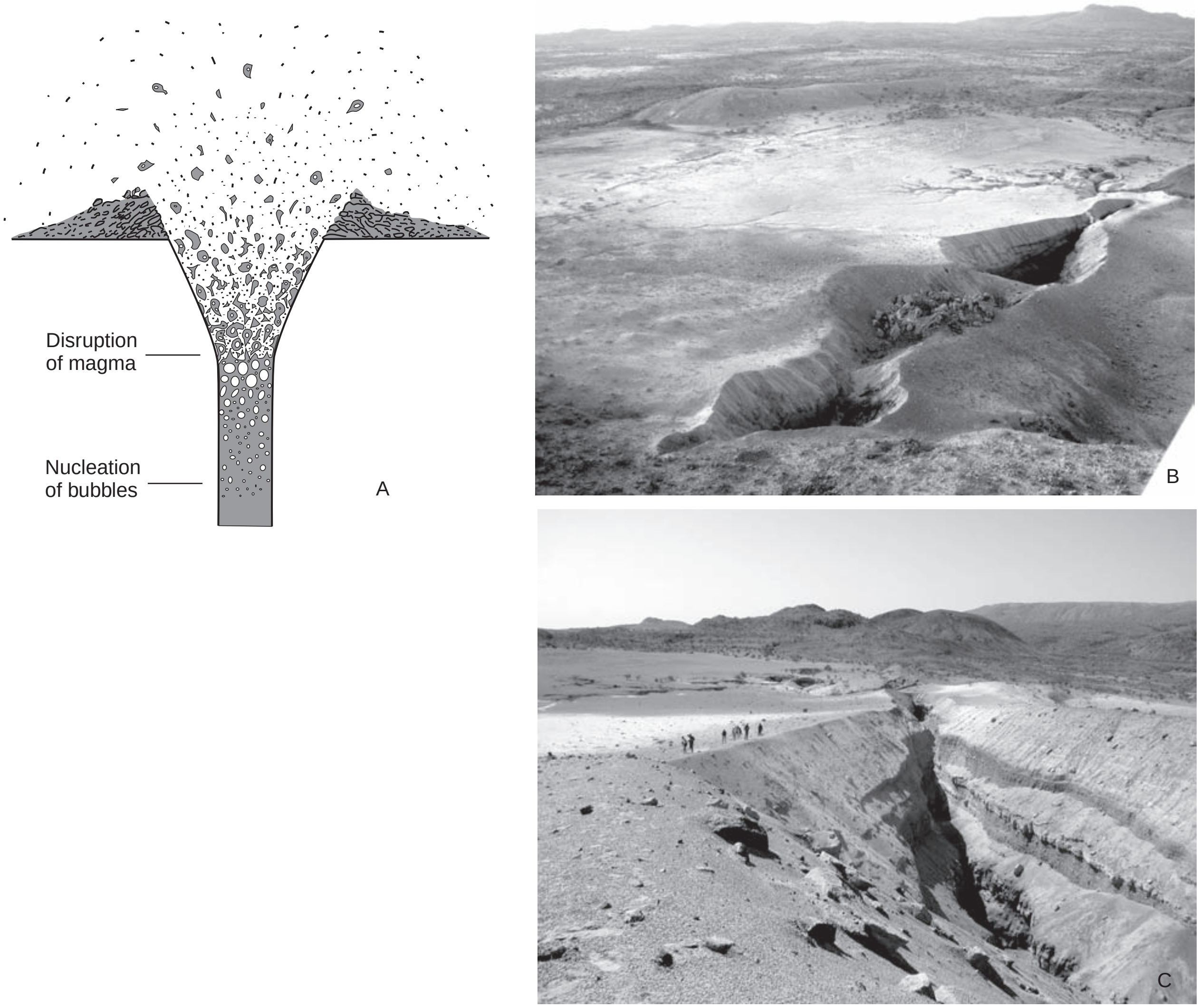


Fig. 3.3 (**A**) Magma rising in a volcanic feeder pipe will, at some depth depending on the initial volatile content, become supersaturated with volatiles (mainly H_2O and CO_2) and bubbles will nucleate. As magma continues to rise, bubbles grow larger through continued exsolution of volatiles and through expansion due to decompression. Bubbles eventually touch and disrupt the magma into frothy particles that are ejected from the vent at high velocities (see Fig. 3.4). (**B**) A 500-m-long, 60-m-deep, north–south trending vent formed by explosive eruption of rhyolitic magma near the Dabbahu volcano in the Afar region of Ethiopia on September 26, 2005 (see Fig. 15.19 for location). The photograph was taken looking toward the southeast by Asfawossen Asrat on October 7, 2005. The volcanic ash and blocks of rock ripped from the walls of the vent litter the surroundings. The eruption ended after several days with the emplacement of a small rhyolitic pumice dome near the center of the vent. (**C**) View looking south from the rhyolite dome on October 16, 2005, with people for scale (photograph by A. R. Philpotts).

that of the gas fraction is ρ_g, and the weight fraction of gas is f, the bulk density of the magma, ρ_b, is given by

$$\frac{1}{\rho_b} = \frac{f}{\rho_g} + \frac{1-f}{\rho_l} \tag{3.6}$$

The approximate density of the gas can be determined from the ideal gas law ($PV = nRT$)

$$\rho_g = \frac{PM}{RT} \tag{3.7}$$

where P is the pressure, M the molecular weight of the gas, R the gas constant, and T the absolute temperature (see Problem 3.5).

While exsolution of gas is taking place, the velocity of the magma increases (Fig. 3.4), in part because of increased buoyancy, but also because of the expansion of bubbles resulting from both decompression of the gas and diffusion of gas out of the melt into the bubbles. Once the bubbles have grown to the stage that they constitute approximately 64% of the volume of the rising magma (Gardner *et al.*, 1996), continued growth becomes difficult because the viscous silicate melt that must be displaced has only the thin tortuous sheets of liquid between the bubbles through which to move. Continued exsolution of gas therefore causes pressure within the bubbles to rise above the ambient pressure. Toward the free upper surface of the magma column this excess pressure causes the larger bubbles to burst. This disrupts the magma and breaks it into frothy particles that are ejected into the atmosphere (Figs. 3.3, 4.37). The term *pyroclastic* (broken by fire) ejecta is used for this material.

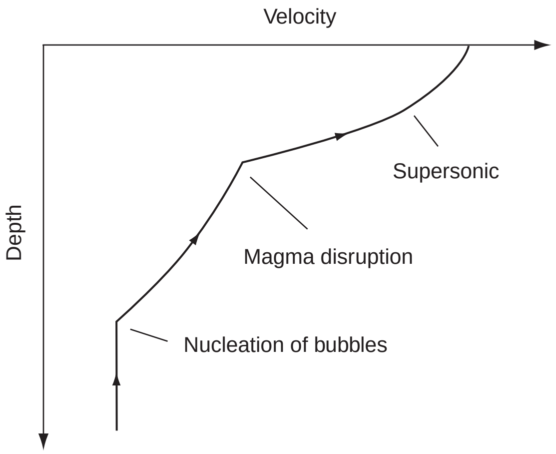


Fig. 3.4 Typical acceleration of vesiculating magma as it rises in a vent. The velocity below the depth at which gas bubbles nucleate (Fig. 3.3) remains constant as long as the conduit diameter is constant. The growth and expansion of gas bubbles cause magma to accelerate as it approaches the surface. Once the magma is disrupted by the gas bubbles, the velocity increases substantially (possibly becoming supersonic) because the low viscosity gas now forms the continuous medium (after Wilson and Head, 1981).

Up to the time of disruption, magma consists of a continuous silicate liquid enclosing bubbles of gas. After disruption, the gas forms the continuous phase, with globules of silicate melt suspended in the gas. The viscosity of the rising column of magma consequently decreases dramatically following disruption, and the flow velocities can increase to supersonic speeds in the vent (Fig. 3.4). Of course, the gas is what actually moves with this velocity, but it carries, in suspension, the smaller pyroclastic particles (<2 cm). Larger particles will have lower velocities depending on their terminal velocity of fall in the rising stream of gas.

The violence of the disruption is determined in part by the viscosity of the original magma (Gardner *et al*., 1996). The low viscosity of basaltic magmas allows gas bubbles to coalesce and rise to the top of a column of basaltic magma during periods of quiescence or slow flow and form large bubbles (see Fig. 4.14) that burst and throw out, for short distances, relatively large blobs or *bombs* of *scoriaceous* (vesicular) basalt (see Fig. 4.21). The magma beneath, meanwhile, has degassed itself and will erupt as a peaceful flow during the next period of activity. This alternation between periods of splatter and quiet flow is known as *Strombolian* activity. In contrast, the high viscosity of rhyolitic magmas impedes the separation of gas bubbles, so that during eruption the bursting of large numbers of evenly spaced small bubbles brings about a more thorough disruption of the magma into many small pieces of *pumice*, which can be ejected violently high into the Earth's atmosphere – the *plinian* type of eruption (Fig. 3.3). Even volcanoes erupting high-viscosity magmas can alternate between explosive and effusive activity (Melnik and Sparks, 2005; Adams *et al*., 2006).

The initial volatile content of a magma also plays an important role in determining the degree to which disruption will affect the magma. In general, basaltic magmas have much lower volatile contents than do rhyolitic ones, and thus basalts reach the disruption stage only at very shallow depths (see Problem 3.5). For example, typical Hawaiian basaltic magmas are estimated to contain about 0.45 wt% H_2O, in which case nucleation of gas bubbles does not even begin until the magma is within 300 m of the Earth's surface. Even so, this amount of gas is sufficient to disrupt the magma at the surface and produce "fountains of fire" over 500 m high. Nonetheless, the general lack of vesicles in dikes beneath eroded Hawaiian volcanoes and other terranes indicates that vesiculation cannot play a role in bringing basaltic magmas up from depth. Rhyolitic magmas, with their higher concentration of volatiles, may be affected by this process, at least at shallow crustal depths.

3.5 TECTONIC PRESSURE ON MAGMA

Our familiarity with squeezing toothpaste from a tube makes the concept of igneous intrusion by tectonic squeezing of a magma chamber a simple one to appreciate. As a result, it is readily invoked, despite the common lack of any supporting evidence. In the case of a squeezed toothpaste tube, flow is caused by a strong pressure gradient resulting from the tube being open at one end. If magma does not have easy access to some lower-pressure region, the magma chamber would simply be deformed by tectonic forces, and there would be no actual intrusion of magma. In the deeper crust and upper mantle, where most magmas originate, solid rocks react to long-term stresses as if they were extremely viscous liquids; under such conditions, there seems little likelihood of tectonic forces causing intrusion. But if movements were sufficiently rapid to cause rupturing of the crust, or if some other mechanism allowed magma to extend itself over a considerable vertical distance, tectonic forces could cause magma to flow toward the surface.

Tectonic stresses, however, play an important role in controlling the orientation of fractures that form when magma rises buoyantly into the lithosphere (Rubin, 1995; Gudmundsson, 2000). The parallel arrangement of multiple dikes to form a *dike swarm* perpendicular to the direction of tectonic transport in zones of crustal extension is perhaps the most striking example of this effect. The lower part of the oceanic crust consists, in large part, of such dikes that parallel the mid-ocean ridges (see Section 4.7). Dikes can also form radial swarms around igneous intrusions or on a still larger scale around hot spots (Ernst *et al*., 2001).

3.6 INTRUSION RATES OF NEWTONIAN MAGMA IN LAMINAR FLOW

The rate at which magma flows depends on the pressure gradient, magma viscosity, the shape of the conduit, and, if rocks are to be displaced by the intruding magma, the rate of deformation of the enclosing rocks. Many steeply dipping regional dikes of basaltic composition have intruded tensional fractures along which there may have been minimal displacement of the intruded rocks by the magma, whereas many large bodies of granitic magma have forcefully displaced the country rocks. Rates of intrusion of basaltic dikes are determined largely by the viscosity of the magma and the conduit shape, whereas the rate of deformation of surrounding rocks may determine the rate of intrusion of granite. Because rocks deform much more slowly than do liquids, rates of intrusion involving forceful displacement of country rocks are very low, ranging perhaps from meters to millimeters per year. Indeed, many bodies of granitic magma may even continue to rise slowly after they have completely solidified, as long as they are surrounded by denser rock (Ramberg, 1981). These slow rates contrast markedly with the rates at which low-viscosity basaltic magma can rise in fissures or volcanic pipes, unimpeded by any necessity to force aside large quantities of country rocks.

Because volcanoes erupt enormous volumes of lava in very short periods of time through small feeder pipes, magma velocities must be very high. For example, on August 21, 1963, over 7×10^5 m^3 of lava was extruded in a period of 12 hours from the Alae pit crater, on the east rift zone of Kilauea, Hawaii; the estimated maximum rate of extrusion was 11.5×10^4 $\mathrm{m}^3\ \mathrm{h}^{-1}$. The largest eruption in historic times, the 1783 fissure eruption at Lakagigar (Laki) in Iceland, is estimated to have poured out 18 $\mathrm{km}^3\ \mathrm{h}^{-1}$ of basalt from a few main fountains along a 10-km-long fissure. Volcanic feeder pipes and dikes that have been exposed by erosion may have diameters or widths of many hundreds of meters, but most are much smaller, and in many cases narrow with increasing depth. The width of dikes in mantle peridotites are ~0.1 m, in eroded Hawaiian volcanoes ~1 m, in the Scottish Tertiary ~4 m, and in Proterozoic dike swarms ~30 m (Rubin, 1995). It is necessary to conclude that magma velocities within feeder conduits can be rather high. Icelandic feeder dikes, for example, average only about 4 m in width, and the estimated average velocity within the dike that fed the 1783 Lakagigar eruption is calculated to be 6 $\mathrm{cm\ s}^{-1}$, but at times it probably surged to ten times this value (Thorarinsson, 1968).

Reasonable values of flow velocity in volcanic feeders can be calculated if the physical properties of the magma and the dimensions of the pipe are known. To illustrate this, we will consider steady-state laminar flow in a vertical, cylindrical pipe of radius r_0 (Fig. 3.5). The magma has a velocity υ in the z direction (upward), and we make the reasonable assumption that it is an incompressible, Newtonian liquid with density ρ_m and viscosity η.

We start by considering the forces acting in the z direction on a small cylindrical volume of magma with radius r and length dz. The force acting on the volume due to the pressure gradient in the z direction in the surrounding fluid is the difference between the pressure on the top ($P_{z+\mathrm{d}z}$) and the base (P_z) of the volume times the cross-sectional area; that is, $-\mathrm{d}P\pi r^2$. Gravitational attraction on the volume gives a negative force in the z direction of $-\rho_\mathrm{m}\pi r^2\mathrm{d}zg$. The shear stress acting on the outer surface of the cylinder, due to the upward movement of magma within the cylinder, gives rise to a force of $\tau\pi 2r\mathrm{d}z$. It is important to note that, because the flow is laminar, the shear stress, τ, and the pressure, P_z, are functions only of r and z, respectively. The steady-state flow of the magma means that there is no acceleration, and the sum of the forces in the z direction must be zero. We can, therefore, write

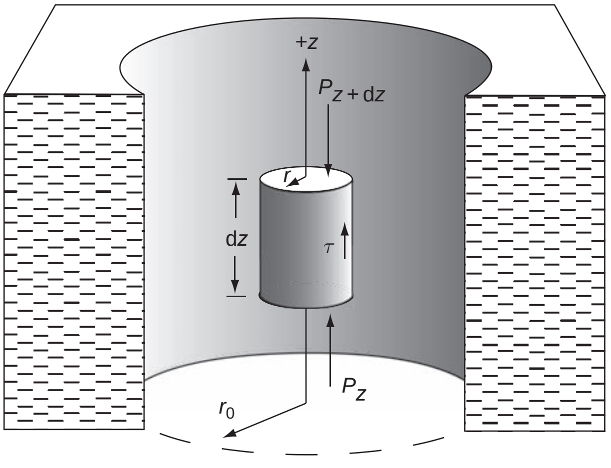


Fig. 3.5 A small cylindrical volume of magma with radius r and length dz within a volcanic feeder pipe of radius r_0. The steady-state laminar flow of magma within the cylinder results in a shear stress (τ) on the surface of the cylinder. The positive direction of z in this case is upward. See text for discussion.

$$-\mathrm{d}P\pi r^2 - \rho_\mathrm{m}\pi r^2\mathrm{d}zg + \tau\pi 2r\mathrm{d}z = 0$$

Substituting η dυ/dr for τ (Eq. (2.8)) and rearranging, gives

$$\frac{\mathrm{d}\upsilon}{\mathrm{d}r} = \frac{1}{2\eta}\left(\frac{\mathrm{d}P}{\mathrm{d}z} + \rho_\mathrm{m}g\right)r$$

which, upon integration, gives

$$\upsilon = \frac{1}{2\eta}\left(\frac{\mathrm{d}P}{\mathrm{d}z} + \rho_\mathrm{m}g\right)\frac{r^2}{2} + c$$

The integration constant can be evaluated, because $\upsilon = 0$ where $r = r_0$; hence,

$$\upsilon = -\frac{1}{4\eta}\left(\frac{\mathrm{d}P}{\mathrm{d}z} + \rho_\mathrm{m}g\right)\left(r_0^2 - r^2\right) \tag{3.8}$$

If the magma is rising solely due to buoyancy, the ratio dP/dz is simply the pressure gradient induced in the magma by the density of the intruded rocks. If the country rocks have a density ρ_c the pressure gradient, dP/dz, will simply be $-\rho_\mathrm{c}g$. Hence, the terms in the first set of parentheses in Eq. (3.8) reduce to $(g\rho_\mathrm{m} - g\rho_\mathrm{c})$. From this it follows that, if the densities of the magma and the intruded rock are the same, the flow velocity will be zero; if the density of the magma is less than that of the surrounding rock, upward flow will occur.

The group of terms in the first set of parentheses of Eq. (3.8) can also be considered to represent the gradient of pressure within the magma that is in excess of the magma-static pressure gradient ($\rho_\mathrm{m}g$); that is, the part of the gradient that can cause flow. If we indicate the gradient of excess pressure as dP /d , Eq. (3.8) can be reduced to

$$v = -\frac{\mathrm{d}P_{\mathrm{ex}}}{\mathrm{d}z\,4\eta}(r_0^2 - r^2) \tag{3.9}$$

Equation (3.9) indicates that the velocity profile across the volcanic pipe is parabolic, with the central part moving the most rapidly. In studying volcanic processes, however, knowledge of the average velocity, $\bar{v}$, is commonly more useful than the maximum velocity. This can be obtained by considering the volume of material that passes any cross-sectional area of the pipe in a unit time. This quantity is known as the *flux* (*J*) and is given by the *Hagen–Poiseuille law*,

$$\begin{aligned} J &= \int_0^{r_0} v 2\pi r\,\mathrm{d}r = \int_0^{r_0} -\frac{\mathrm{d}P_{\mathrm{ex}}}{\mathrm{d}z\,2\eta}(r r_0^2 - r^3)\pi\,\mathrm{d}r \\ &= -\frac{\pi r_0^4}{8\eta}\frac{\mathrm{d}P_{\mathrm{ex}}}{\mathrm{d}z} \end{aligned} \tag{3.10}$$

The average velocity is then

$$\bar{v} = \frac{J}{\text{area of pipe}} = \frac{J}{\pi r^2} = -\frac{r_0^2}{8\eta}\frac{\mathrm{d}P_{\mathrm{ex}}}{\mathrm{d}z} \tag{3.11}$$

which is one-half of the maximum velocity at the center of the pipe. Calculations of actual velocities are left to the reader to do in Problem 3.6 to 3.10. Note that the flux referred to in Equation (3.10) is the total flux through the pipe and hence has units of $\mathrm{m^3\ s^{-1}}$. Elsewhere in the book, the flux is commonly given as the flux passing through a square meter of cross-sectional area and hence has units of $\mathrm{m^3\ m^{-2}\ s^{-1}}$.

In this analysis of the flow of magma in a volcanic feeder pipe, the flow has been assumed to be laminar. Experiments indicate that when the value of $2r\rho\bar{v}/\eta$ exceeds 2300 (terms as defined above), a small initial disturbance grows and leads to turbulence, whereas below this value, it is damped out, and laminar flow prevails. This group of terms, known as the *Reynolds number*, and symbolized by Re, is a dimensionless number (see below) that gives a measure of the balance between inertial and viscous forces acting on the fluid.

$$\mathrm{Re} \equiv \frac{2r\rho\bar{v}}{\eta} \qquad \mathrm{m}\frac{\mathrm{kg}}{\mathrm{m^3}}\frac{\mathrm{m}}{\mathrm{s}}\frac{\mathrm{ms}}{\mathrm{kg}} \tag{3.12}$$

The value of 2300 is said to be the *critical Reynolds number* for flow in a cylindrical conduit. Problem 3.6 illustrates how this number can be used to determine when turbulent flow might be present in a magmatic pipe. If turbulence does occur, the analysis of feeder flow above would be invalid. But with the high viscosities of magmas, especially those rich in silica, laminar flow is likely to be more common than turbulent flow. Turbulent flow might, however, be expected for large radius conduits up which very fluid basalt rises (Shaw and Swanson, 1970), and komatiitic magmas with their extremely low viscosities (<0.3 Pa s) almost certainly would flow turbulently (Huppert and Sparks, 1985). Turbulent flow will be considered in Section 3.8.

In the previous treatment, magma in a volcanic feeder pipe was taken to be in a steady state of laminar flow. In reality, the pipe must have a beginning, possibly where it taps a magma chamber, and at this point the previous treatment is no longer valid. Magma entering the pipe does not immediately have a well-developed parabolic velocity profile. In fact, the magma will have to travel some distance along the pipe before attaining the steady-state velocity profile previously calculated. This distance, known as the *entrance length*, is of great importance in a wide range of fluid mechanics problems, including the flow of human blood through the aorta.

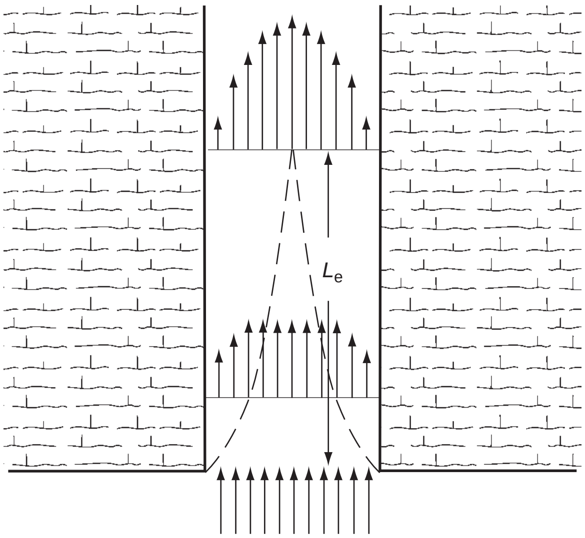


Fig. 3.6 Magma on first entering a feeder pipe from a magma chamber has the same velocity at all points across the pipe, indicated here by the lengths of the arrows. As the magma rises in the pipe, a boundary layer develops in which velocities are decreased by the viscous drag against the walls. Simultaneously, magma in the central part accelerates to maintain a constant flux of magma along the pipe. After a distance known as the entrance length, L_e, the boundary layer completely fills the pipe, and from thereon the velocity profile is parabolic.

The entrance length can be appreciated by referring to Figure 3.6, which illustrates a volcanic feeder pipe extending up from a magma chamber. At the instant magma enters the pipe, the velocity is constant across the pipe because there has not yet been any viscous drag along the walls. But as soon as magma begins traveling up the pipe, a boundary layer develops, due to the viscous drag on the pipe walls. The velocity of magma near the walls decreases, but because the average velocity must remain constant along the pipe (assuming constant radius), the magma in the center must simultaneously increase its velocity. This process continues until the boundary layer completely fills the pipe. Then, flow is said to be fully developed, and, from there on, the velocity profile remains constant.

The entrance length, symbolized by L_e is given by

$$L_e \equiv 0.1150(\mathrm{Re})r \tag{3.13}$$

where *r* is the radius of the pipe. Because the Reynolds number already includes the radius of the pipe, the entrance length varies as the square of the radius. The entrance length for a large-diameter pipe could therefore be considerable, and a parabolic velocity profile might not be achieved in the distance available (Problem 3.6d). This could be of importance in calculations regarding velocities of intrusion and to theories regarding the migration of crystals away from the walls of magma conduits (Chapter 10).

Another entrance length that is of importance in considering the intrusion of magma is known as the *thermal entrance length* (Lister and Kerr, 1991; Rubin 1995). As magma rises from depth into cooler rocks, heat is lost and eventually the magma solidifies. The distance that it can travel before it solidifies is the thermal entrance length (L_T). Calculating L_T is a complicated problem, because as the magma's temperature goes down, its viscosity increases exponentially. Also, as magma solidifies inward from the walls, its average velocity decreases in proportion to the square of the radius of the pipe (or square of the width of a dike). An order of magnitude for L_T can be obtained by multiplying the solidification time by the average velocity of the magma. In Chapter 5, we will see that the time required for a solidification front to advance a distance r is proportional to r^2 divided by the thermal diffusivity k (Eq. (5.15)); that is $t = \text{constant} \times r^2/k$. If we multiply this time by the average flow velocity (Eq. (3.11)), we obtain

$$L_T \propto \frac{r^4}{k\eta}\frac{dP_{ex}}{dz} \tag{3.14}$$

The same relation holds for a dike except that we replace the radius, r, with the dike's half width w (Fig. 3.7). Because the average velocity and the solidification time are both proportional to the square of the radius (or width for a dike), the entrance length is proportional to the fourth power of this dimension. Consequently, a small increase in the radius of a pipe or the width of a dike allows the magma to travel much farther before solidifying.

The expression for the velocity profile across a dike can be derived in the same way as that for a circular pipe. It will not be derived here, but the reader is asked to derive it in Problem 3.11. With luck, you should end up with Eq. (3.15). In the field, we would measure the width of a dike as the distance from one contact to the other. In calculations dealing with dikes, however, this distance is commonly designated $2w$ (Fig. 3.7); that is, w is the distance from the center of the dike to a contact (see, for example, Lister and Kerr (1991) and Rubin (1995)). By defining w in this way, it is analogous to the radius of a pipe. If we define x as the perpendicular distance from one contact of the dike toward the other, and z is positive upward and we deal only with vertical steady-state laminar flow (Fig. 3.7), the velocity as a function of x is given by

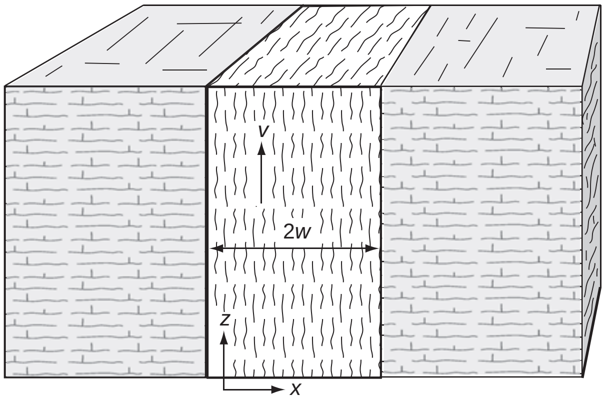


Fig. 3.7 Dike of width $2w$ up which magma flows with a velocity v in the vertical + direction.

$$\upsilon = -\frac{1}{2\eta}\left(\rho_m g + \frac{dP}{dz}\right)(2wx - x^2) \tag{3.15}$$

The similarity to pipe flow should be evident (see Eq. (3.9)). The velocity profile is again parabolic. The average velocity in a vertical dike is given by

$$\bar{\upsilon} = -\frac{1}{3\eta}\left(\rho_m g + \frac{dp}{dz}\right)w^2 \tag{3.16}$$

Equation 3.15 can be generalized for a sheet-like body with any angle of dip θ by multiplying the ρg term by $\sin\theta$. Thus, the term in the first set of parentheses becomes ($\rho g + dP/dz$) for a vertical dike and (dP/dz) for a horizontal sill. The reference direction, of course, would have to be changed so that in each case z is still in the direction of flow.

As with flow in pipes, too rapid flow of magma in a dike leads to turbulence, which may, in turn, affect such processes as crystal–liquid separation and cooling. The Reynolds number for flow in a dike is defined as

$$\text{Re} \equiv \frac{\rho 2w\bar{\upsilon}}{\eta} \tag{3.17}$$

where $\bar{\upsilon}$ is the average flow velocity (see Problem 3.17). Again, the critical value marking the onset of turbulence is 2300.

A Reynolds number can also be formulated for a lava flow, which is referred to as *open channel flow*. This type of flow experiences viscous drag only from the base and sides – the top of the flow is a free surface. This is taken into account by defining the Reynolds number as

$$\text{Re} \equiv \frac{4r_h\rho\bar{\upsilon}}{\eta} \tag{3.18}$$

where r_h the *hydraulic radius*, is the cross-sectional area of the channel divided by the wetted perimeter. For a lava flow that spreads over a great area, the thickness is very small compared with the width. The hydraulic radius thus reduces to the thickness of the flow, a dimension that is easily measured, even in poorly exposed terranes. When the Reynolds number is less than 500, flow is laminar in an open channel, and when greater than 2000, it is turbulent. Between these values is a transition, with the type of flow depending on the detailed nature of the channel bed (Shaw and Swanson, 1970). Perhaps the most important effect of the change from laminar to turbulent flow in a lava is the dramatically increased cooling rate. But because of the high viscosities of most lavas, turbulence is a rarity, except in very thick basaltic flows, komatiitic flows, and highly gas-charged ash flows (see Chapters 4 and 5). Where flows do move in a turbulent manner, their average velocity is given by

$$\bar{\upsilon} = \left(\frac{2g\sin\theta\, h}{k}\right)^{1/2} \tag{3.19}$$

where θ is the surface slope, h the thickness of the flow, and k is a friction coefficient, which has values between about 0.01 and 0.06, depending on the roughness of the surface.

In Section 3.2, we examined the role of buoyancy in the intrusion of magma. It was pointed out that the simplistic model of columns of magma balancing columns of juxtaposed rock (Fig. 3.1(A)) was misleading because it was static. For magma to intrude from a source region toward the surface of the Earth, it must create a fracture, force aside the country rock and overcome the viscous friction on the conduit walls. Having discussed the flow of magma in conduits, we are now in a position to say what extra force must be applied to a magma to overcome the viscous forces. For laminar flow we can rearrange Eqs (3.11) and (3.16) to obtain the excess pressure gradient needed to drive flow in a pipe or dike:

$$\text{Pipe flow: } \frac{\mathrm{d}P_{\mathrm{ex}}}{\mathrm{d}z} = -\frac{8\eta\bar{v}}{r^2} \quad \text{Dike flow: } \frac{\mathrm{d}P_{\mathrm{ex}}}{\mathrm{d}z} = -\frac{3\eta\bar{v}}{w^2} \tag{3.20}$$

From these equations, it can be seen that if the magma's viscosity increases, a greater pressure gradient is needed to force the magma along the conduit. For the dike to intrude into the rock it must also force aside the country rock. This involves the elastic properties of the rock, which we can define in terms of two parameters, the shear modulus and Poisson's ratio. The *shear modulus*, μ, is the ratio of the shear stress to the shear strain, and *Poisson's ratio*, v, is the ratio of the transverse to longitudinal strains produced by a stress. If we define the pressure in the magma that forces aside the rock (i.e. overcomes the elastic strength of the rock) as ΔP_{e}, then the dike will have the shape shown in Figure 3.8. If we define the dike's width as $2w$ and its breadth as $2b$, then

$$w = \Delta P_{\mathrm{e}} b \left(\frac{1-v}{\mu} \right) \tag{3.21}$$

For crustal rocks, the value of $(1 - v)/\mu$ would be ~0.05 GPa^{-1}, so Eq. (3.21) becomes on rearranging

$$\Delta P_{\mathrm{e}} = 20w/b\,\mathrm{GPa}$$

But the aspect ratio, w/b, for most dikes is between 10^{-2} and 10^{-4}. Consequently, the pressure required to open a dike is small compared with the pressure derived from buoyancy ($\Delta\rho gz$) and for dikes of great lateral extent (b is large), ΔP_{e} approaches zero. For this reason, the buoyant force on a magma is balanced primarily by the viscous drag as defined by Eq. (3.20). For a full discussion of this topic, see Lister and Kerr (1991) and Rubin (1995).

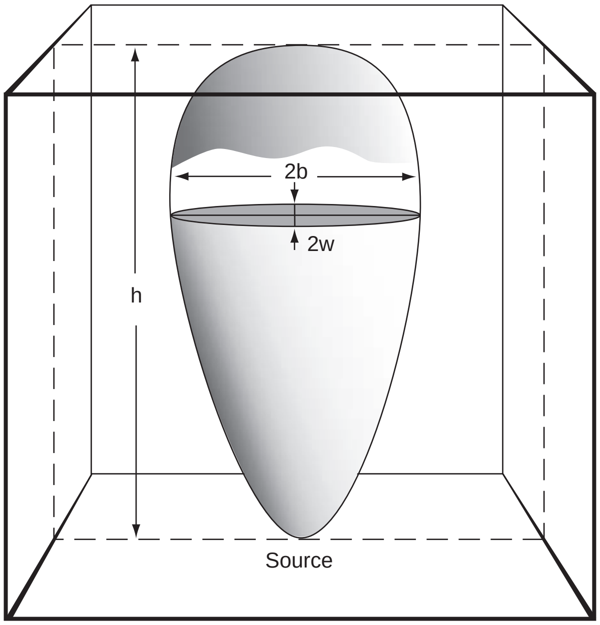


Fig. 3.8 Shape of a dike rising into the lithosphere. Its width (2*w*) has been exaggerated relative to its breadth (2*b*) and height (*h*). Typical ratio of *b* ^{2} ^{4} (after Lister and Kerr, 1991, and Rubin, 1995).

Before leaving the discussion of laminar flow rates through conduits, it is worth remarking on one important aspect that will be encountered again in later chapters. The flux of magma passing along a pipe is given by Eq. (3.10). It consists of two parts: the first group of terms expresses the ease with which the conduit allows material to pass, whereas the second part gives the gradient that provides the driving force for movement. All transport phenomena, whether involving fluid, heat, electricity, or diffusing ions, obey similar laws. For comparison with Eq. (3.10), here are three other transport laws that will be encountered later:

$$\text{Fluid flow} \quad J_x = -\frac{\pi r^4}{8\eta}\frac{\partial P}{\partial x} \quad \text{Hagen–Poiseuille law} \tag{3.10}$$

$$\text{Fluid flow} \quad J_x = -K_{\mathrm{H}}\frac{\partial h}{\partial x} \quad \text{Darcy's law} \tag{3.26}$$

$$\text{Heat flow} \quad J_x = -K_{\mathrm{T}}\frac{\partial T}{\partial x} \quad \text{Fourier's law} \tag{5.3}$$

$$\text{Diffusion} \quad J_x = -D\frac{\partial c}{\partial x} \quad \text{Fick's law} \tag{5.50}$$

In each of these equations, J_x is the flux or amount of the quantity transported per square meter per second in the x direction. The proportionality constants are K_{H}, the hydraulic conductivity, K_{T}, the thermal conductivity, and D, the diffusion coefficient. The gradients are expressed as partial derivatives of pressure, hydraulic or piezometric head (h), temperature, and concentration with respect to x, because only the gradients in the x direction cause transport in that direction. The hydraulic gradient in Darcy's law ($\partial h/\partial x$) is directly related to the fluid pressure gradient driving flow (Section 3.9). Similar partial derivatives can be written for the gradients in the y and z directions when the flux in those directions is considered.

3.7 FLOW RATE OF A BINGHAM MAGMA

In the derivations of flow rates through conduits, magma was assumed to be a Newtonian liquid. This was valid as long as the magma was largely liquid, but when it contains a significant number of crystals or gas bubbles, it can behave as a Bingham liquid; that is, it has a yield strength, which must be overcome before flow can occur. This type of behavior is expected if magmas continue to flow as they crystallize. Because this late-stage flow may affect the appearance of the final rock, it is important for us to examine how flows of Bingham and Newtonian liquids differ. We will do this for the case of a dike.

The shear stress, τ, at any point in a vertical dike with the dimensions and reference directions shown in Figure 3.7 is given by

$$\tau = -\left(\rho_{\mathrm{m}} g + \frac{\mathrm{d}P}{\mathrm{d}z}\right)(w - x) \tag{3.22}$$

From Figure 2.5 we can substitute $\tau_0 + \eta \mathrm{d}v/\mathrm{d}x$ for τ (τ_0 is the yield strength) as long as $\tau > \tau_0$ and $x < w$. This gives, on rearranging and integrating, an expression for the flow velocity of a Bingham liquid in a dike:

$$v = -\frac{1}{\eta}\left[\left(\rho_{\mathrm{m}} g + \frac{\mathrm{d}P}{\mathrm{d}z}\right)\left(wx - \frac{x^2}{2}\right) - \tau_0(w - |w - x|)\right] \tag{3.23}$$

which is valid as long as $\tau > \tau_0$. The shear stress in a flowing dike decreases from a maximum at the margins to zero at the center. At some distance in from the margins, the shear stress must drop below the yield strength of the Bingham liquid, and from there through to an equivalent point on the other side of the dike no additional flow is possible. The result is a central sheet that moves with constant velocity and produces *plug flow* (Fig. 3.9(A)).

Common examples of the flow profiles across Newtonian and Bingham liquids are provided by the shape of aloe vera gel and toothpaste when extruded from a tube. Aloe vera gel is a clear viscous Newtonian liquid, and when it is squeezed from a tube, it forms a parabolic velocity profile (Fig. 3.9(B)). Toothpaste, in contrast, is a Bingham liquid because of the fine abrasive particles that it contains in suspension (commonly ground-up feldspar), and as a result, it has a high yield strength and extrudes from a tube as a plug (Fig. 3.9(C)). A dike or volcanic feeder pipe containing large numbers of phenocrysts or vesicles would be expected to flow like the toothpaste; that is, the central part would flow as a plug and all of the shearing of the liquid would occur in the marginal zones. Vesicles, which make ideal strain markers, would be

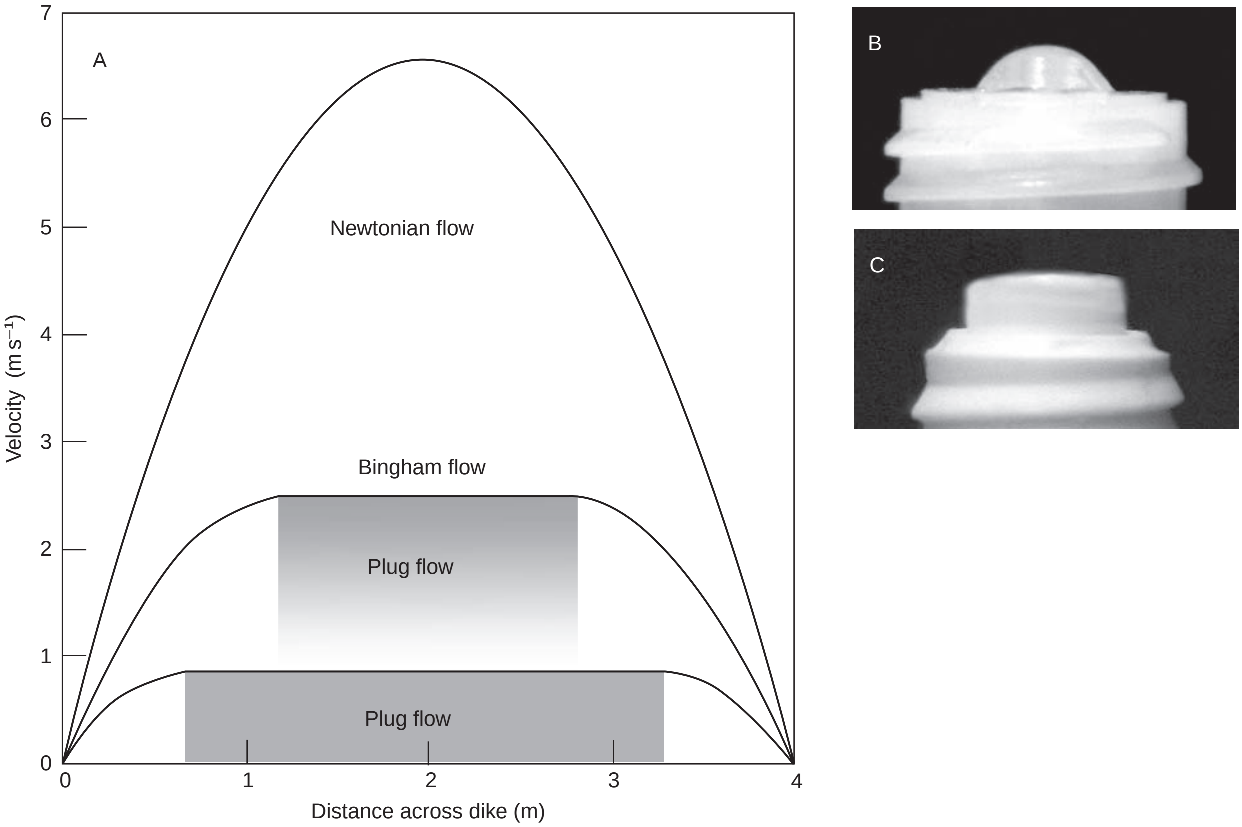


Fig. 3.9 (**A**) Laminar flow velocities in a 4-m-wide basaltic dike, in which the magma with a density of 2.6 Mg m^{-3} is buoyantly emplaced in response to a density contrast of 100 kg m^{-3}. The upper curve is for Newtonian behavior with a viscosity of 300 Pa s. The lower two curves are for Bingham behavior, where the viscosity is still 300 Pa s but the yield strength of the magma is 750 Pa (middle) and 1250 Pa (lower). When the magma behaves as a Bingham liquid, its velocity is diminished and a central portion (shaded) flows as a plug, the width of which increases as the yield strength of the magma increases. The maximum shear stress developed at the margin of this dike under its specified conditions is 1960 Pa. Should the yield strength of the magma exceed this value, no flow would be possible. (**B**) Newtonian flow of aloe vera gel squeezed from a tube. (**C**) Bingham plug flow of toothpaste squeezed from a tube.

spherical across the central part of the body and would be deformed into ellipses only near the contacts.

The same type of Bingham flow will occur in pipe-like conduits and lava flows, which will have a central or upper part, respectively, that flows with constant velocity. The fraction of a pipe, dike, or lava flow that moves as a plug depends on the yield strength of the magma; if it is high, most of the magma may flow as a single plug (see Problem 3.14). The yield strengths of magmas are poorly known, but they can vary over many orders of magnitude. For example, values as high as 10^6 Pa have been estimated for dacite flows on the surface of the Earth.

The change from Newtonian to Bingham behavior has a number of important consequences. Most obvious is that flow velocities are diminished because the yield strength of a magma provides an additional resistance to flow. Processes that depend on the shearing of liquid against a contact, such as alignment of phenocrysts or stretching of gas bubbles parallel to contacts, and movements of crystals away from conduit walls (flowage differentiation; Section 14.6) will not operate within the region of plug flow. Finally, the sinking of dense crystals through a less dense Bingham liquid will not take place if the crystals do not exert sufficient force to overcome the yield strength of the liquid (see Section 14.3).

3.8 INTRUSION RATES OF TURBULENT MAGMA

The discussions above have dealt only with laminar flow. In passing, however, it was mentioned that the very high Reynolds numbers for low-viscosity komatiites and rhyolitic ash flows and thick basaltic flows and dikes indicate that these magmas may have flowed turbulently. These are important types of magma, and we should therefore have some means of evaluating the rate at which they rise from their source or are erupted on the surface. Derivations of the equations of flow of turbulent liquids are not as simple as those for laminar flow, but the interested reader can find them in any of the standard texts on fluid mechanics (e.g. Tritton, 1977). Application of these equations, however, is simple.

If magma is assumed to rise buoyantly in a vertical dike of width *2w*, the average turbulent flow velocity, $\bar{v}$, is given by

$$\bar{v} = \left(\frac{g\,\Delta\rho\,2w}{k\,\rho_\mathrm{m}}\right)^{1/2} \tag{3.24}$$

where $\Delta\rho$ is the difference between the densities of country rock and magma, ρ_m is the density of the magma, and k is a friction coefficient, which has values between about 0.01 and 0.06, depending on the roughness of the wall (Huppert and Sparks, 1985). This relation is valid as long as the Reynolds number, which is defined as $\rho\bar{v}2w/\eta$, is greater than about 2000. The vertical flux of turbulent magma per unit length of dike, J, is simply the product of the average velocity and the width,

$$J = \left(\frac{g\,\Delta\rho}{k\,\rho_\mathrm{m}}\right)^{1/2}(2w)^{3/2} \tag{3.25}$$

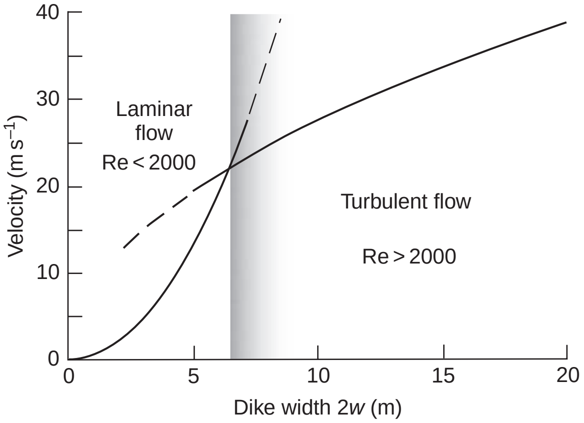


Fig. 3.10 Variation in average flow velocity in basaltic dikes as a function of width. The basaltic magma is assumed to have a viscosity of 300 Pa s and a density of 2.6 Mg m^{-3}, and to be buoyantly emplaced in response to a density contrast of 200 kg m^{-3}. Because of laminar flow, velocities in narrow dikes increase rapidly with increasing dike width. When widths are greater than about 8 m, the Reynolds number exceeds 2000 (assuming a friction factor of 0.01), and flow becomes turbulent. Velocities in wider dikes therefore do not increase as rapidly with increasing dike width.

The most significant difference between the expressions for the average velocity of turbulent and laminar flow of magma in a dike is the way in which dike width affects velocity; in turbulent flow, the velocity is proportional to the square root of the width, whereas in laminar flow it is proportional to the width squared. This is brought out in Figure 3.10, where the average velocity in vertical dikes of basaltic magma having $\eta = 300$ Pa s, $\rho_\mathrm{m} = 2.6$ Mg m^{-3}, and ρ_c (country rock) = 2.8 Mg m^{-3} is plotted as a function of dike width (2*w*). In dikes with a width up to about 8 m, Reynolds numbers are <2000 and flow is laminar; in wider dikes, flow is turbulent and then velocities increase only slightly with increasing width (Problem 3.17).

One of the most important consequences of turbulent flow is that temperatures are maintained approximately constant throughout the magma. Turbulence transfers heat rapidly to margins by convection, where it can heat country rocks or be radiated into space in the case of a lava flow. In laminar flow, heat is transferred by conduction, which is a slow process. The heat effects on the surroundings are consequently less pronounced, and cooling rates are very much slower. This point is discussed in Chapter 5.

3.9 FLOW THROUGH A POROUS MEDIUM

In our discussion so far, we have assumed that magma has risen from reservoirs at depth. These reservoirs, however, do not instantly come into existence but instead form slowly from melt that initially is a dispersed phase. Normally, the crust and upper mantle are solid, and only in special tectonic settings do conditions of temperature and pressure allow melting to occur. Because of the large heat of fusion of rocks, melting is only partial; that is, the amount of heat available is only sufficient to cause partial melting. Melting occurs where different minerals come in contact (Fig. 3.2(B)) and provide low melting mixtures (eutectic, Chapter 10). This liquid forms a complex network of tubes and films between the grains of the rock (Section 21.2). This liquid must migrate and segregate to form bodies of magma. To do so the liquid must flow through a *porous medium*.

The classic work describing porous flow is that of Henry Darcy (1856) the town engineer of Dijon, France (of mustard fame). He showed that the flux of water passing through a cylinder filled with sand from the River Saône, which flows near Dijon, was proportional to the pressure applied to the water. The proportionality constant, which is known as the *hydraulic conductivity,* K_{H}, depends on the permeability of the material, on the acceleration of gravity, and on the density and viscosity (η) of the fluid. Hydrologists usually write Darcy's law in terms of the hydraulic gradient ($\partial h/\partial z$) and the hydraulic conductivity

$$J_z = -K_{\mathrm{H}}\left(\frac{\partial h}{\partial z}\right)$$

where h is the hydraulic or piezometric head. However, this form is strictly valid only if the fluid has constant density, as is appropriate for very shallow groundwater flow modeling (Bear and Verruijt, 1987). For deeper, higher temperature systems, we cast Darcy's law directly in terms of pressure gradient. Written in this way, Darcy's law for flow of a fluid with density ρ in the vertical direction (z, increasing upward) is

$$J_z = -\frac{k}{\eta}\left(\frac{\partial P}{\partial z} + \rho \mathrm{g}\right) \tag{3.26}$$

where k is the permeability, $\partial P/\partial z$ is the pressure gradient applied to the fluid in the vertical direction, and ρg is the pressure gradient produced in the fluid by its own mass in Earth's gravity field (see Chapter 21). The negative sign is required because a fluid flows from high pressure to low pressure, that is, the gradient must be negative. The acceleration of gravity is actually a vector directed downward, but it is adequate to treat it as a constant here.

The Darcy flux given by Eq. (3.26) is the volume of fluid passing through a given cross-sectional area of porous material per time and hence has units of $\mathrm{m}^3{}_{(\mathrm{fluid})}\,\mathrm{m}^{-2}{}_{(\mathrm{area})}\,\mathrm{s}^{-1}$. It may be tempting to try and cancel these units to arrive at $\mathrm{m\ s}^{-1}$. However, it is important to emphasize that the flux does not give the actual velocity of the fluid moving through the individual pores, because the pores constitute only a fraction of the total cross-sectional area. The fraction of this area through which the fluid moves is known as the *porosity* (ϕ). The average velocity of the liquid flowing through these pores can be obtained by dividing the flux by the porosity; that is

$$\bar{\upsilon}_z = -\frac{k}{\phi\eta}\left(\frac{\partial P}{\partial z} + \rho \mathrm{g}\right) \tag{3.27}$$

If the fluid is rising due to buoyancy and the difference in density between the surrounding rock and the fluid is Δ , Eq. (3.27) can be rewritten as

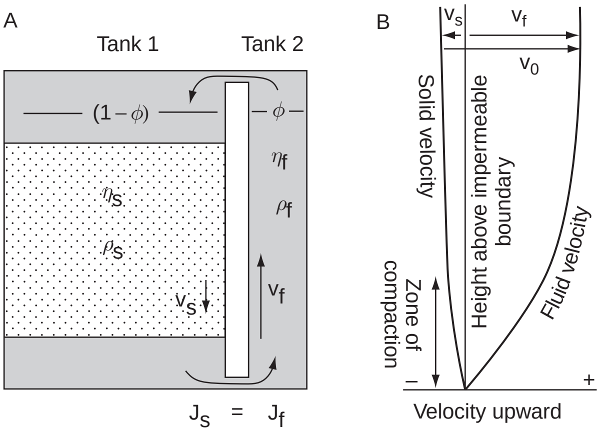


Fig. 3.11 (**A**) Simple model to illustrate the fluxes of solid and fluid during compaction. As a slug of solid material (stippled) sinks in one tank, an equal volume of fluid (shaded) rises in the other. The velocities in the two tanks depend inversely on their cross-sectional areas, 1 - ϕ and ϕ. By analogy, the downward flux of solids in a partially molten mantle with a porosity of ϕ expels upward an equal flux of liquid through its pores (see text for discussion). (**B**) Velocity profiles of fluid and solids through a pile of partially melted rock undergoing compaction against an impermeable lower boundary. Compaction occurs only in a narrow boundary layer at the base. Elsewhere the velocity of the solids and the fluid are almost constant (after McKenzie, 1984).

$$\bar{\upsilon}_z = \frac{k}{\phi\eta}\Delta\rho \mathrm{g} \tag{3.28}$$

However, the fluid can achieve this velocity only if the density contrast between the rock and the fluid can be converted into a pressure gradient. For this to happen, the rock has to be able to deform, and this is a slow process. The problem can be appreciated by considering the simple model shown in Figure 3.11(A), which shows two adjoining tanks that are connected by pipes. The larger tank is filled mainly with solid, which has a high density (ρ_{s}) and an extremely high viscosity (η_{s}); the smaller tank is filled with low-density (ρ_{f}), low-viscosity (η_{f}) fluid that is able to flow into and out of the larger tank via the connecting pipes. The cross-sectional areas of the smaller and larger tanks are ϕ and $(1-\phi)$ respectively. Although the solid and fluid tanks are separate in this model, in a zone of partial melting the fluid would actually be rising through pore spaces (ϕ) in the solids. However, by grouping together all of the pores and showing them as the single tank of rising fluid in Figure 3.11(A), it is easier to appreciate the balance that must exist between the solids and the fluid.

In the model, the solids in the larger tank are denser than the fluid in the smaller tank and so the system starts to convect in a counterclockwise direction. Because the two tanks are connected, the sinking of solids in the large tank is balanced by the influx of fluid from the smaller tank; that is, the flux in the two tanks must be equal. Since the tanks have different cross-sectional areas ($\phi \ll 1-\phi$), the fluid velocity (υ_{f}) is proportionately more than that of the solids (v_{s}). However, the solids flow very slowly because of their extremely high viscosity. Therefore, although the fluid moves faster than the solids because of the relatively low value of ϕ (porosity), the flow rate of the fluid is controlled by the velocity of the solids.

In a zone of partially melted rock, the buoyant rise of melt must be balanced by an equivalent downward flux of solids, but if the zone of melting has a floor, the solids must compact against the floor and expel the liquid upward. McKenzie (1984) has mathematically analyzed the porous flow of fluids expelled from zones of partially melted rock undergoing compaction. In the simple model shown in Figure 3.11(A), the solids simply formed a slug of highly viscous material that sank and displaced the fluid upward in the adjoining tank. In a zone of partially melted rock, however, the solids form a porous medium that sinks as the fluid migrates upward through its pores. If the zone of partial melting is floored by an impermeable boundary of solid rock containing no melt, as presumably happens with increasing depth in the Earth, the downward moving solids must compact against this boundary. Although the bottom of a partially melted zone would be gradational, analysis is simplified by considering it an impermeable plane. McKenzie shows that a zone of partially melted rock with initially homogeneous porosity resting on such a boundary will undergo compaction only in a thin basal boundary layer. At the boundary, the velocity of the solids and liquid must be zero, but above it, the solids sink toward the boundary causing compaction and expelling the liquid upward. The velocities of the solids and the liquid increase away from the boundary, but after a short distance, which McKenzie refers to as the *compaction length*, the velocities approach steady values (Fig. 3.11(B)). Above this level, the solids all sink with the same velocity, so there is no compaction in this part of the section. Also, in this zone, the upward velocity of the liquid relative to the solids (υ_0) is just sufficient to support the weight of the liquid and the solids. This relative velocity is given by

$$\upsilon_0 = \frac{k_\phi(1-\phi)\Delta\rho g}{\phi\eta} \tag{3.29}$$

The compaction length (δ_c) is

$$\delta_c = \left[\frac{(\zeta + \frac{4}{3}\mu)}{\eta} k_\phi\right]^{1/2} \tag{3.30}$$

where ζ and μ are the effective bulk and shear viscosities, respectively, of the solid matrix (both ~10^{15} Pa s), and η is the viscosity of the liquid. The effective bulk and shear viscosities of the matrix are decreased from what they would be if the porosity were zero by a factor of $(1-\phi)$. The permeability (k) is also a function of the porosity (ϕ). Note that the compaction length is independent of density. The velocities shown in Figure 3.11(B) are valid only during the initial stage of compaction. As compaction eliminates melt from the bottom of the pile, the impermeable boundary must move upward with time.

An analogy to the downward moving solids would be a line of evenly spaced vehicles moving at high speed on a highway toward an accident where the traffic was stopped. There would be no "compaction" of the vehicles until they neared the accident site, and then the distance between them would rapidly decrease and a pileup would result. A common example of the upward flux of fluid balancing the weight of the solids is provided by quicksand or the gliding puck in an air hockey game. Such systems are said to be *fluidized*. In quicksand, however, the particles become completely separated by the rising fluid and hence quicksand has low viscosity, whereas in partially melted rock with porosities of less than 20%, the solids remain interconnected and thus preserve a high viscosity.

McKenzie shows that the rate of compaction and expulsion of liquid in zones of partial melting are controlled by the properties of the solid matrix. This requires knowledge of the permeability and the bulk and shear viscosity of the matrix. Darcy's law provides only a macroscopic quantification of flow through a porous medium. It does not deal with the shape of the individual channels through the porous medium, and this shape determines, in a complex manner, how the permeability is affected by the porosity. For example, if the channels were all cylindrical tubes, then $k \propto \phi^2$ (Section 21.3), but if they were all planar cracks, $k \propto \phi^3$ (Guéguen and Palciauskas, 1994). Although analysis of flow through simple geometrical channels has provided useful insights, it falls far short of dealing with the complexities of real rocks. Experimental studies in which mantle rocks have been partially melted have provided valuable information on porosity, especially of how the wetting of crystals by the liquid affects the shape and connectivity of pores (e.g. Wark *et al*., 2003). Theoretical studies and computer simulations of complex flow patterns have gone a long way to bridging the gap between the macro- and microscopic scales of flow (van Genabeek and Rothman, 1996; Cheadle *et al*., 2004; Hersum *et al*., 2005). In addition, noninvasive methods involving the use of high-resolution X-ray computed tomography (CT) has allowed the channels through which flow takes place to be imaged. Just as CT scans are used in the medical field to image the human body, they have been used in petrology to image channels of granitic material formed during partial melting of metamorphic rocks (Brown *et al*., 1999) and channels of liquid in basaltic crystal mush (Philpotts *et al*. 1999). All of these different approaches are leading to a better understanding of the porosity of partially melted rocks.

We can conclude that porous flow in zones of partial melting is a slow process that is dependent on the rate of deformation of the solid matrix.

3.10 CHANNEL FLOW

Field, experimental, and theoretical studies reveal that in zones of partial melting, porous flow is soon replaced by channelized flow, especially if accompanied by deformation. Once channels develop, the flow velocity of liquid dramatically increases.

Normally the oceanic crust is subducted beneath continents, but on rare occasions it is thrust onto continents (obducted) to form a suite of rocks known as ophiolites (Section 15.2). These rocks provide a rare glimpse through the oceanic crust down into the underlying mantle where partial melting took place. Ophiolites have a typical stratigraphy from ocean-floor sediments at the top down through basaltic volcanic rocks, many of which form pillows

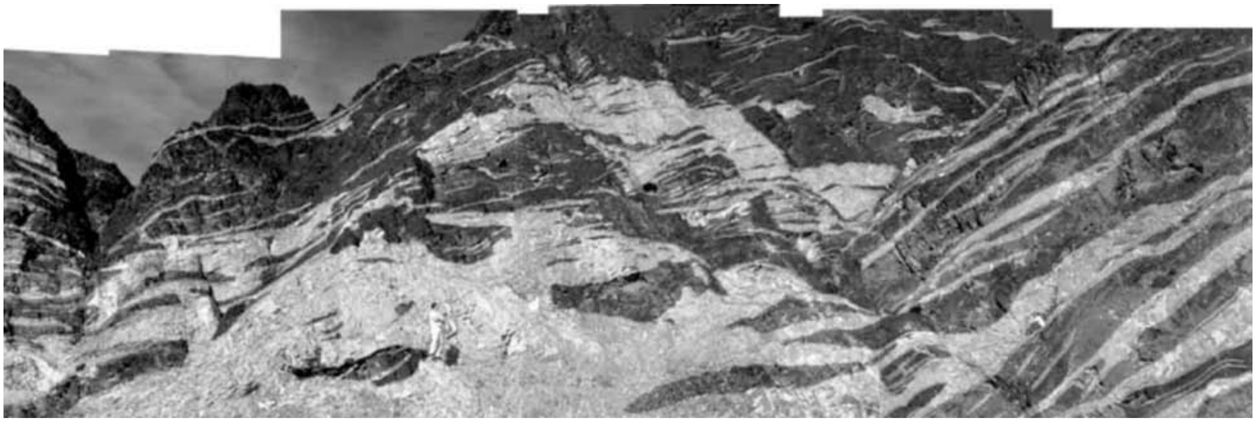


Fig. 3.12 Photomosaic showing light-colored veins of dunite cutting dark peridotite (harzburgite) in the Muscat Massif, Oman. The geologists to the left and right of center are 12.5 m apart. (Photograph from Braun and Keleman, 2002. Published with permission of the American Geophysical Union.)

indicative of extrusion in water, through dike rocks into underlying bodies of gabbro, and eventually into peridotites – rocks composed essentially of olivine and pyroxene. These latter rocks are believed to be part of the upper mantle that underwent partial fusion to produce the magmas that rose to form the overlying gabbroic and volcanic rocks. One of the most striking features of these mantle rocks is the abundance of olivine-rich veins ranging in width from ~1 cm to ~100 m (Fig. 3.12), which have been interpreted to have formed through channeled flow of melt leaving the source region.

The host rock for the veins is a variety of peridotite known as *harzburgite*, consisting predominantly of olivine and orthorhombic pyroxene. It is interpreted to be the refractory residue from which the basaltic liquids were extracted. The veins are composed almost entirely of olivine, which forms a rock known as *dunite*. Partial melting of peridotite produces a liquid that is saturated in both olivine and pyroxene, but when the liquid begins to rise, the decrease in pressure causes the liquid to become undersaturated in pyroxene (see Section 10.6, Fig. 10.8). The liquid then begins to dissolve pyroxene and precipitate olivine but it also increases the amount of liquid. This, in turn, increases the porosity, which increases the permeability. Kelemen *et al*. (1995, 1997) and Braun and Kelemen (2002) have proposed that this increased permeability leads to greater melt flux, which in turn leads to more solution of pyroxene, precipitation of olivine, and increased porosity and permeability. This positive feedback, which they refer to as a *reaction infiltration instability*, results in the formation of highly porous conduits bordered by olivine. These, then, are the dunite veins cutting the peridotite. Numerical modeling of this process shows that melt being expelled from partially melted peridotite undergoing compaction naturally organizes itself into channels (Spiegelman *et al*., 2001). Braun and Kelemen (2002) show that with the observed widths and frequencies of dunite veins in the Oman peridotite, the flux of liquid would have been great enough to create the thickness of the crust even if it had formed at a rapidly spreading ocean ridge.

Stevenson (1989) showed, on theoretical grounds, that partially melted rock undergoing deformation is unstable with respect to small-scale redistribution of the melt into veins (approximately meter scale) if the shear viscosity of the matrix is a function of the porosity, which it appears to be. Once formed, these veins may grow and intersect to form a network that allows the melt flux to increase still more (Maaløe, 2003). Experiments have shown that when partially melted peridotite is sheared, the melt does indeed segregate into thin sheets that are oriented ~20° to the shear plane (Holtzman and Kohlstedt, 2007). Theoretical calculations also indicate that shear will cause melt to migrate into lenses at a low angle (15–20°) to the shear plane (Holtzman *et al*., 2005; Katz *et al*., 2006). The development of these melt-rich bands increases the permeability of the partly melted rock but it also makes the permeability anisotropic. Melt flows faster parallel to bands than it does perpendicular to them. This anisotropy may actually lead to the focusing of fluid flow in the mantle toward mid-ocean ridges.

The importance of stress-driven melt segregation is beautifully illustrated in continental crustal rocks that have undergone partial fusion. At the highest grades of metamorphism in orogenic zones, partial melting has created rocks known as *migmatites* – mixed igneous/metamorphic rocks. One of the most striking features of these rocks is that the low melting light-colored granitic fraction (*leucosome*) is invariably segregated into sheets or channels that permeate the dark refractory residue (*melanosome*) on all scales from millimeters to meters (Figs. 3.13, 23.18). The movement of liquid from tightly compressed fold limbs into more open hinges or into shear zones is evidence that deformation accompanied the fusion.

Differences in the degree of X-ray attenuation between leucosome and melanosome has allowed the three-dimensional smaller-scale distribution of melt segregations in migmatites to be imaged in great detail through the use of nondestructive X-ray CT scans (Brown *et al*., 1999). Although a migmatite preserves only the last fluid flux prior to solidification, the CT scans do indicate that the channels tend to be planar and parallel. The 3-D CT image also allows the tortuosity of flow paths through the migmatite to be determined. Defining the tortuosity as the ratio of the flow path length to the straight-line

Fig. 3.13 Migmatite, a mixed igneous and metamorphic rock consisting of light-colored (leucosome) lenses and veins of granite, which formed by partial melting of the rock, and a dark grey (melanosome) refractory host. Aided by shear, the molten granitic fraction moved into fold hinges and shear zones. Width of field ~1 m.

length, Brown *et al*. (1999) found the average tortuosity in one sample they studied to be 3.6.

Shear stresses are far greater during formation of migmatites than they are during partial melting beneath mid-ocean ridges. Shear therefore plays a more important role in segregating liquid in migmatites (Burg and Vigneresse, 2002; Rabinowicz and Vigneresse, 2004). When stress is applied to a partially molten migmatite, strain occurs mainly in the melt fraction, because of its much lower viscosity (~10^7 versus ~10^{15} Pa s). The strain rate in the melt is greater than in the solid matrix, and consequently the melt moves more rapidly and tends to segregate, especially into shear zones (Fig. 3.13). Formation of these shear zones decreases the overall viscosity of the migmatite, which leads to still more strain partitioning into the melt. As the strain rate increases, the melt may become non-Newtonian and decrease its viscosity (shear thinning; see Section 2.4 and Fig. 2.5). As shear continues, melt is actually removed from the rock between the veins. This further increases the strain partitioning into the veins of melt and produces a positive feedback. This feedback results in pulses of expulsion of melt. Should the percentage of melt go above 20% to 30%, the solids lose cohesion and the bulk viscosity decreases by about 10 orders of magnitude. The field evidence from migmatite zones and theoretical studies indicate that shear channeling must play an important role in segregating bodies of granitic magma.

In summary, we can say that in zones of partial melting, liquid first forms as a dispersed phase along grain boundaries. Driven by buoyancy, this interstitial liquid moves slowly by porous flow as the solid matrix compacts. The flow, however, soon becomes channelized into small veins that connect to form a drainage network from the zone of partial melting. Although compaction is a viable means of extracting basaltic liquid from partially melted peridotite beneath mid-ocean ridges, it is not likely to remove highly viscous granitic melts from partially melted lower continental crust. There, shearing of the rock must play an important role in segregating the partial melts.

Fig. 3.14 Wine draining down the sides of a glass in a series of regularly spaced dribbles is an example of a Rayleigh–Taylor instability. The regular spacing of many features in nature is a result of such instability due to density differences (see Figs. 3.15 and 3.16).

3.11 DIAPIRIC INTRUSION OF MAGMA

One other mechanism of igneous intrusion that has been invoked, especially for the rise of granitic magma, involves *diapirs*. Diapirs form in response to a density inversion. A low-density layer overlain by denser material is gravitationally unstable, and given appropriate viscosities will try to invert. It typically does so by developing waves along the interface between the layers that grow into larger dome-like structures – diapirs. This phenomenon is referred to in fluid mechanics as a *Rayleigh–Taylor instability*, named after two of the early researchers in the field. Numerous examples of this phenomenon are found in nature. The common thundercloud, which results from cold dense air overlying warm less dense air, is perhaps the most familiar. When wine is swished around in a glass and drains back down the sides, it does so in a series of regularly spaced dribbles or "windows," the result of a Rayleigh–Taylor instability (Fig. 3.14). Sedimentary beds of salt (a low-density mineral) commonly form diapirs that intrude overlying beds of denser sedimentary rocks. Because magmas are less dense than equivalent solid rock, melting within the Earth typically sets up the conditions that would make diapiric rise possible. The book by Ramberg (1981) provides an excellent treatment of the subject; it is well illustrated with many scale models of geologically interesting examples of diapirs.

The sequence of photographs in Figure 3.15 illustrates the development of diapirs within low-density motor oil overlain

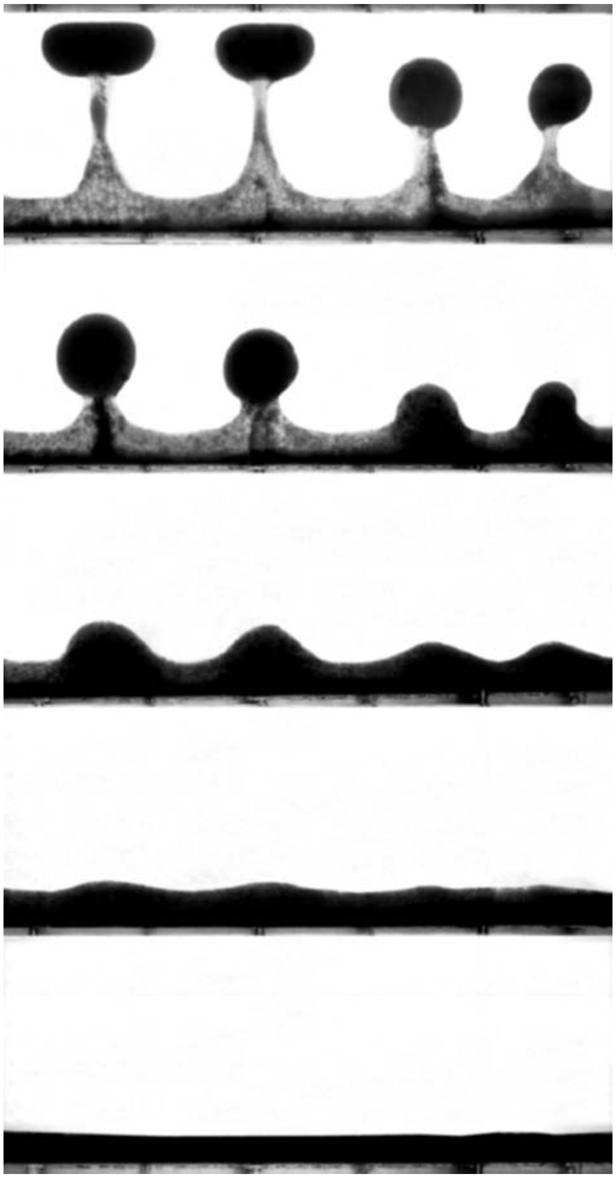


Fig. 3.15 Progressive stages in the development of buoyantly rising domes of oil into overlying denser honey. The wavelength of the disturbance on the oil–honey interface, which determines the eventual spacing of the domes, is determined by the viscosity contrast between the two liquids and the thickness of the oil layer and is an example of a Rayleigh–Taylor instability. Diapiric domes of salt, batholiths of granitic magma, and bodies of andesitic magma beneath island arcs may rise in a similar manner.

by denser honey. Figure 3.16 shows a similar-appearing set of diapirs developed between magmas of nepheline monzonite and alkali basalt composition in a horizontal sill near Montreal (Philpotts, 1972). Both figures illustrate an important feature of diapirs; that is, diapirs form in groups that have a characteristic spacing or wavelength. Regardless of the number of times the oil and honey model of Figure 3.15 is operated, the spacing between the diapirs remains the same, as long as the temperature and quantities of the two fluids are kept constant. If the thickness of the upper layer is great enough so that the upper surface does not interfere with waves that develop on the interface between the two fluids, Selig (1965) showed that a characteristic wavelength develops, which depends only on the thickness of the buoyant layer and the ratio of the viscosities. If the interface is perturbed by a range of wavelengths, one wavelength grows faster than all the others and, with time, dominates the pattern. This wavelength (λ) is given by

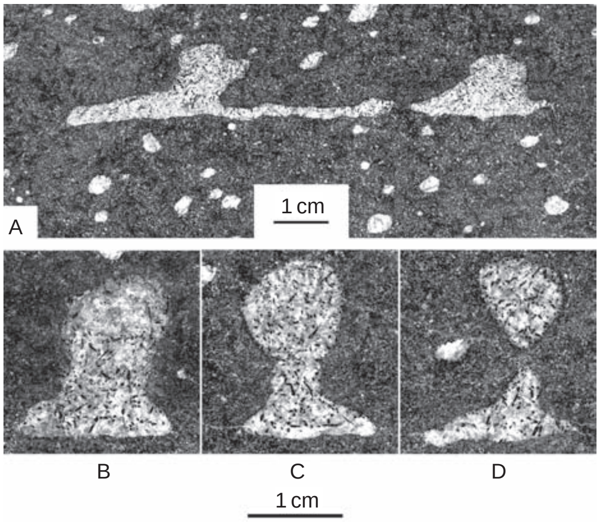


Fig. 3.16 Small diapiric domes of nepheline monzonite that rose from a thin layer of felsic magma in the upper part of a basaltic alkaline sill at Ste. Dorothée, Quebec. Serial sections through the small dome on the right in (**A**) are shown in photographs (**B**) to (**D**). The sheet of felsic liquid was less dense than the overlying basaltic magma; it was therefore gravitationally unstable and tried to rise as diapirs toward the top of the sill (after Philpotts, 1972). On a larger scale, granite batholiths may have the same general form.

$$\lambda = 2.92h\left(\frac{\eta_t}{\eta_b}\right)^{1/3} \tag{3.31}$$

where h is the thickness of the buoyant layer and η_t and η_b are viscosities of the top and bottom layers, respectively. Thus, if the viscosity of the bottom layer is decreased while the other parameters remain constant, the spacing between diapirs would increase. This is intuitively what one would expect, because the more fluid the lower layer, the greater is its lateral mobility.

The simple relation between viscosities, source-layer thickness, and diapir spacing can be used to determine any one of these properties if values of the others are known. For example, Philpotts (1972) used the spacing between the diapirs (~6 cm) developed on the thin layer (~0.25 cm) of nepheline monzonite in an alkali basalt sill (Fig. 3.16) to determine their viscosity ratio to be ~600 (basalt/monzonite). At a very different scale, Marsh and Carmichael (1974) determined that the thickness of the source layer of andesitic magmas along the top of the Benioff zone must be about 20 m to account for the typical 70-km spacing between volcanoes along many island arcs.

Despite the success of the diapiric model in explaining the regular spacing of salt domes, serious problems arise when it is applied to igneous bodies that have risen any distance through the lithosphere (Spera, 1980; Marsh, 1982; Mahon *et al*., 1988). As a diapir of magma rises into the progressively cooler lithosphere, it must remain at least partially molten if its viscosity is to be low enough to allow ascent. The high viscosity of the lithosphere (~10^{20} Pa s) causes ascent rates to be slow, which, in turn, allows heat to be lost to wall rocks and causes the magma to solidify. Marsh and Kantha (1978) calculate that for a 6-km spherical diapir of magma to rise to the surface of the Earth without solidifying, its ascent rate would have to be greater than 3 m a^{-1}. If magmatic diapirs are to ascend at these rates, they must do so by reducing the viscosity of the lithsophere through which they travel. This is most easily done by partial melting of the wall rocks (Marsh, 1982). But this, in turn, causes some of the magma to crystallize, thus reducing the amount of liquid that will reach the surface of the Earth. Repeated passage of diapirs through the same part of the lithosphere could raise wall rock temperatures and by so doing increase the percentage of liquid in each successive diapir.

The problems associated with the ascent rate have sparked a debate over whether granitic magma rises through the crust as diapirs or as dikes (Petford and Clemens, 2000). The passage of a diapir should leave a complex record of deformation in the surrounding rocks (Dixon, 1975; Cruden, 1990), which some claim is lacking. Others point out that the "stalks" or "tails" of diapirs are not found. In some circular bodies that have been interpreted as diapirs, based on their shape, a lack of a vertical lineation has been used to argue that no vertical movement occurred in the magma during the final emplacement. These bodies are interpreted to form through *ballooning*; that is, growing laterally with little vertical motion rather than as diapirs (e.g. Molyneux and Hutton, 2000).

In discussing the intrusion of magma, it is important to distinguish between the ascent of magma and its final emplacement. Rocks provide information about only the final stages of this ascent. Most of the textures and structures in igneous rocks are formed during the final emplacement stage; little information is preserved from the ascent stage (phenocrysts are an exception). In the oil and honey model in Figure 3.15, the sequence of photographs shows the ascent of the oil diapirs. However, only the diapir at the left end of the upper photograph is likely to resemble a diapir in nature, because its upward ascent has been stopped by the roof of the container. In nature, a diapir will stop rising when the overlying rocks can no longer be pushed aside. At this level, the buoyant magma will spread laterally. This is similar to magma spreading laterally at the level of neutral buoyancy, except that the barrier in this case is one of viscosity rather than density. In the upper photograph of Figure 3.15, the two diapirs on the right were still rising; the next one to the left had just arrived at the roof and was beginning to spread laterally; and the one on the left, which arrived at the roof first, shows considerable lateral spreading. If still later photographs had been included, they would show all of the diapirs flattened against the roof, and all of the stalks would have disappeared.

McCaffrey and Petford (1997) show that the majority of large bodies of granite can be shown through geophysical means to be rather flat and tabular with their thickness (T) to length (L) being related by a power law ($T = 0.6L^{0.6}$). They believe that the magma spread laterally from relatively local sources, which were probably dikes (Petford *et al*., 2000). However, the diapir in the upper left of Figure 3.15 can be seen to have been changing into a tabular body with the oil being fed from a small central feeder. Although granitic magma can ascend through dikes much faster than it can in diapirs (factor of ~10^6), is there evidence for such dikes? In deeply eroded parts of the continental crust, large regional basaltic dikes are common and attest to the passage of huge volumes of basaltic magma. However, no equivalent granitic dikes are found.

3.12 FACTORS IN SOURCE AFFECTING THE SUPPLY OF MAGMA

The rates at which magmas move are determined by numerous factors, some of which are dependent on the magma itself, such as density, viscosity, and yield strength, whereas others are determined by the environment, such as gravitational acceleration, lithostatic pressure, and conduit shape and roughness. We have also seen that although many magmas flow in a laminar fashion, ones with low viscosities may flow turbulently, especially when in wide conduits or thick flows. For each of the various types of flow, steady-state velocities can be determined. It would be a mistake, however, to conclude that magmas must always flow with these velocities. All magmatic activity is transient; it has a beginning and an end. There will be periods during which flow will be either increasing to the steady-state condition or diminishing toward the end of the activity. These periods of flux may, in fact, constitute the largest fraction of any particular igneous episode. It is important therefore to consider what might affect the rates of flow during these stages of activity.

Once steady-state flow has been reached during an episode of magmatic activity, it can be expected to continue as long as magma is available and the driving pressure remains the same. Should either of these factors decrease, magmatic flow would decrease or even stop. Because the driving pressure on a magma is most likely the result of buoyancy, any factors affecting the density of magma could cause fluctuations in flow rate. Apart from this, fluctuations are likely to result from changes in the availability of magma.

Bodies of magma are formed by the partial melting of mantle or crustal rocks followed by the segregation and separation of the liquid from its refractory residue. Magma containing a large fraction of refractory crystals does not have the bulk physical properties (low density and viscosity) necessary for it to ascend rapidly through conduits. When rocks partly melt, the first-formed liquid occupies spaces between grains of different minerals, but because of wetting properties, the liquid spreads along grain edge intersections. The rock therefore mops up the liquid as it is formed, in much the same way that a sponge absorbs water. Magma can flow only slowly between grains at this stage because the channels are narrow and tortuous. Furthermore, surface tension forces are significant at this scale. The most likely type of movement at this stage is the diapiric rise of an entire mass of partly melted rock. Only when magma is segregated from its refractory residue does viscosity become low enough for magma to rise rapidly through conduits.

A greater degree of melting is necessary before liquid is able to segregate and migrate upward. This degree of melting, which is analogous to the quantity of water a sponge can absorb before becoming saturated, is between 5% and 22%, depending on the surface tension between liquid and crystals. Once sufficient melt is produced to wet grain edges, segregation of liquid can take place. Here, shearing can greatly augment the segregation process. Rapid ascent through conduits may still not be possible because this early segregated liquid is likely to carry high percentages of crystals and thus behave as a Bingham liquid. Rapid flow would not occur until the fraction of liquid increased to the point where buoyant forces could overcome the yield strength of the liquid (see Problem 3.14). Once this happens, magma ascent is likely to be relatively rapid compared with that during the segregation stage (Petford and Clemens, 2000). Because of the different rates of these processes, intrusion of magma will be episodic and flow rates highly variable.

3.13 EVIDENCE OF FLOW IN ROCKS

Although magmatic flow plays such an important role in the formation of igneous bodies, little evidence of that flow is preserved in igneous rocks. This is in large part due to the fact that most textures and structures in igneous rocks develop only during the final stages of solidification, and by that time most flow has ceased. It is also becoming increasingly apparent that many slowly cooled igneous rocks have textures that are very similar to metamorphic rocks in that they have undergone recrystallization, albeit in the presence of a liquid. This means that some early formed textures can be modified or destroyed.

One place where evidence of flow can commonly be found is near igneous contacts. This is because the shear of magma increases toward contacts, and cooling rates are greater, so that primary textures are more likely to be preserved. Most magma, especially near contacts, flows in a laminar fashion. If the magma is composed largely of liquid, it will have Newtonian behavior. However, as it begins to crystallize, or if it is carrying a significant percentage of phenocrysts, the behavior is likely to be Bingham. In either case, we can expect the shear to be greatest at the contact.

If we were able to draw an inert marker circle in magma near the contact of a dike (Fig. 3.17), laminar flow would shear the circle into an ellipse and a line AB would be rotated to A′B′. Vesicles come close to providing such a marker. Although surface tension tries to keep bubbles spherical, this force is small except on very small bubbles (see Eq. (12.44)). Vesicles near contacts of dikes and in lava flows are deformed by flow. Figure 3.18(A) shows the contact of a dike that contains no phenocrysts but does contain a large number of vesicles (white) that have been deformed into ellipses by the upward flow of magma. Note that the largest vesicles are the most elliptical and the smallest are almost spherical. The largest vesicles formed first and record the most deformation; the smallest ones probably formed after the dike stopped flowing and hence are spherical.

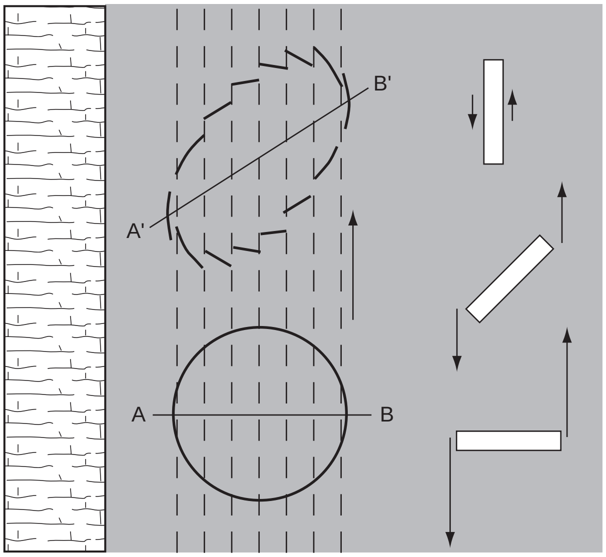


Fig. 3.17 Laminar flow of magma near the wall of an intrusion deforms a circle, such as that made by a vesicle, into an ellipse. Crystals are rotated by the shearing motion of the laminar flow. The crystal rotation velocity is determined by the difference in shear between the ends of the crystals (magnitude indicated by length of arrows). Because elongate crystals rotate more slowly when they parallel the flow plane, they commonly show a preferred orientation parallel to the flow planes.

If the dike contains phenocrysts, the shear of the liquid causes these crystals to continuously rotate. If the crystals are elongate or plate-like, such as plagioclase crystals, the rate of rotation is not constant. When an elongate crystal is perpendicular to the planes of laminar flow (Fig. 3.17) it experiences the greatest differential shear and rotates rapidly. As it rotates into the plane of flow, the differential shear decreases and its rotation velocity decreases. Although crystals that are unimpeded by coming in contact with other objects continue to rotate, there is more chance of finding crystals lined up parallel to the flow plane than at other angles because of the minimum rotation velocity in this orientation. This results in elongate and platy crystals commonly being aligned parallel to contacts (Fig. 3.18(B)). Note that this does not indicate the flow direction of the magma. Elongate minerals can develop a lineation, which gives a flow direction but not the sense of flow. For example, a vertical lineation could result from either upward or downward flow.

A rotating phenocryst may have its motion impeded by coming in contact with other phenocrysts. This results in crystals stacking against each other in an *imbricate* fashion

(Fig. 3.18(C)). Such a pattern does indicate the flow direction of magma. Crystals can even be broken during this stacking process and the relative motion of the fragments can also indicate the flow direction (Fig. 3.18(D)).

Magma solidifies over a range of temperature (~200 °C) and as it does, the composition of the residual liquid changes its composition, often significantly. For example, in many basalts, the final liquid is of granitic composition. If flow continues during solidification, this residual liquid may migrate into zones of relative low pressure caused by the shear of crystallizing matrix. Figure 3.18(E) shows plagioclase phenocrysts aligned parallel to the contact of a basaltic dike. Thin stringers of granitic material (light color) trail off from opposing ends of the phenocrysts indicating the direction of flow.

Away from contacts and in the interior of large igneous bodies, rocks may exhibit foliations and lineations that have formed through magmatic flow. In coarse-grained rocks, these features can be mapped in the field (e.g. Molyneux and Hutton, 2000). In finer-grained rocks, however, time-consuming microscopic studies may be necessary to determine preferred orientations of crystals. One technique that has been used widely to infer magmatic flow directions, especially in finer-grained rocks, involves measuring a rock's anisotropy of magnetic susceptibility (AMS). The strength of magnetization induced in a sample when placed in a magnetic field can be measured in minutes and with such high precision (1:10 000) that very small degrees of anisotropy can be detected (Tarling and Hrouda, 1993). Although the AMS fabric of a rock is easily obtained, using it to determine the magmatic flow directions is often difficult because of uncertainty surrounding the origin of the AMS. In rocks that contain a significant amount of magnetite, the shape and distribution of the magnetite grains probably controls the AMS. However, in most igneous rocks, magnetite is one of the last minerals to crystallize; in addition, it is an isometric mineral. Therefore, it would not be expected to record magmatic flow directions. Hargraves *et al*. (1991), however, have suggested that early formed silicates that do preserve a record of the flow form a template around which the late crystallizing magnetite forms. There is no doubt that AMS fabrics, such as imbricate patterns of the AMS ellipsoid on either side of dikes, do record magmatic flow directions. Used with caution, this technique can provide valuable textural information relating to flow (Philpotts and Philpotts, 2007).

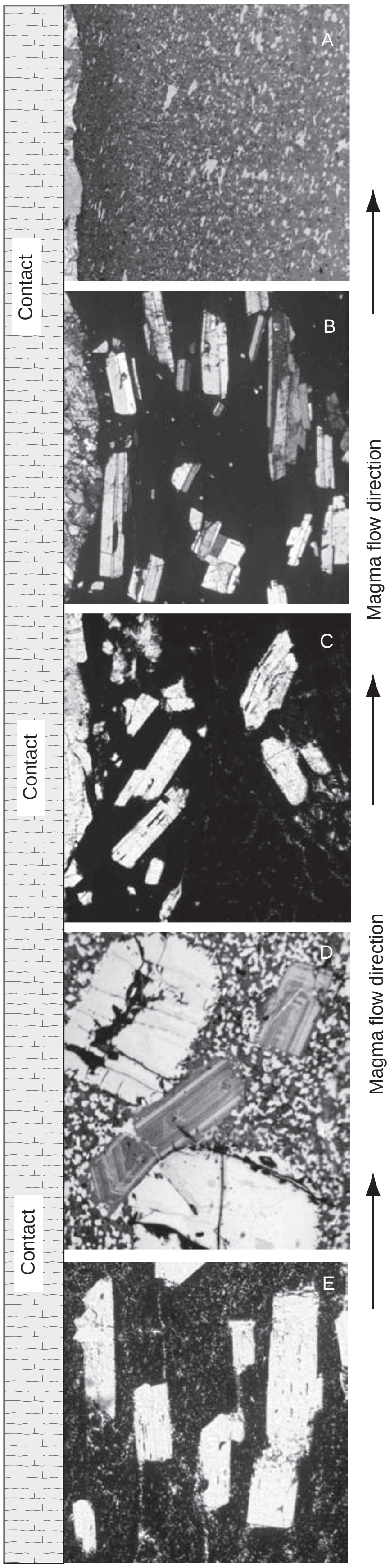


Fig. 3.18 (**A**) The larger vesicles near the margin of a camptonite dike indicate upward flow of magma. Small vesicles are spherical and formed after flow had stopped (Philpotts and Philpotts, 2007). (**B**) Plagioclase crystals aligned parallel to the margin of basaltic dike. (**C**) Plagioclase crystals stacked in an imbricate fashion by the upward flow of magma. (**D**) Oscillatory zoned plagioclase phenocryst (gray) broken by orthopyroxene phenocrysts (white) with sigmoid distribution of fragments indicating magmatic flow direction (HF-fumed polished section in reflected light). (**E**) Granitic residual liquid (light color) trailing from diagonally opposite sides of plagioclase phenocrysts, indicating the shear direction in the margin of a basaltic dike. All photomicrographs are several millimeters wide. Those in (

3.14 PROBLEMS

3.1 Axial magma chambers (AMC) have been seismically identified along the length of the Galapagos spreading axis (Blacic *et al*., 2004). They tend to be shallower in the east near the Galapagos hot spot, where they are 1–2.5 km beneath the seafloor, and deeper to the west where they are up to 4.5 km below seafloor. If the AMC is 3 km below seafloor, what thicknesses of extrusive rock, with density of 2.15 Mg m^{-3}, would you expect to overlie intrusive rocks, with a density of 2.9 Mg m^{-3}, if the magma has a density of 2.75 Mg m^{-3}, and buoyancy provides the driving force for the magma?

3.2 In the example worked out in Section 3.2, magma of density 2.9 Mg m^{-3} that was able to rise buoyantly just to the surface of the Earth through a mantle of density 3.3 Mg m^{-3}, 10 km of lower crust with density 3.0 Mg m^{-3}, and 25 km of upper crust with density 2.75 Mg m^{-3} was shown to have originated at a depth of 41.9 km beneath the surface. In this example, the magma was taken to be incompressible, but if the coefficient of compressibility, β, is 5.0×10^{-11} Pa^{-1}, what would the depth to source be if the surrounding rocks are again taken to be incompressible and g to remain constant over the depth considered?

(Before doing any problem, it is worth anticipating the answer. For example, in this case, will the depth to the source be greater or less than that in the example where the magma was taken as being incompressible? In working a problem, equations are often derived that can be solved only by trial and error (as in this case). By substituting estimated values for the unknown and noting the resulting error, a correct value can soon be obtained. The *Solver* function in Microsoft Excel provides another way of solving such equations.)

3.3 **(a)** Epp (1984) states that Mauna Loa towers 8192 m above the floor of the Pacific Ocean, but only 4169 m of this is above sea level. He estimates that about 150 m of ocean-floor sediment covers the oceanic crust, which is 6.5 km thick in this region. The densities of the various materials, in Mg m^{-3}, are: seawater 1.0, sediments 2.0, crust 2.84, mantle 3.1, and magma 2.9. Calculate the depth of origin below sea level of an incompressible magma that rises buoyantly to the summit of Mauna Loa.

(b) Magma erupting onto the floor of the caldera of Kilauea, to the southeast of Mauna Loa, rises only 1240 m above sea level. If densities and thickness of materials beneath this volcano are similar to those beneath Mauna Loa, what difference in the depth to source region might be expected between these two volcanoes? Does this calculated depth for the source of magma agree with evidence from seismic disturbances associated with Hawaiian eruptions? (See Section 4.3.)

3.4 Assume that magma is able to rise buoyantly through the lithosphere just to the surface of the Earth, and that the lithosphere is stratified with respect to density as indicated in the following table. Plot a graph showing the excess pressure, ΔP, in a column of magma as a function of depth for **(a)** a komatiitic magma with density 2.8 Mg m^{-3}, **(b)** a basaltic magma with density 2.7 Mg m^{-3}, **(c)** an andesitic magma with density 2.6 Mg m^{-3}, **(d)** a rhyolitic magma with density 2.4 Mg m^{-3}. The density of each magma is taken to remain constant with depth. On the same graph, plot a line marking the maximum tensile strength of the rocks, which can be taken to be about 10 MPa. At what depth might plutonic bodies of each of these magmas form?

Depth (km)	Density (Mg m^{-3})	ΔP (Pa)
		– 0 +
0		
	2.3	
3		
	2.75	
25		
	3.0	
35	3.3	

3.5 The maximum amount of water soluble in a basaltic melt at normal magmatic temperatures (~1200 °C) increases with pressure according to the approximate relation

$$\text{wt\% } H_2O \text{ in basaltic magma} = 6.8 \times 10^{-6} P^{0.7}$$

where P is measured in pascal.

(a) If a rising basaltic magma with a density of 2.7 Mg m^{-3} becomes saturated in water at a depth of 5 km, construct a graph showing the weight fraction of exsolved water as the magma continues toward the surface, assuming that the magma contains the saturation amount at all depths and the magma's density and temperature remain constant. Add to this graph a similar plot for an 800 °C granitic magma with a density of 2.4 Mg m^{-3} that also becomes water saturated at a depth of 5 km. The wt% H_2O that can be dissolved in a granitic melt as a function of pressure is given approximately by

$$\text{wt\% } H_2O \text{ in granite magma} = 0.411 \times 10^{-3} P^{0.5}$$

(b) Weight fractions can be converted to volume fractions by dividing by the densities. The density of water can be calculated by assuming that water obeys the ideal gas law ($PV = nRT$) where n is number of moles, R the gas constant, and T the absolute temperature. Note that the pressure on the surface is 10^5 Pa. Calculate at what depth the volume fraction of H_2O gas reaches 0.64 and disrupts the basaltic and granitic magmas. What conclusion might you draw concerning the nature of the eruptions of the two magmas?

3.6 During the 1959 eruption of Kilauea Iki, a fountain of lava rose from the vent to a height of 450 m, and on one occasion, to 580 m. The ponding of this lava within the crater (Fig. 2.2) allowed for accurate monitoring of the volumes of erupted material in much the same way as a measuring cylinder is used to determine volumes of liquid in the laboratory. The flux of lava was estimated to have varied from 100 to 300 $m^3\ s^{-1}$. Although there were periods of more violent eruption brought about by escaping gas, the steady fountaining is thought to have simply resulted from the velocity of ejection of magma from the orifice.

- **(a)** Using the height of 450 m and neglecting air resistance, calculate the maximum velocity of ejection of lava from the orifice. Assuming a parabolic velocity profile across the orifice (i.e. fully developed flow), determine the average velocity of ejection.
- **(b)** Assuming a cylindrical feeder pipe, use the average velocity from part (a), along with the rate of ejection of 100 $m^3\ s^{-1}$, to determine the diameter of the pipe.
- **(c)** If the magma had a density of 2.75 Mg m^{-3} and viscosity of 300 Pa s, would you expect turbulent or laminar flow in a feeder pipe with the diameter calculated in part (b)?
- **(d)** If the diameter of the feeder pipe calculated in part (b) remains constant down to the magma chamber, how far would the magma have to rise before fully developed flow was achieved in the pipe? Was the assumption of parabolic velocity profile in part (a) justified?

3.7 If the diameter of the Kilauea Iki feeder pipe was 1.65 m, what would the (**a**) maximum and (**b**) average velocities have been for the magma if it rose in response to a pressure gradient induced solely by the loading of solid basalt ($\rho = 3.0$ Mg m^{-3}) on the magma ($\rho = 2.75$ Mg m^{-3})? The viscosity of the magma is 300 Pa s.

3.8 If the pressure gradient causing the eruption of 100 $m^3\ s^{-1}$ of lava from Kilauea Iki in 1959 was due solely to the loading of solid basalt on the magma, what diameter would you expect the feeder pipe to have? The viscosity and density are the same as in Problem 3.7.

3.9 The 1959 eruption of Kilauea Iki developed a deep lava lake that eventually covered the vent. During periods of quiescence, the lava from the lake drained back down the vent at the incredible rate of 425 $m^3\ s^{-1}$. Assuming that the lava attained a constant rate of back flow, and the pressure on the base of the column was the same as that on the top (1 atm, or 10^5 Pa), the only driving force would have been that of gravity. Calculate the diameter of the vent using a magma density of 2.75 Mg m^{-3} and viscosity of 300 Pa s.

3.10 Discuss differences in the calculated diameters of the Kilauea Iki feeder pipe in Problem 3.6, 3.8, and 3.9. Indicate which method of calculation is likely to give the most accurate value. How does the power to which the radius is raised in the various equations affect the sensitivity of the results?

3.11 Derive Eq. (3.15), which expresses the ascent velocity of magma at any point across a vertical dike that is in steady-state laminar flow. Begin by considering the forces acting on a small magma volume, dx dy dz.

3.12 Starting with Eq. (3.15), derive an expression for the average laminar flow velocity in a dike.

3.13 Prove mathematically that the maximum velocity in a dike does, indeed, occur at its center.

3.14 What fraction of a 10-m-wide dike of dacite that behaves as a Bingham liquid would flow as a plug if the magma has a yield strength of 10^4 Pa and the magma is buoyantly emplaced in response to a density contrast between magma and country rocks of 300 kg m^{-3}?

3.15 Starting with Eq. (2.13), which expresses the laminar flow velocity as a function of height within a lava flow, derive an expression for the average laminar flow velocity of a lava.

3.16 Why are each of the transport laws (Hagen–Poisseuille, Darcy's, Fourier's, and Fick's) preceded by a negative sign?

3.17 Jurassic basalts in eastern North America, associated with the initial opening of the Atlantic Ocean, are compositionally similar to those still being erupted today at the Mid-Atlantic Ridge in Iceland. In both areas, erosion has exposed dikes that were feeders to fissure eruptions. In Iceland these dikes average 4 m in width, but in eastern North America many are as much as 60 m wide. Assuming that magma rose buoyantly in both areas and that the average density of the intruded lithosphere was 3.0 Mg m^{-3} in Iceland and 2.9 Mg m^{-3} in eastern North America, determine whether the flow was laminar or turbulent in these two areas, and calculate the average flow velocities in the respective dikes. The basaltic magma in both areas can be taken to have had a density of 2.6 Mg m^{-3}, and a viscosity of 300 Pa s, and the friction factor in the dikes was 0.01.

3.18 Volcanoes along the Aleutian Arc are spaced ~70 km apart, whereas those in the adjoining Kamchatka belt are only 30 km apart. Assuming the viscosity of the lithosphere in the two regions to be similar, speculate on what might account for the different spacings.

4 Forms of igneous bodies

4.1 INTRODUCTION

Bodies of igneous rock are referred to as *extrusive* when formed on the Earth's surface, and *intrusive* when formed within the Earth. Intrusive bodies are further subdivided into *plutonic* – large intrusions formed at moderate to great depth, and *hypabyssal* – small intrusions formed near the Earth's surface. Hypabyssal rocks cool rapidly, and most resemble volcanic rocks more than plutonic ones. Extrusive bodies of igneous rock are relatively well understood, for many of them have actually been studied during their formation. The same cannot be said for plutonic bodies, where even their form may be uncertain and their formative processes often conjectural. We, therefore, start with a discussion of extrusive bodies.

Extrusive bodies

When thinking of volcanic forms, it is natural to picture a volcano such as Mount Fuji. Although such spectacular edifices are common, the largest volumes of volcanic rock in the geological record erupted to form less striking features. By far the most voluminous volcanic rocks on Earth are the basaltic flows covering the ocean floor. Most of these were erupted from fissures along mid-ocean ridges where they poured out to fill the axial valleys. Fissure eruptions of basalt also constitute the most voluminous rocks on continents. Coming in a close second with regard to volume are huge sheets of rhyolitic volcanic ash. Nine percent of the world's population lives within 100 km of an active volcano, and although many large eruptions in historic times have had disastrous effects on humans (de Boer and Sanders, 2004), no eruption has been on the scale of the large ones found in the geologic record. Such large eruptions are bound to recur, and when they do, we will be grateful for all the research that has been done in predicting eruptions. Figure 4.1 is a cartoon depicting the approximate relative volumes of the different types of volcanic edifices.

4.2 FLOOD BASALTS

In zones of crustal extension, basalt is commonly erupted from long fissures. This type of eruption is continually occurring along mid-ocean ridges. At many times in the Earth's history, however, basalt has poured out on continents in such large quantities as to flood thousands, and possibly millions, of square kilometers. These areas, which are referred to as *large igneous provinces* (LIP), are thought to form above mantle plumes (Saunders, 2005). Rates of extrusion are sufficiently great and viscosities low enough to allow basalt to spread almost horizontally and to flood the local topography. For this reason, these lavas are referred to as *flood basalts*. They also give the topography a plateau appearance, which explains why they are also referred to as *plateau basalts*. Where erosion has cut down through these flows, slopes tend to be stair-like, with steep cliffs in the lower part and gentler slopes in the upper part of each flow (Fig. 4.2). The Dutch word for stairs is "trap" (Swedish – *trappa*) and, consequently, this type of basalt is commonly referred to as *trap rock*, especially in the quarrying industry where it makes the most durable aggregate of all rock types.

Flood basalts are the most voluminous type of extrusive rock. The eruption at the end of the Cretaceous of the Deccan trap of western India, which covers an area of at least 500 000 km^2 and averages 600 m thick, has been blamed for climatic changes that may have caused the extinction of dinosaurs (Wignall, 2005). In the western United States, the Miocene Columbia River basalts in the state of Washington cover an area of 200 000 km^2 and in places are more than 1500 m thick (Fig. 4.2). To the southeast, in Idaho, the Snake River plain is underlain by 50 000 km^2 of basalt. In other areas, erosional remnants, such as the 4500-m-thick sequence of Precambrian basalts on the Keweenaw Peninsula and north shore of Lake Superior, are all that remain to attest to the former existence of these great sheets. Perhaps the most extensive of all were those of the Tertiary Brito-Arctic, or Thulean province (Fig. 15.12) now exposed as erosional remnants in Northern Ireland, Scotland, Iceland, and Greenland. In addition to these major fields, there are many smaller areas of flood basalt, such as those of the various Triassic–Jurassic basins of eastern North America (Fig. 4.49). Fortunately for the human race (but not for petrologists), no flood basalts have erupted during historic times. The 1783 fissure eruption of Lakagigar, Iceland (L in Fig. 4.3 inset), which is one of the largest eruptions witnessed, killed 75% of the livestock in Iceland and one quarter of its population as a result of famine. With a volume of only 14 km^3, however, it is small by comparison with those in the geologic record (Guilbaud *et al*., 2005).

The most voluminous production of fissure eruptions is from the Earth's system of mid-ocean ridges. Parsons (1982) estimates that their annual production, in the form of lavas and underlying feeder dikes, has been constant for the last 180 Ma, and is ~24 km^3 (6.7×10^{13} kg a^{-1}). This value has

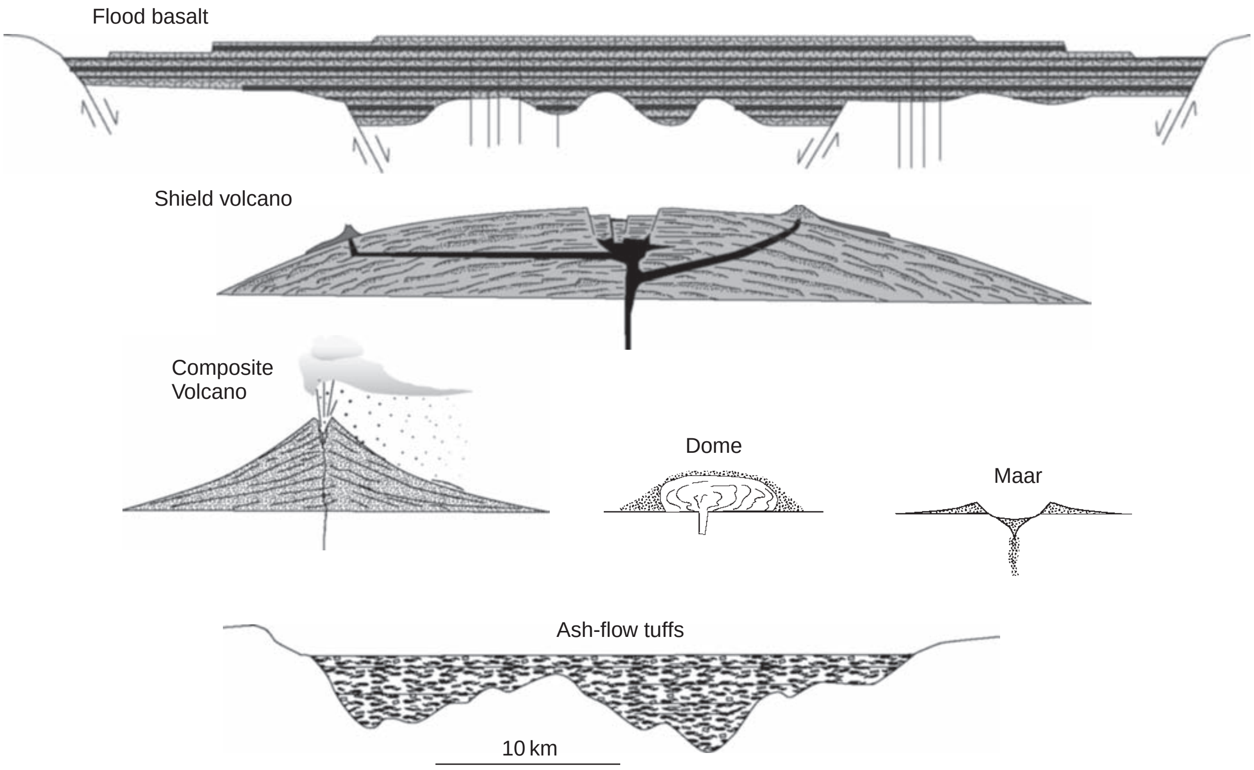


Fig. 4.1 Forms of the major volcanic structures and their approximate sizes. *Flood basalts* and *shield volcanoes* are composed of basaltic rocks, *composite volcanoes*, of andesitic rocks, and *domes* and *ash-flow tuffs*, of rhyolitic or other feldspar-rich rocks. *Maars* are explosion craters surrounded by debris from the vent.

been confirmed with recent detailed data from the ocean floor (Rowley, 2002). Because the basaltic rocks are erupted along mid-ocean ridges, they are referred to as *mid-ocean ridge basalt* or *MORB* for short. Although created at mid-ocean ridges, MORB is transported across ocean floors by plate motion. Consequently, most of the ocean floor is underlain by MORB, which makes it the most abundant rock type on Earth. Despite the rapid production of MORB, most ocean floor is subducted back into the mantle within 200 Ma, so that MORB is not an abundant rock type in the older geologic record. Only rarely in subduction zones do some ocean floor rocks get incorporated and preserved with continental rocks (obduction). These produce the ophiolites, which are discussed in Section 15.2. Where the Mid-Atlantic Ridge traverses Iceland, the basalts that elsewhere are erupted beneath the ocean can be seen on land. Figure 4.3 is a view looking north from the Krafla volcano (K on inset map) along this rift as it exits the northern side of Iceland. In the foreground are lavas that were erupted from fissures during the most recent period of activity (1975 to 1984). Only one-third of the magma flux during this period actually erupted; the remainder moved northward, with velocities up to 0.5 m s^{-1}, in a dike swarm that extended at least 30 km (Sigurdsson and Sparks, 1978; Gudmundsson, 2000).

Flood basalts owe their great thicknesses mostly to the accumulation of large numbers of thin flows, but some thick flows may also be present. Individual flows typically range in thickness from meters to tens of meters and more rarely to 100 m (Fig. 4.2). The Greenstone flow in the Keweenaw Peninsula, which is up to 600 m thick (Fig. 4.7(B)), is possibly the world's thickest. It also crops out on Isle Royale on the north side of Lake Superior, indicating that at the time of eruption it formed a gigantic lava lake possibly the size of the present Lake Superior. Flood-basalt flows extend laterally for many kilometers and in some cases, several hundred kilometers. Commonly, the lateral extent of a flow away from its source is approximately proportional to the third power of its thickness near the source. Volumes of individual flows average several tens of cubic kilometers but can be hundreds or even thousands of cubic kilometers.

For sheets of lava to spread over great areas, they must do so before cooling makes them too viscous to flow. In light of the calculated flow rates of various types of lava in the problems of Chapter 2, it is not surprising to find that basalts are the only lavas to form these laterally extensive flows. However, it remains a puzzle how even basaltic lava can spread for hundreds of kilometers without solidifying. Shaw and Swanson (1970) proposed that the flux of magma from feeder dikes must have been in excess of 1 km^3 day^{-1} for each kilometer of dike length, and that the lavas spread as thick turbulent flows. Self *et al*. (1996) proposed that the laterally extensive flows thickened and extended themselves by a process of inflation, with magma intruding beneath a solid crust. The advantage of this mechanism is that the magma remains insulated from its surroundings and therefore does not cool as rapidly. Ho and Cashman (1997) showed through melting experiments that the temperature of the 30- to

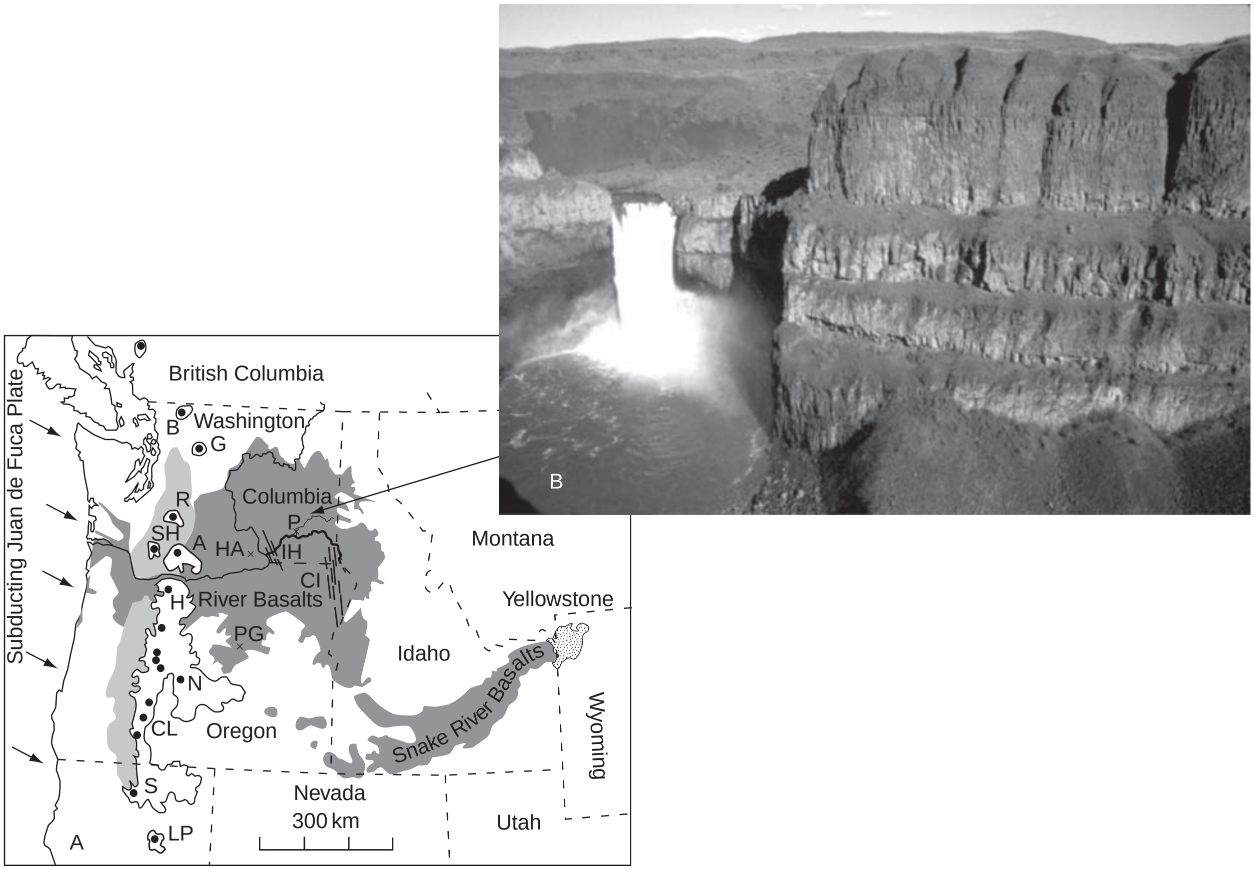


Fig. 4.2 (**A**) Volcanic rocks of the northwestern United States. Lavas from the older Cascade composite volcanoes are shown in light gray just to the west of the younger High Cascades whose lavas are outlined in white, with the main composite volcanoes being marked by circles and letters: Baker (B), Glacier (G), Rainier (R), St. Helens (SH), Adams (A), Hood (H), Newberry (N), Crater Lake (CL), Shasta (S), and Lassen Peak (LP). These volcanic rocks were formed above the subducting Juan de Fuca plate. Flood basalts of the Columbia and Snake Rivers are shown in dark gray. The Ice Harbor (IH) and Chief Joseph (CJ) dike swarms were feeders to the Columbia River basalts. (**B**) The 60-m-high Palouse Falls (P) shows a section through several of the flood-basalt flows. The Hanford Nuclear Site (HA) and Picture Gorge (PG) are referred to in other figures. The rhyolitic rocks of Yellowstone mark the present position of a hot spot whose trace has produced the linear belt of Snake River Basalts as the North American plate has moved westward. (Sources of data: Waters, 1955; Smith and Christiansen, 1980; Reidel *et al*., 1994.)

Fig. 4.3 View looking north along the Krafla rift system from which basaltic lava erupted between 1975 and 1984. The width of the down-faulted rift system, which is part of the Mid-Atlantic Ridge crossing Iceland (see inset map), can be seen extending from the left to the right side of the photograph; the higher hills to the left and right are older flood-basalt flows. The inset map shows the younger volcanic rocks, which occupy the rift where the Mid-Atlantic Ridge crosses Iceland. The locations of Krafla (K) and the 1783 Laki fissure eruption (L) are shown.

Fig. 4.4 Photomicrograph of basalt from the center of one of the thick Columbia River Basalt flows, the 70-m-thick Cohassett flow, sampled beneath the Hanford Nuclear Site (HA in Fig. 4.2). The sample shows randomly oriented elongate laths of plagioclase and pyroxene in glass. Had the lava still been moving when the plagioclase crystallized, the crystals would have become aligned. This typical texture of flood basalts indicates they are emplaced rapidly. Partially crossed polars. Width of field 1 mm.

Fig. 4.5 Plan view of columnar joints formed during cooling and shrinkage of a flood basalt at the "Church Floor", Kirkjubæjarklaustur, Iceland. (Photograph by Doreen Philpotts.)

70-m-thick Ginko flow of the Columbia River Basalt Group decreased <20 °C in flowing 500 km from its source near Palouse Falls (Fig. 4.2) to the Pacific coast south of Portland, Oregon. This indicates that the lava had to travel in an insulated jacket (hence by inflation) and must have been in laminar flow, because turbulence would have caused greater heat loss.

Most flood basalts have a characteristic texture in which randomly oriented plagioclase crystals either are embedded in or are wrapped around pyroxene crystals (Fig. 4.4). The lack of alignment of plagioclase crystals indicates that flow ceased before most crystallization took place. This requires rapid emplacement. Ho and Cashman (1997), for example, calculate that the Ginko flow took only six days to travel 500 km. The random orientation of crystals is one of the reasons why trap rock is so tough and makes such desirable aggregate for mixing with asphalt for road surfacing.

As flood-basalt flows cool, they shrink and develop a characteristic set of fractures which, because they break the rock into regular five- or six-sided columns, are known as *columnar joints* (Figs. 4.5, 4.6). Because the stress necessary to fracture the rock is generated by cooling (and the coefficient of thermal expansion), the fractures propagate perpendicular to the cooling isotherms. Cooling from the upper surface is influenced by major fractures into which water may percolate and cause additional cooling when the water is converted to steam. These "cold fingers" distort the isotherms, which cause the downward-propagating fractures to radiate out from these fingers (Fig. 4.6). In the lower part of the flow, heat is lost by conduction into the underlying rocks. The cooling isotherms consequently parallel the lower surface of the flow and the fractures propagate up from the base as regular columns. Because cooling from the surface of a flow, especially if augmented by evaporation of water, is faster than heat conduction into underlying rocks, the downward and upward propagating fractures meet at approximately one-third the height of the flow.

The striking difference between the fractures in the lower and upper parts of flows is used to subdivide flows. By analogy with Greek architectural elements, the regular columns extending up from the base of a flow are referred to as the *colonnade* and the upper zone of downward-splaying fractures, as the *entablature* (Fig. 4.6). In some flows, the vesicular flow top may also exhibit an upper colonnade. Although this division is based on the way in which the rock fractures, we will see in Section 14.5 that there are important differences in the way the magma crystallizes in these zones, the entablature undergoing rapid disequilibrium crystallization, whereas the colonnade crystallizes more slowly under nearly equilibrium conditions where magmatic differentiation processes may play a role. Long and Wood (1986) have argued that the fracturing and crystallization that characterize the entablature are the result of quenching when water flows across the surface of a flow. Some flows contain more than one level of entablature, which they attribute to repeated flooding of the surface of the flow. Lyle (2000) has shown that these multitiered flows are usually covered with fluvial sediments, which supports the water-quenching hypothesis.

At pressures below the brittle–ductile transition, the tensile strength of rock increases with confining pressure. Rock near the base of a thick flow is therefore stronger than near the top, and a larger volume of rock must shrink on cooling to produce the stress necessary to cause fracturing. The diameter of the columns in the lower part of a flow is therefore greater than in the upper part. Similarly, the size of columns in thick flows is greater than in thin flows (Fig. 4.7). In the 600-m-thick Greenstone flow of the Keweenaw Peninsula, Michigan, the columns are ~10 m in diameter, which is similar to the size found in some intrusive bodies of rock.

Although columnar joint surfaces are relatively smooth, many exhibit a small amount of relief, which breaks the surface into narrow (centimeter to decimeter) strips

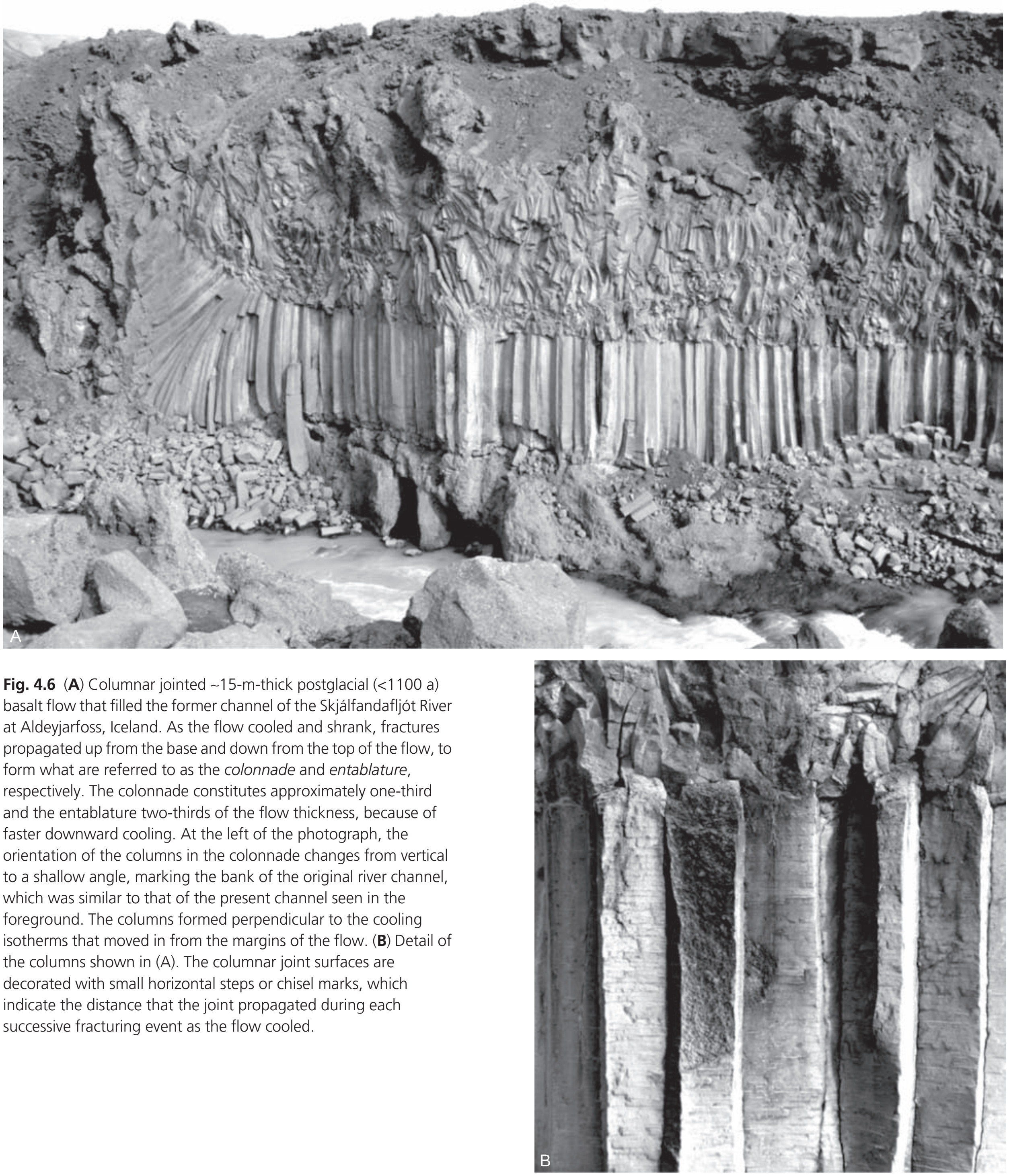


Fig. 4.6 (**A**) Columnar jointed ~15-m-thick postglacial (<1100 a) basalt flow that filled the former channel of the Skjálfandafljót River at Aldeyjarfoss, Iceland. As the flow cooled and shrank, fractures propagated up from the base and down from the top of the flow, to form what are referred to as the *colonnade* and *entablature*, respectively. The colonnade constitutes approximately one-third and the entablature two-thirds of the flow thickness, because of faster downward cooling. At the left of the photograph, the orientation of the columns in the colonnade changes from vertical to a shallow angle, marking the bank of the original river channel, which was similar to that of the present channel seen in the foreground. The columns formed perpendicular to the cooling isotherms that moved in from the margins of the flow. (**B**) Detail of the columns shown in (A). The columnar joint surfaces are decorated with small horizontal steps or chisel marks, which indicate the distance that the joint propagated during each successive fracturing event as the flow cooled.

perpendicular to the long axis of the column (Figs. 4.6(B), 4.8). In the colonnade, these strips, which are commonly referred to as *chisel marks*, give the joint surface the appearance of siding on a house; that is, at the base of a strip, the joint surface curves out abruptly from the general plane of the joint and then curves back slowly until it parallels the general joint surface; then another step occurs, and the pattern keeps repeating. DeGraff and Aydin (1987) have shown that these steps are formed by individual fracture events that occur successively as the column propagates into the flow. As stress builds up in the cooling flow, a fracture nucleates and spreads rapidly in a horizontal direction until it terminates against another fracture. It also propagates into the flow, but only for a short distance, because on advancing, the fracture approaches hotter rock that has not yet cooled enough to develop sufficient stress to fracture. The fracture therefore terminates along an isotherm. Following some period of cooling, stress builds up and another fracture nucleates, and

Fig. 4.7 The column diameters in the colonnade of flows depends on the thickness of the flow. In the 15-m thick flow of Figure 4.6, joint diameters are ~0.5 m. (**A**) In the ~50-m-thick Columbia River flood basalt at Sentinel Gap on the Columbia River just north of the Hanford Nuclear Site (HA in Figure 4.2), the column diameters are ~1 m (note five-foot-two person, 1.57 m, for scale). These columns exhibit a common pinch and swell feature, a shape that is not found on tectonic joints. (**B**) In the 600-m-thick Greenstone flow of the Keweenaw Peninsula, Michigan, the column diameters are almost 10 m. The view shown here is looking northward, up a single column near the base of the flow, which dips to the north under Lake Superior.

Fig. 4.8 Columnar joint surfaces are commonly segmented into horizontal steps that resemble the siding on a house; that is, at the base of each step, the fracture surface curves out abruptly and then gently tapers back in to parallel the overall columnar joint surface. Each step marks the distance a fracture moves into the flow during an individual fracture event. At the end of each fracture episode, cooling must build up stresses before another fracture is nucleated.

the process repeats itself. The chisel marks therefore indicate the stepwise propagation of the columnar joints into the flow. The direction of propagation can be determined from the asymmetry of the steps; that is, in the lower part of the flow where the joints propagate upward, the steps appear like the siding on a house (see the colonnade in Figs. 4.6(B) and 4.8), whereas in the entablature where the joints propagate downward the “siding” is upside down.

In deeply dissected flood-basalt provinces, large numbers of parallel dikes are exposed, which were the feeders to the flows. The total thickness of these dikes indicates considerable crustal extension. For example, in parts of western Scotland and in Iceland the dikes account for crustal extensions of about 5%. This extension, of course, is associated with the opening of the North Atlantic. On continents, extension of the lithosphere results in the formation of rift valleys, and many flood basalts occur within these valleys. Figure 4.58, for example, shows flood basalts on the east side of the Afar Rift, which are cut by normal faults formed as the Arabian plate moved away from the Nubian plate. We will see in Chapter 23 that the upwelling of the mantle beneath zones of extension leads to partial melting and formation of flood-basalt magmas. If buoyancy is the force responsible for the ascent of these magmas, the Earth’s surface should subside as the basalts accumulate in thickness. This is borne out by field observations. For example, deep drill holes through the Columbia River basalts indicate that the lowest members are now below sea level, but these flows were clearly erupted on land well above sea level as indicated by associated sediments, the lack of pillows, and the fact that they flowed hundreds of kilometers toward the sea. These deepest flows must therefore have subsided as they were covered by successive flows.

Rift valleys are commonly occupied by lakes, and when basaltic lava enters water, it tends to divide and bud into large sack-like bodies that are referred to as *pillows* (Fig. 4.9). Much of the basalt on the ocean floor is consequently pillowed. Water causes rapid cooling of the advancing surface of the flow, which is quenched to a glass. Pressure in the flow’s interior builds and eventually bursts the glassy selvage, and a lobe of lava extrudes and necks off to form a pillow that rolls down the front of the advancing flow. Although an approximately one-centimeter-thick rind of glass forms on pillows almost immediately, their interior remains molten for up to several hours, depending on the size of the pillow. Consequently, they are deformable, and they mold themselves to the shape of earlier pillows on which they come to rest. Pillows therefore typically have a tail on their lower side where they sag between earlier solidified pillows, whereas

Fig. 4.9 Pillowed Talcott basalt at Meriden, Connecticut, was erupted into a lake that occupied the Hartford Mesozoic Basin, one of the many basins formed in eastern North America during the initial opening of the Atlantic Ocean. The pillows sank into the soft mud on the lake bottom, displacing it upward between other pillows. The way the pillow just above the center of the photograph sags between the two underlying pillows shows that it was still deformable when it came to rest on the underlying pillows. Dark glassy selvages can be seen on the pillows. Some of this glass broke off the pillows as they deformed to produce a porous breccia between the pillows, which was later cemented with quartz, calcite and zeolites (white).

their upper surface is convex. This shape can be used to determine top directions in folded volcanic rocks (Fig. 4.10).

Pillows may vary in size and shape from buns to mattresses, and show a variety of internal structures, including radial fractures, amygdales that may be concentrically arranged or radially elongated, or large empty cavities which are formed by shrinkage of the lava on crystallization or leaking of lava from the pillow. Because of the molding prior to solidification, pillows generally fit together rather closely (Fig. 4.10). They may, however, be surrounded by fragments of pillow selvage, sedimentary material, or later cavity-filling minerals, such as chalcedony, quartz, calcite, and zeolites (Fig. 4.9).

Flood-basalt magmas contain dissolved gases (predominantly H_2O, CO_2, and S), which form *vesicles* upon eruption. Flows, especially thick ones, cool sufficiently slowly that most of the vesicles are able to float and form vesicular flow tops, leaving the main part of flows massive and nonvesicular. Some thick flows have multiple vesicular zones, which are interpreted to have formed from bubbles rising from successive pulses of magma that inflated the flow (Self *et al*., 1996). Vesicles can be quenched in the lower chilled margin of these flows. These are often pipe-like, and when filled with later minerals that are often white (e.g. calcite, quartz, zeolites), they resemble an old-fashioned clay-pipe stem; they are consequently called *pipe-stem vesicles* or *pipe-stem amygdales*. They never occur in the glassy lower contact of a flow but only start where the flow had time to crystallize. They are, therefore, formed from gas that exsolves as the lava crystallizes. When anhydrous minerals such as olivine, pyroxene, and plagioclase crystallize from basalt, whatever gases are present increase their concentration in the residual liquid, which becomes supersaturated and bubbles must form. Once the first bubbles nucleate, exsolving gas diffuses to them and as the solidification front advances into the flow, the bubbles grow as tubes that are perpendicular to the solidification front (Philpotts and Lewis, 1987). Similar tubes of air can be found in ice cubes formed in your refrigerator, except these tubes are radial, reflecting the concentric solidification of the ice cube. If lava is still moving as bubbles grow, the bubbles grow with a tilt in the direction of flow, forming a useful flow-direction indicator (Fig. 4.11).

Fig. 4.10 Vertically dipping Precambrian pillow lavas, Ungava, northern Quebec. These are part of a thick sequence of ocean-floor basalts that were obducted onto the continent. The convex upper surface of the pillows and the tail on their lower side indicate that the original top of these lavas was to the right.

Pipe-stem vesicles rarely extend into a flow for more than a few decimeters, but above this vesicular zone, the massive lava commonly contains vertical pipes of vesicular basalt, which are typically several centimeters in diameter, are spaced ~0.5 m apart, and extend vertically for several meters, These are referred to as *vesicle cylinders* (Fig. 4.12). Unlike pipe-stem vesicles, which are commonly tilted by the flow of the lava, vesicle cylinders are almost always vertical and therefore must have formed after the lava stopped moving. In addition to being highly vesicular, these pipes are composed of basalt that is richer in Fe, Ti, Na, K, and P than the host basalt (Goff, 1996). These elements, which do not enter the early crystallizing minerals, are concentrated in the residual liquid to a level that indicates approximately one-third crystallization of the host basalt (see Problem 10.25 for the comparative compositions of cylinder and host basalt). The pipes are formed as this residual liquid is expelled upward by the compacting crystal mush. In the upper part of thin flows, the vesicle cylinders connect with horizontal sheets of vesicular basalt. These sheets commonly develop diapirs on their upper surface, which rise into the overlying basalt and are often capped by a large gas cavity, which may be filled with secondary minerals (Fig. 4.12(B)). Many of the beautiful amethyst "cathedral" geodes seen in mineral shops were formed in this way in the 133-million-year-old Parana flood basalts of southern Brazil, Paraguay, and Uruguay, which were formed during the breakup of Gondwanaland (Section 15.4). Thick flows also contain horizontal vesicular sheets, but the connection with the vesicle cylinders in the base of the flow can often not be made.

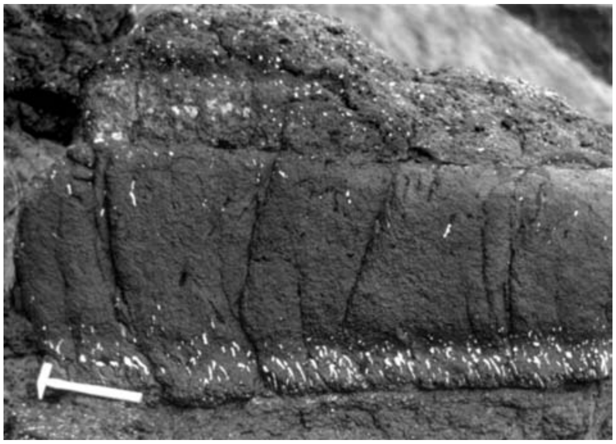


Fig. 4.11 Meter-thick basalt flow with amygdaloidal lower and upper zones and a massive core, County Antrim, Northern Ireland. Amygdales at the base form tilted pipes, indicating that the lava flowed from right to left.

4.3 CENTRAL VOLCANOES

Central volcanoes include all those structures formed by volcanic processes associated with a main central vent or magma chamber (Fig. 4.1). They are classified on the basis of their form and type of volcanic activity (MacDonald, 1972). The forms are determined largely by the type of activity, which in turn is determined almost entirely by the composition of magma involved. Basalt is very fluid and flows rapidly, even on gentle slopes, to form what are known as *shield volcanoes*. Magmas with higher silica contents are more viscous and cannot flow as far from central vents. In addition, such magmas are more explosive and the resulting volcanoes, which are composed of both lavas and fragmental material, are referred to as *composite*. Silica-rich magmas, such as rhyolitic ones, are so viscous that they do not generally flow far from their vent; instead, they develop large blisters or *domes* above the vent. Explosive activity is so common with these silica-rich magmas that the bulk of eruptive material may be in the form of extensive fragmental deposits (*tephra*).

Shield volcanoes are built almost entirely from basaltic flows, which emanate from summit craters or from fissures on the flanks of the volcanoes. The low-viscosity basalt moves rapidly down slopes and accumulates on the outer parts of the shield. For example, some of the Hawaiian flows have traveled more than 50 km and yet may be only 5 m thick. Slopes rarely exceed 10° and are commonly convex upward, giving rise to a shield-like form in cross section (Fig. 4.13(B)). But the symmetry of many of these volcanoes

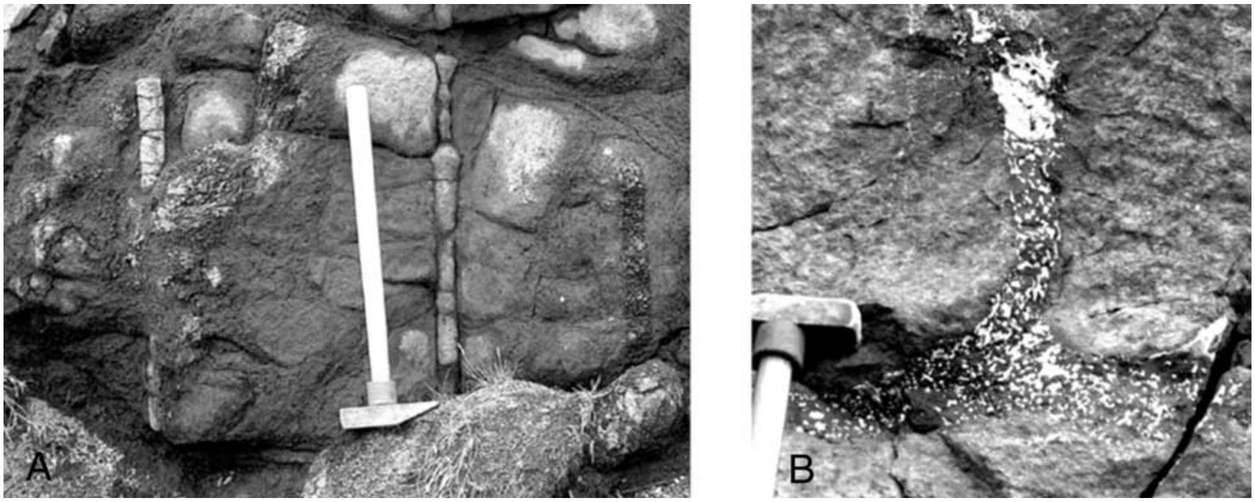


Fig. 4.12 (**A**) Vesicle cylinders in the Jurassic North Mountain flood basalt in the Bay of Fundy, Nova Scotia. The base of the flow is just beneath the tufts of grass at the bottom of the photograph. On horizontal surfaces, the cylinders are seen to have a regular spacing of ~0.5 m. (**B**) The vesicle cylinders connect with horizontal sheets of vesicular basalt in the upper part of the flow. Most of these are less than 10 cm thick. They commonly have diapiric structures on their upper surface, at the top of which can be large vesicles or amygdales.

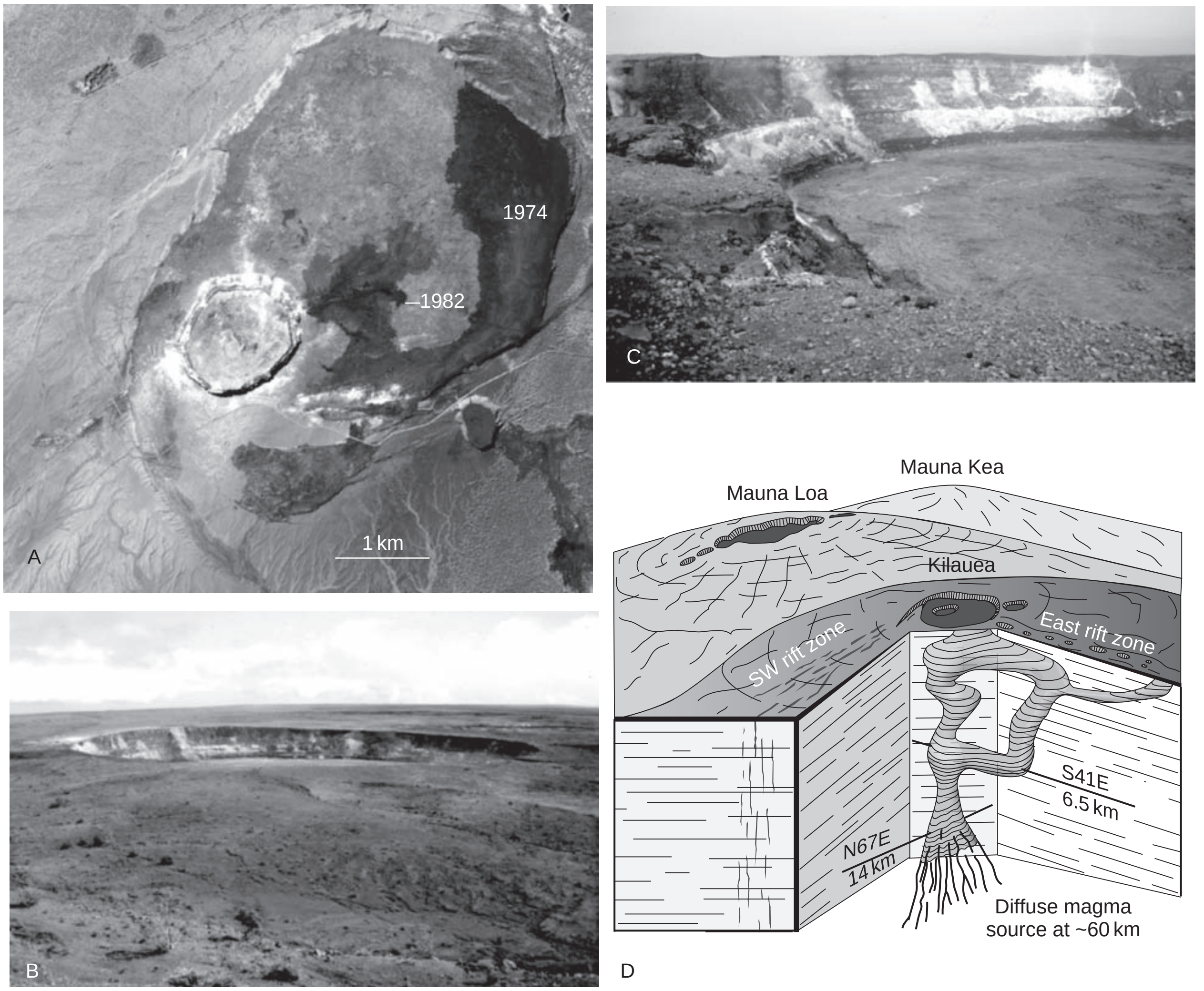


Fig. 4.13 (**A**) Satellite image of the 165-m-deep Kilauea caldera and the smaller Halemaumau Crater, at the summit of Kilauea volcano, Hawaii. Image is from NASA's Visible Earth catalog of images (http://visibleearth.nasa.gov). (**B**) Halemaumau Crater in the floor of the Kilauea caldera. Note the gentle convex horizon, caused by the typical shape of a shield volcano. Large blocks of rock littering the floor of the Kilauea caldera were blown from Halemaumau in the first half of 1924 when subsidence of the crater floor allowed groundwater to gain access to hot rock below. These explosions increased the crater diameter from 425 m to its present 900 m. (**C**) Following the 1924 explosions, the floor of Halemaumau was 410 m below the caldera floor, but since then, seven periods of flooding with basalt have raised the level to its present 85 m depth. The prominent bench on the crater wall was formed during a high stand of lava in 1968. The present crater floor is covered by basalt erupted in 1974. The caldera walls are covered with yellow and white sulfur and sulfate minerals deposited by fumerolic activity. The flower wreath, or *lei* in Hawaiian, in the foreground is an offering to Pele, the goddess of fire, who is believed to make Kilauea her most recent residence. (**D**) Schematic view of the Island of Hawaii, with the magma conduit beneath Kilauea being exposed in the cutaway. (Drawn from sections presented by Ryan *et al.*, 1981. See Okubo *et al.*, 1997, for more recent seismic tomography.)

is destroyed by the extrusion of large amounts of material from fissures on the flanks of the volcano. Also, the summits of many of these volcanoes have been down-faulted along concentric fractures to produce *calderas* (Figs. 4.13, 4.24). These form when the roof of a near-surface magma chamber collapses due to magma migrating either into the flanks of the volcano or up along the faults bounding the caldera. Subsequent eruptions are commonly localized along the faults bounding the caldera (1974 flow in Kilauea, Fig. 4.13 (A)), and active lava lakes are often formed within the caldera itself (Fig. 4.14). Active lava lakes are connected to the underlying magmatic plumbing, which replenishes the heat needed for convection. This differs from large ponds of lava such as that which filled the Kilauea Iki crater in 1959 (Fig. 2.2), which have no additional input of energy after the initial filling.

The Hawaiian volcanoes are excellent examples of shield volcanoes. Mauna Loa forms a giant shield, rising over 8 km above the ocean floor (see Problem 3.3), with only the upper 4.2 km exposed above sea level (Fig. 4.13). A caldera, more than 200 m deep and 15 km in circumference, crowns its summit. During recent years, careful measurements with seismographs and tiltmeters on the Island of Hawaii have revealed much about the behavior of such volcanoes, in

Fig. 4.14 Sixty-meter-wide lava lake at summit of Erte'ale shield volcano in the Afar region of Ethiopia in October, 2005. This lava lake has been active since 1967 and possibly since the beginning of the twentieth century. The magma rises on the left where the large bubble of gas is bursting. As it spreads from this point, fractures in the rapidly formed crust continually open and expose molten basalt. Crust traverses the lake in ~15 minutes and sinks at the crater wall on the right. The author, Anthony Philpotts, in the foreground.

particular of Kilauea, a volcano situated on the eastern flank of Mauna Loa (Richter *et al*., 1970; MacDonald, 1972; Ryan *et al*., 1981; Tilling and Dvorak, 1993; Okubo *et al*., 1997).

The conduit through which magma ascends beneath Kilauea has been mapped by locating seismic hypocenters that are associated with the fracturing of rock produced by excess pressures in the buoyantly rising magma (Ryan *et al*., 1981). More recently, three-dimensional seismic P-wave traveltime tomography has been used to image the magma conduits (Okubo *et al*., 1997). The seismicity defines a southward-dipping zone that extends to a depth of 60 km. The zone narrows and becomes better defined toward the top of the mantle. At a depth of 14 km, it is elliptical in plan, with maximum and minimum dimensions of 3.8 km and 2.1 km, respectively (Fig. 4.13(D)). This zone is not filled with liquid but must consist, instead, of a large percentage of solid rock that is capable of being fractured by the magma pressure. The elliptical shape of the conduit, which has its long dimension striking N 67 ° E, reflects stress inhomogeneities in the upper mantle resulting from plate tectonic motions. On rising through the oceanic crust, the cross-sectional area of the conduit decreases slightly, and between 8.8 and 6.5 km the major dimension of the elliptical section rotates in a clockwise direction (in map view), becoming essentially parallel with the strike of the east rift zone of Kilauea (S 41 ° E). From a depth of 6.5 to 5.7 km, seismic activity is largely absent,

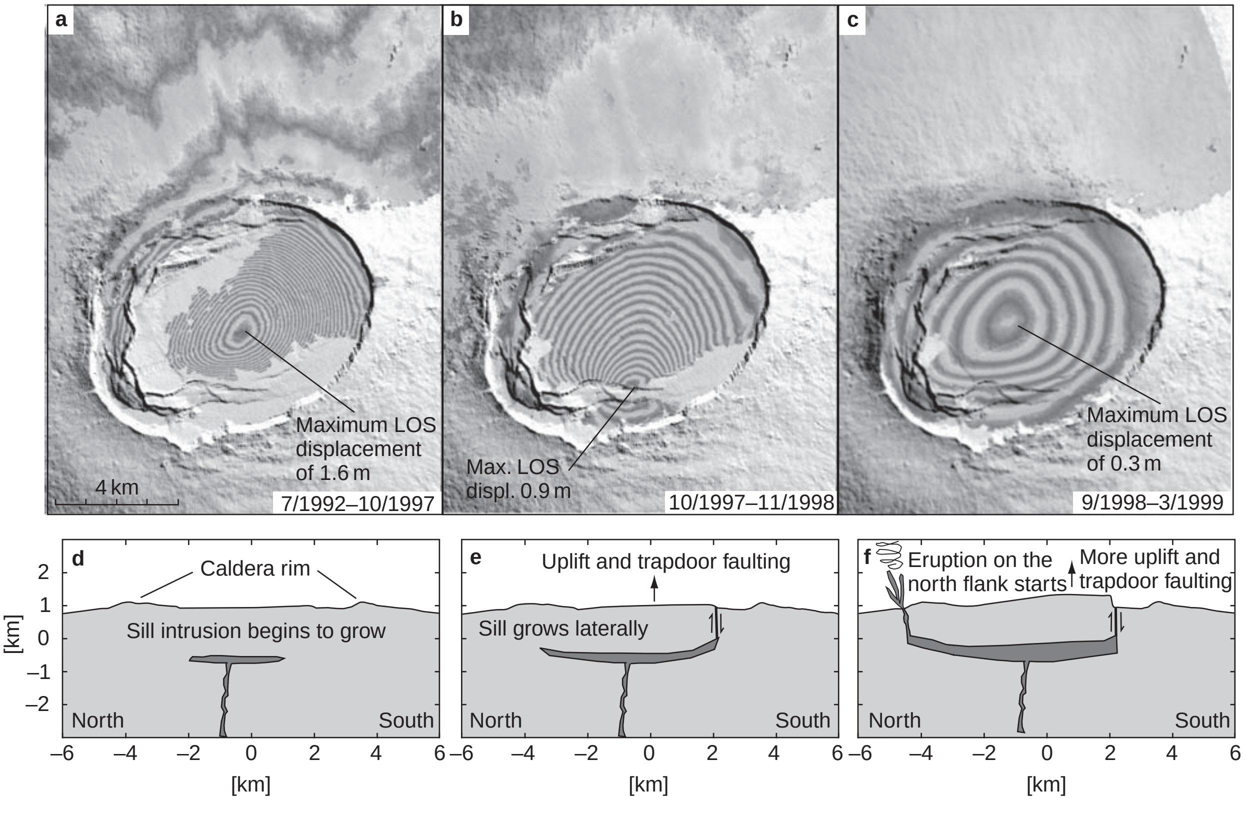


Fig. 4.15 Deformation of the caldera floor of Sierra Negra volcano on Isabela Island, Galapagos, as evidenced from a series of satellite radar interferograms (**a** to **c**) for the three periods indicated. Each color cycle (see www.cambridge.org/philpotts), shown as concentric gray lines, represents 5 cm line of sight (LOS) displacement up toward the satellite. The growth of a sill in (**d**) to (**f**) is modeled to account for the deformation. See text for discussion. (From Jónsson *et al*., 2005. Published with permission of Elsevier Press.)

indicating that the ratio of magma to rock in this zone is sufficiently large to allow for easy access of ascending magma. An offshoot from the main conduit at this level feeds a conduit beneath the east rift zone. The aseismic zone is interpreted to be the base of a magma chamber. Tomography indicates a localized high-velocity zone between 5 and 11 km beneath Kilauea, which is interpreted to be an accumulation of olivine from this magma chamber. The top of the chamber, which is at a depth of about 1.5 km, is a seismically active zone, presumably because of intense dike activity. The filling of the magma chamber inflates the volcano by as much as several meters, with subsidence occurring only when there is an eruption or magma moves laterally along the conduit into the east rift zone. The flow of magma into fractures in the roof of the chamber develops stresses at the tip of the cracks, which causes further fracturing. The alternating fracturing and injection produces a pulsating seismic wave known as *harmonic tremor*. These waves cease abruptly once magma breaks through to the Earth's surface.

Despite an apparent rather constant annual flux of about 0.1 km^3 of magma into the base of Kilauea, the exact timings, locations, and chemical compositions of eruptions are highly variable. This is to be expected, however, in view of the rather complex plumbing system beneath the volcano. New batches of magma may rise directly to the surface, they may mix with earlier batches that have cooled and changed composition, or they may move laterally into the east rift conduit to be extruded at a later date. Complexities in older shield volcanoes indicate that Kilauea is a perfectly normal example of this type of volcano. The most recent eruption of Kilauea, which began in January 1983 at Puʻu ʻOʻō in the east rift zone, appears to be fed by magma that is flowing directly into the east rift zone and bypassing the summit reservoir (Garcia *et al*., 2000).

Although the timing and location of Kilauean eruptions is variable, investigations by the U.S. Geological Survey (Klein, 1984) have made it possible to forecast eruptions of this volcano at the 99.9% confidence level. This is done by monitoring surface tilt, rate of tilting, seismicity, and 14-day Earth tidal cycles. Of these parameters, the tilt, which is measured with a simple water tube, gives the earliest warnings and is effective up to 30 days prior to an eruption. The frequency of earthquakes, on the other hand, is effective only within 10 days of an eruption. The apparent constant flux of liquid into the magma chamber results in the volcano being on the verge of eruption at all times. Consequently, minor disturbances, such as a maximum in the 14-day Earth tide cycle, can be sufficient to trigger eruption.

Land-based monitoring of active volcanoes is now augmented by satellite *Interferometric Synthetic Aperture Radar* (*InSAR*). By comparing satellite radar images of a volcano taken at different times, changes in elevation can be shown as interference patterns that allow ground deformation to be monitored at the centimeter scale. This technique has tremendous potential, because even remote volcanoes that have no ground monitoring devices can still be monitored for potential activity. For a description of the technique and application to the monitoring of volcanoes refer to the USGS Volcano Hazards Program web site (http://volcanoes.usgs.gov).

Figure 4.15 shows a series of interferograms of the caldera on the Sierra Negra Volcano on Isla Izabela in the Galápagos (Jónsson *et al*., 2005). Each color cycle (shades of gray in Fig. 4.15) represents 5 cm line-of-sight (LOS) displacement. Between 1992 and 1997, the floor of the caldera domed upward ~1.6 m symmetrically about its center. Between 1997 and November 1998, an additional 0.9 m of uplift occurred but the maximum uplift moved to a fault on the south side of the caldera. This resulted in faulting, with the caldera floor behaving like a trapdoor. Between 1998 and 1999 the caldera floor was raised symmetrically an additional 0.3 m. Jónsson *et al*. (2005) successfully modeled this deformation with a sill or laccolith intruding at a depth of ~2 km beneath the caldera floor. The trapdoor fault on the south side of the caldera prevented the sill from extending farther to the south. They predicted that if inflation continued, the sill might erupt to the north. Inflation did continue, with the total amount since 1992 reaching 4.9 m by October 2005, when an eruption started from a fissure along the north side of the caldera (Chadwick *et al*., 2006).

Fig. 4.16 (**A**) The large slump scarp on the south side of the Island of Hawaii is covered by multiple pahoehoe and aa flows from the Pu’u ‘O’ō vent in the east rift zone of Kilauea volcano. Here the smooth ropy surface of a still hot, steaming pahoehoe flow covers an older aa flow with typical rough clinkery surface. (**B**) Detail of the pahoehoe flow overlying the aa flow.

Fig. 4.17 Freshly extruded pahoehoe lava from the May 1987 Pu'u 'O'ō eruption on Kilauea's east rift zone, Hawaii. Various scales of ropy structures indicate variable temperatures and rates of extrusion. Note that lava that rose to fill the crack in the brittle crust on the right was cooler and more viscous than the earlier lava and solidified before it could spill onto the surface of the flow.

In addition to InSAR, infrared monitoring of Earth by Moderate Resolution Imaging Spectroradiometers (MODIS) carried on NASA's Earth Observing System (EOS) satellites Terra and Aqua provide near-real-time (every 48 hours) 1-km pixel images of the entire Earth on which any hot spots, such as fresh volcanic eruptions or forest fires, can be seen. The Hawai'i Institute of Geophysics and Planetology has created an automated system that maps the global distribution of these hot spots, which can be seen on their web site at http://modis.higp.hawaii.edu. The Ozone Monitoring Instrument (OMI) carried on NASA's EOS/Aura satellite provides sensitive daily global measurements of SO (http://aura.gsfc.nasa.gov). While designed to monitor the release of anthropogenic SO_2, the data provide a ready means of detecting new volcanic eruptions, which typically release large amounts of SO_2.

Lava flows on shield volcanoes are generally of two types, pahoehoe and aa (Fig. 4.16). *Pahoehoe* flows have smooth surfaces, which may be wrinkled into ropy-looking masses by the continued flow of hot fluid lava beneath the partly solidified crust. The hot lava issues forth as tongues along the leading edge of the flow. Cracks formed in the crust are rapidly healed by lava welling up and pouring out onto the surface (Fig. 4.17). The fluid cores of pahoehoe flows can drain completely if the supply of fresh lava is cut off, leaving

Fig. 4.18 Lava draining from the interior of a lava flow can leave a lava tube, the sides of which commonly have benches marking periods during which the draining lava remained at a particular level for some considerable time, and stalactite-like dribbles of lava (lavacicles) hang from the roof. Pu'uhonua o Hōnaunau National Historic Park, Hawaii.

Fig. 4.19 Large blister formed on the surface of a pahoehoe flow caused by pressure of lava in the flow's interior; Kamoamoa, Hawaii.

behind empty *lava tubes* which can be many meters in diameter and of considerable length (Fig. 4.18). If these tubes become blocked at their lower end, pressure can build up in the tube and eventually dome and burst the surface of the flow (Fig. 4.19). In some cases, the interior of a flow can drain away almost completely, lowering the surface of the flow as lava issues from the distal ends. This is evident in many Hawaiian flows where tree molds record the former height of the flow as it moved through a forest (Fig. 4.20). In contrast, the surface of an *aa* flow is broken into a mass of clinkery fragments, each of which has an extremely rough surface, from which project many small spines of glass. This rubble is carried along on the surface of the flow until it reaches the leading edge, whereupon it tumbles down the steep front face and is overridden by the flow. This conveyor-type motion results in rubbly material occurring in the upper and lower parts of aa flows.

The decrease in confining pressure on magma as it moves toward the surface causes dissolved gases to come out of solution and form bubbles, which are known as *vesicles* (see Section 3.4). These may later be filled with such minerals as zeolites, chlorite, calcite, or forms of silica, which are deposited from late magmatic solutions or circulating groundwater. These filled vesicles are called *amygdales* because of their resemblance to almonds, the Greek word for which is *amygdalos*. The diminutive term *amygdule* is also used, although commonly with little regard to absolute size. Rocks containing amygdales are said to be *amygdaloidal*. The minerals of amygdales usually reflect the composition of the host rock. For example, those in nepheline-bearing basalts commonly contain zeolites, whereas those in more silica-rich basalts may contain chalcedony or quartz.

Both pahoehoe and aa flows contain abundant vesicles or amygdales. These are mostly concentrated in the upper parts of flows where the buoyantly rising bubbles encounter cooler, more viscous lava. The central part of flows generally contains far fewer vesicles and in thick flows may be completely devoid of them. This is due to the rapid rise of bubbles through basaltic magma. For example, a 5-mm-diameter bubble rises $1.3\,\mathrm{m\,h^{-1}}$ in a lava of density $2.75\,\mathrm{Mg\,m^{-3}}$ and viscosity of 100 Pa s. Pipe-stem vesicles (Fig. 4.11) may be present in the base of a flow.

Compositionally, pahoehoe and aa flows are very similar. In fact, a pahoehoe flow may change into an aa one as it flows away from its source. One striking difference between them is in the form of their vesicles (MacDonald, 1972). In pahoehoe, vesicles are mostly spherical, indicating that the bubbles were growing or expanding at the time of solidification of the lava. But vesicles in aa have highly irregular shapes, which indicate that growth of the bubbles had ended prior to solidification. This implies that pahoehoe lavas have had less time to degas and are closer to their source. This is also supported by the higher temperatures of pahoehoe flows. The rate of flow, however, is an additional factor, for ropy lava, if moved too rapidly, can develop a clinkery surface.

Lava that comes in contact with water is rapidly quenched to a glass that may fragment into sand-size particles. Rocks formed from this material are called *hyaloclastites* from the Greek for glass (*hyalo-*) and to break (*clast*). They differ from volcanic ash in that the fragments are equant and generally lack the cuspate shape of the glass shards formed from the lava trapped between expanding bubbles of gas (Fig. 4.41) that are formed when lava is disrupted during eruption from a vent.

A number of minor rock types containing more silica than basalt can occur on shield volcanoes. These lavas are derived from basalt by various processes of differentiation operating beneath the volcano. Because of their higher silica content, and consequently higher viscosity, they may form small, steep-sided cones on the main shield. Similarly shaped cones may also be formed from cinders ejected during periods of more violent disruptive eruption, especially where groundwater or seawater gains access to an active vent and

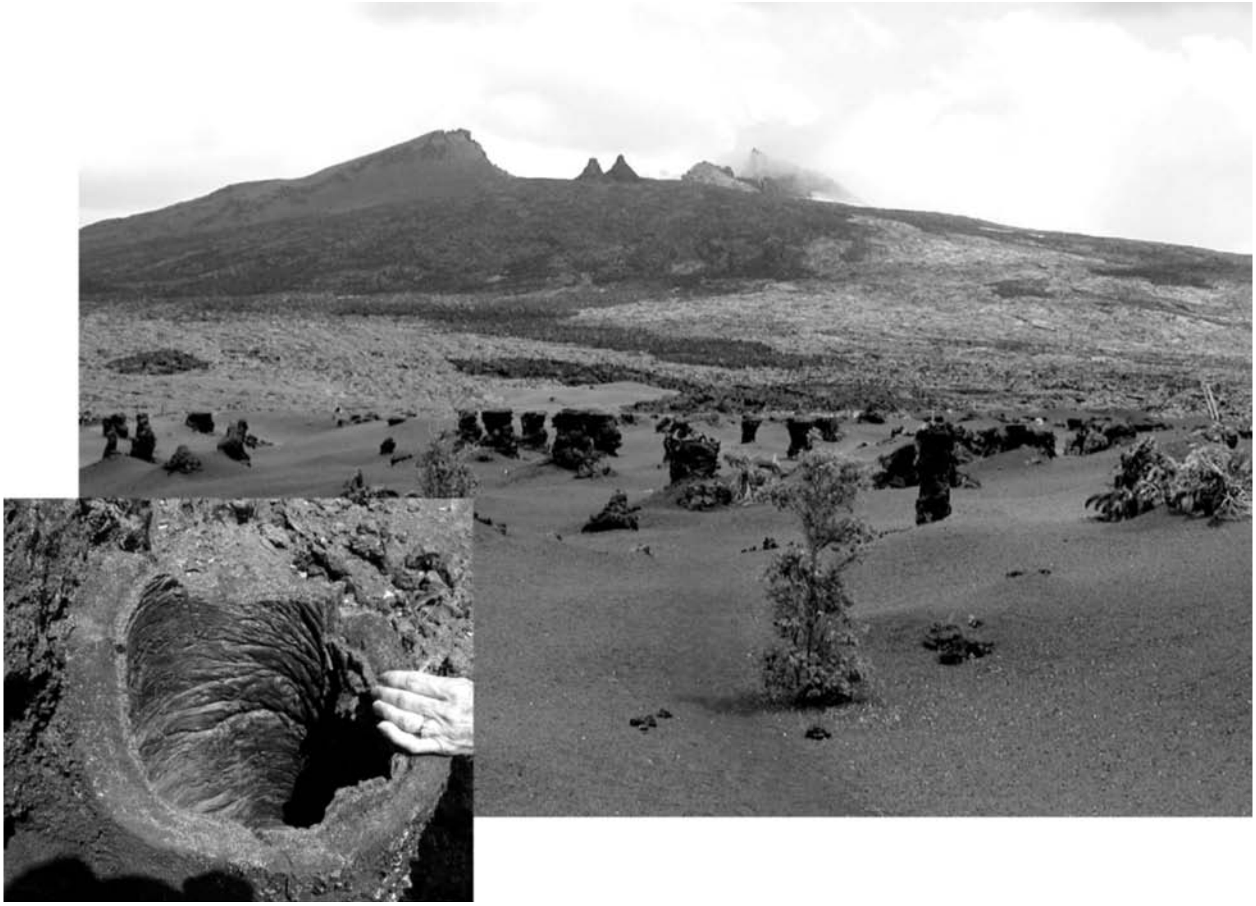


Fig. 4.20 Pu'u 'O'ō, Hawaii, initially formed as a large cinder cone on the east rift zone of Kilauea from which large amounts of ash were deposited on the surroundings, seen in the foreground around tree molds formed when an earlier pahoehoe flow moved through a forest; this flow eventually drained away, but the molds (see inset) record the former height of this flow. Later activity at Pu'u 'O'ō developed a lava lake from which both pahoehoe and aa flows have descended the slopes and covered much of the earlier ash. Two spatter cones can be seen near the summit vent.

Fig. 4.21 Cinder cones in the summit basin of Haleakala, Maui. Inset shows a bomb and cinders found on the floor of the basin.

is rapidly converted to steam (Fig. 4.21). Blobs of lava known as *bombs* may be ejected from these vents and because they are often still partially molten on landing, they are commonly flattened and resemble cow pancakes (inset in Fig. 4.21). If the liberation of gas is not intense enough to produce cinders, large blobs of magma may be ejected a short distance from the vent. This material is still molten when it lands and builds a steep-sided *spatter cone* around the vent (Fig. 4.22). Despite the common occurrence of these more explosive features on shield volcanoes, the amount of fragmental material in shield volcanoes is relatively small compared with the amount found in composite volcanoes.

Fig. 4.22 Spatter cones on Bartholomew Island, Galapagos. In the background, a large aa flow can be seen extending out into the sea from the shield volcano on Santiago Island.

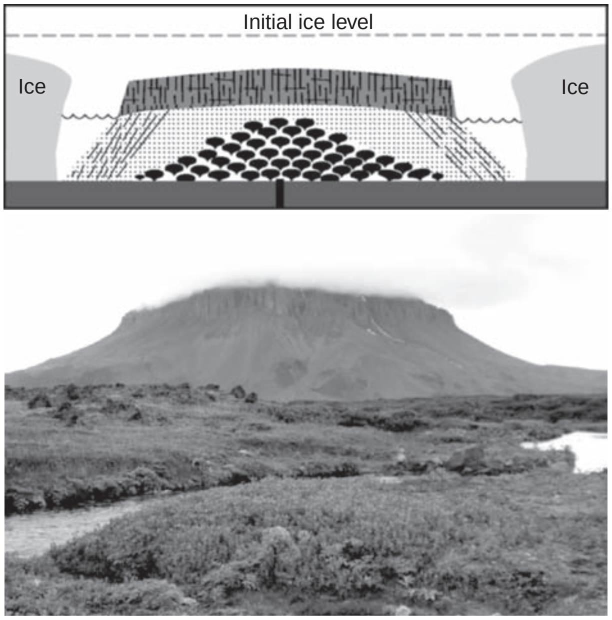


Fig. 4.23 The 1682-m-high Herdubreid Mountain, Iceland. This flat-topped volcano is an example of a tuya, which forms beneath glacial ice. The cross section above gives a tuya's typical stratigraphy. When eruption begins, ice melts to form a subglacial lake into which the lava flows to form pillows. As the pillow pile grows higher, hyalocastites start forming. Eventually when the surface of the glacier is breached, subaerial eruptions produce massive lavas that form a resistant flat-topped cap to the volcano.

Lava can erupt beneath glacial ice, which results in the construction of a characteristic volcanic form known as a *tuya* (Fig. 4.23). One of the most active volcanoes in Iceland is Grímsvötn, which lies beneath Vatnajökull, Europe's largest glacier. Although the Grímsvötn volcano remains hidden beneath ice, we can observe its effects on the glacier and surmise what is happening beneath the ice. In addition, with the shrinkage of the ice sheets since the close of the Pleistocene, many flat-topped volcanoes have been exposed in Iceland that were clearly formed beneath the ice. Studies of the structure of these volcanoes combined with the observations from Grímsvötn have provided understanding of this type of ice-contact volcanism (Hoskuldsson and Sparks, 1997).

When eruption starts, melting of the ice results in the formation of a subglacial lake. InSAR satellite images of the surface of the glacier over Grímsvötn during the 1996 eruption provide a detailed record of the deformation of the surface of the glacier (Gudmundsson *et al*., 2002). If water in the subglacial lake is not able to drain away, the lake grows larger, but eventually it floats the glacier off its base and the meltwater escapes, forming devastating floods known as *jökullaups*. The 1996 jökullaups from Grímsvötn destroyed the main highway and bridges along the south coast of Iceland. When lava erupts into the subglacial lake, it forms pillows (Fig. 4.23). As melting continues and the pile of pillows grows higher, the level of the volcanic vent is raised and much of the erupted lava splinters into small glassy particles forming hyaloclastite. Some of this fragmental material forms steeply sloping deposits around the earlier pile of pillows. Eventually, melting extends the lake to the surface of the glacier, where boiling can get rid of the water; alternatively, the lake may drain away through subglacial channels. In either case, when eruption eventually occurs subaerially, massive lava flows form, which create the flat top to the tuya (Fig. 4.23).

A number of flat-topped volcanic structures on Mars have been interpreted to be tuyas and to have formed when the Martian polar caps were more extensive (Allen, 1979). Even the largest volcano in the solar system, Olympus Mons (Fig. 4.24), may owe its shape in part to growth beneath an ice sheet. Olympus Mons is three times higher than Mount Everest, towering 27 km above the surrounding plain. It forms a 500-km-wide shield, which is surrounded at its base by a remarkable ~10-km-high escarpment. Some voluminous aureole deposits extend out from the escarpment for another ~500 km. At the volcano's summit is a 70-km-wide caldera, which has several smaller calderas within it. Helgason (1999) has proposed that early in the growth of Olympus Mons, an ice sheet formed on the volcano, possibly due to its increased elevation. Further eruptions occurred beneath the ice cap, causing a greater vertical accumulation of lava, in the manner that tuyas form on Earth. During this period of growth, pillow lavas and hyaloclastites would have formed. Finally, the volcano grew high enough to rise above the ice and form subaerial lava flows. Many of these lavas flowed down the slopes over earlier-formed less competent material, much of which may have been mixed with ice. The result of this loading was to cause slumping around the already steep sides of the volcano with the development of the prominent escarpment. The aureole deposits that extend out from the escarpment are probably mixtures of volcanics and melted ice.

Composite volcanoes (Fig. 4.25) are built of both lava flows and fragmental material ejected from vents during periods of explosive activity. These materials are interlayered and, because of their markedly different resistance to weathering, produce very stratified-looking rocks (Fig. 4.26). Consequently, the name *strato volcano* is also used for this

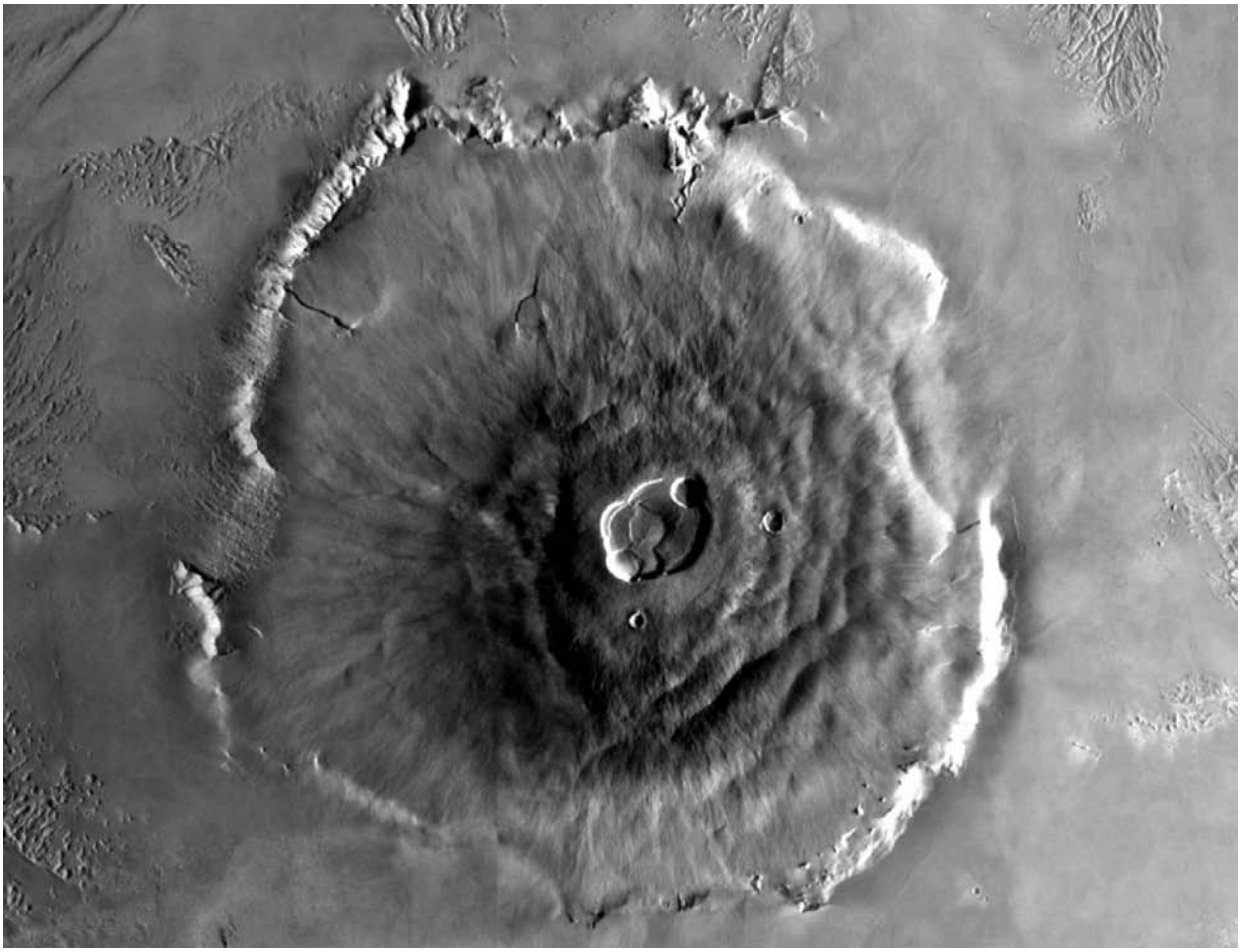


Fig. 4.24 Olympus Mons on Mars is the solar system's largest volcano. It is 500 km in diameter, 27 km high, and has a 70-km-wide caldera at its summit. It is surrounded by a 10-km-high escarpment, which has been interpreted to indicate the volcano may have erupted beneath an ice cap early in its growth (see text for discussion). The image, which is from the NASA's NSSDC Photo Gallery, can be accessed at http://nssdc.gsfc.nasa.gov/image/planetary/mars/olympus_mons.jpg.

type of volcano. They are composed predominantly of andesite, but commonly have early basaltic phases, and toward the end of their life erupt more silica-rich magmas. They have the typical form that most people associate with volcanoes; that is, concave-upward slopes ranging from 10° to 36° with a single crater at the summit (Fig. 4.25). Mount Fuji in Japan is perhaps the best known of these volcanoes, but there are many others, especially in orogenic regions, for example Orsono in Chile (Fig. 4.25), Mayon in the Philippines, Vesuvius in Italy, and Shasta, Rainier, Hood, and St. Helens in the Cascade Range of the northwestern United States (Fig. 4.2). These volcanoes owe their symmetrical shapes to eruptions from a single summit vent. Although this is common, composite volcanoes can erupt from several vents, or the feeder pipe may shift with time so that later cones may be offset from the earlier center of activity (Fig. 4.27). Mount Shastina on the western flank of Shasta is an example of such a younger vent. In still others, a volcanic cone may be situated within the remains of an earlier one that has been largely destroyed by violent explosive activity. This is the case, for example, with Vesuvius, which is located in the center of what remains of the older Monte Somma volcano.

Fig. 4.25 Orsono composite volcano in the Andes of Chile. This volcano last erupted in 1890 and was erupting when visited by Charles Darwin in 1834.

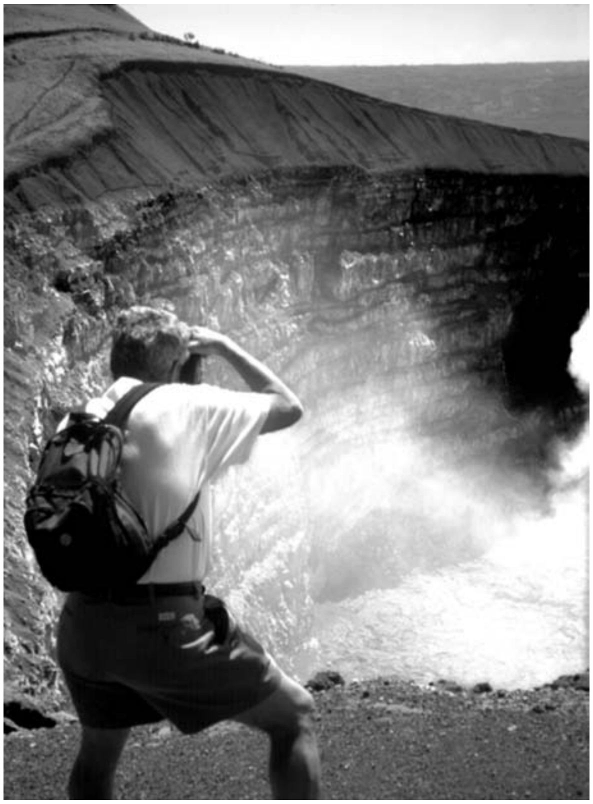


Fig. 4.26 Alternating layers of ash and lava in the crater wall of the Masaya volcano, Nicaragua. The last eruption can be seen to have deposited a thick layer of ash, which eventually will be eroded away if not covered with another lava flow.

Fig. 4.27 Gabillema, a 1459-m-high composite or strato volcano in the rift valley of Ethiopa. Thick viscous lavas can be seen to have descended from the summit vent, but a younger eruption has developed an extremely viscous rhyolite dome and short flow on the near flank of the volcano.

Fig. 4.28 Blocky andesitic lava flow that erupted from the cone in the upper left of the photograph, McKenzie Pass, High Cascades, Oregon.

The increase in slope toward the summit of composite volcanoes may be gradual, but more often there is a relatively constant gentle slope on the lower parts of the volcano followed by rapid steepening near the summit. This morphology is the direct result of the greater accumulation of erupted material near the summit, especially of tephra, which consists of large, heavy fragments near the vent, grading to finer particles farther down the slopes. The outer flanks consist mainly of lava (Fig. 4.27), which commonly forms *blocky* flows that are somewhat akin to aa, but the blocks are more regular in shape and lack the extremely rough surfaces of the clinkery aa (Fig. 4.28). Lesser amounts of aa and pahoehoe may also occur. In addition, the lower slopes may be covered with mudflows of loose tephra washed down from above by heavy rainfalls. Indeed, if it were not for the interlayers of lava in these volcanoes, the steep slopes would soon be reduced by erosion. These mudflows, which are named *lahars* to distinguish them from ones not associated with volcanism, can carry in suspension blocks of material measuring up to many meters in diameter (Figs. 4.29 and 4.30). Lahars pose a serious environmental hazard, especially in areas of heavy rainfall. For example, mudflows were responsible for the complete destruction of the Colombian town of Armero and its more than 20 000 inhabitants when the Nevado del Ruiz volcano erupted in November 1985. Torrential rain and melting snow caused volcanic ash laid down by prehistoric eruptions to become saturated with water; the seismic disturbance produced by the eruption of Nevado del Ruiz was then sufficient to trigger the mud slides.

In general, the rocks of composite volcanoes are more silicic than those of shield volcanoes. Andesites and dacites (see Chapter 6) are the two most common rock types, but smaller amounts of basalt and rhyolite may occur. The magmas from which most of these rocks form are more viscous than basalt, and the lavas, therefore, cannot flow as far.

Moreover, the higher viscosities prevent gas from escaping easily from the magma. Consequently, most bubbles of gas formed during ascent remain in the magma until the time of eruption. The pressure decrease accompanying ascent brings about expansion of the gas and eventual disruption of the magma, which results in the periodic explosive activity characteristic of this type of volcano (Melnik and Sparks, 2005). For example, a bubble of gas in magma extruded onto the surface from a depth of 3 km expands approximately 1000 times in volume (see Problem 4.4). Of course, gas may escape from magma that remains stationary for a sufficient length of time, or, within a volcanic pipe, gas may rise and concentrate in the upper part, leaving behind relatively gas-free magma. This results in periods of explosive activity alternating with periods of lava eruption. Melnik and Sparks (2005) have shown through computer simulations that the interplay between gas exsolution, bubble growth, gas escape through the magma, decompression-induced crystallization, and possible Bingham rheology can lead to periodic variations in magma discharge rate and explosivity that are similar to recorded fluctuations at, for example, the Soufrière Hills on Montserrat (Sparks and Young, 2002).

The type of volcanic activity in which ash is ejected high into the atmosphere is known as *plinian* (Fig. 4.30), after the Roman historian Pliny the Younger, who described in considerable detail the AD 79 eruption of Vesuvius. The ash from this eruption not only buried the city of Pompeii (see Fig. 4.36) but was indirectly responsible for the death of Pliny the Elder,

Fig. 4.29 Volcanic mud flow, or lahar, on the flanks of Mount Hood in the Cascade Range of Oregon (Fig. 4.2).

Fig. 4.30 (**Above**) The catastrophic eruption of Mount St. Helens on May 18, 1980, removed the top and much of the north face (to the left) of the volcano and filled the Toutle River valley with a lahar, which has since been cut into by the river. (**Left**) A 3.5-km-high plinian-type plume of volcanic ash venting from a 200-m-high dome of dacite, which has grown in the crater since the 1980 eruption of Mount St. Helens. (Photograph taken by James Zollweg on June 9, 1982.)

who died of an apparent heart attack while escaping from the eruption with his nephew, Pliny the Younger.

The particles ejected from a volcano during explosive activity vary considerably in size and composition. Many consist of molten material, but others can be of igneous rocks formed during earlier periods of activity, while still others may be of foreign rocks ripped from the walls of the feeder pipe. All of this ejected material is collectively known as *tephra*, which when consolidated forms *pyroclastic* rocks (broken by fire). Ejected bodies composed of molten material may take on streamlined shapes while passing through air; these are known as *volcanic bombs* if their diameters are greater than 64 mm (Figs. 4.21 and 4.31). Smaller particles are called *lapilli*, and the fine material (<2 mm) is referred to as *ash*. Thin filaments of glass known as *Pele's hair* (Fig. 4.31) may also form when droplets of lava separate from their source. Rocks formed from the accumulation of larger fragments are known as *agglomerates*, whereas those formed from ash are known as *tuff*.

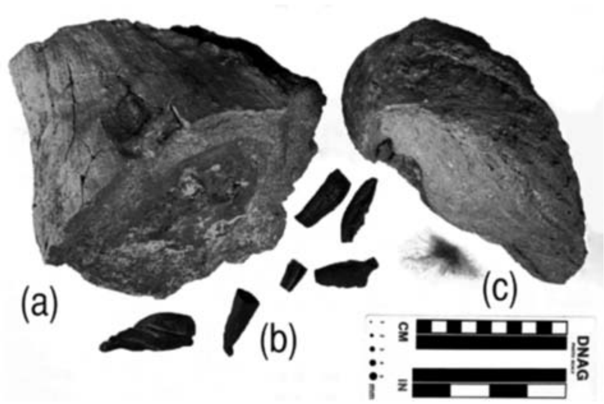


Fig. 4.31 Molten material ejected from volcanoes may take on various forms. Large bodies (>64 mm), known as bombs (**a**), commonly have tails with prominent lines and grooves formed when the particle separated from its source. The bomb shown on the left is broken, exposing concentric layers of vesicles within. Smaller particles (<64 mm, >2 mm), known as lapilli (**b**), are commonly teardrop-like in shape. Fine strands of basaltic glass (**c**) formed during the separation of lapilli and bombs are known as Pele's hair after the Hawaiian goddess of fire.

Even when composite volcanoes are not erupting, fumarolic activity commonly continues long after eruptions have ceased. Gas emitted from *fumaroles*, also known as *solfataras*, is composed largely of steam, most of which is formed from groundwater that has been heated by circulating through hot rock. Fumaroles emit many other gases, which can be derived from magma cooling at depth. These include CO_2, SO_2, H_2S, Cl_2, F_2, and H_2. The rates of gas emission and their composition have been used to predict renewed volcanic eruptions (McGee and Jefferson Sutton, 1994; Francis *et al*., 2000). Fumarolic gases react with water to form acids, which rapidly alter the rocks with which they come in contact. Fumaroles are consequently surrounded by brightly colored alteration haloes (Fig. 4.32). At depth, the escape of these volatile constituents can lead to the formation of hydrothermal ore deposits.

The proportion of lava to pyroclastic rocks in composite volcanoes varies, depending, to a large extent, on the composition of magmas involved. It is not uncommon for the magmas to become more silicic toward the final stages of the life of a volcano, with rhyolite being the last rock type to be erupted.

Fig. 4.32 (**A**) Hot volcanic gases deposited sulfur and altered the rocks surrounding this fumarole near the summit of Etna, Sicily. (**B**) Similar caustic solutions emanating from magma bodies produced this hydrothermal alteration around the rich copper deposits of Bisbee, Arizona.

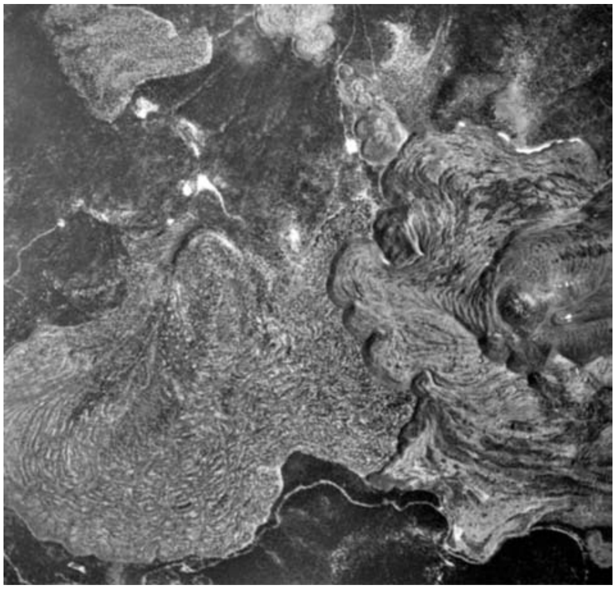


Fig. 4.33 Glass Mountain, Medicine Lake Highlands, California. A number of thick, viscous rhyolite flows emanate from this obsidian dome (center right). Glass Mountain and several smaller domes are localized along a northwesterly trending fracture. An earlier dome and associated flow occur in the center and lower left of the photograph. The apparent ropy surface of the flows is produced by transverse ridges that have a wavelength of many meters. (Photo courtesy of U.S. Forest Service.)

Fig. 4.34 The Big Obsidian Flow, Newberry Volcano, central Oregon (N in Fig. 4.2). The obsidian flow erupted from the margin of the caldera at the summit of the Newberry shield volcano, the largest volcano in the conterminous United States. A small rhyolite dome developed over the vent, and the surface of the 40-m-thick rhyolite flow has been wrinkled into large transverse waves as the lava descended the slope onto the floor of the caldera. Note the many small cinder cones in the background scattered on the broad flanks of the shield volcano. (Photograph courtesy of Oregon Department of Geology and Mineral Industries.)

This change is accompanied by increased explosive activity, and the highly viscous magma, rather than producing flows, may simply rise in the summit crater to form a *dome* (Fig. 4.30).

In an attempt to quantify the explosive power of eruptions, a *Volcanic Explosivity Index* (VEI) has been created (http://volcano.und.edu/vwdocs/eruption_scale.html). This scale ranges from a value of zero where only lava flows are erupted (e.g. Kilauea) to a value of 8 where "mega colossal" eruptions of thousands of cubic kilometers of ash are ejected. Although eruptions with an index of 8 are present in the geologic record, none has occurred in historic times. The 1883 eruption of Krakatoa produced tens of cubic kilometers of ash and is given a VEI of 6. Once the index reaches 4, the plume of ash rises to 25 km where it can have significant effects on Earth's climate.

The types of volcanic structures formed when only highly viscous magmas are erupted are quite different from those already discussed. High viscosities prevent lava flowing rapidly away from a vent; thus, as long as there are no violent explosions, lava accumulates over the vent as a *blister* or *dome* (Figs. 4.1 and 4.33). These generally have the form of a flattened hemisphere, but eruption from a fissure, rather than from a vent or eruption on a sloping surface, may result in the formation of an elongated body. The dome grows by magma intruding into its base; hence, the term *endogenous* dome is used to indicate that it was built from within. As a dome is inflated by the addition of new magma, its cooling surface cracks and develops a talus breccia around the body. If pressure within a dome is sufficiently great, the surface may burst and highly viscous lava extrude and flow down the steep sides of the dome. If a dome is built largely through the extrusion of lava, it is said to be *exogenous*. These domes, however, are less common than endogenous domes. Commonly, the pressure within a dome forces up *spines* of solid rock, which may be up to hundreds of meters in diameter and hundreds of meters high. These become highly fractured on cooling and are soon eroded away. A spectacular spine grew to a height of 350 m in nine months during the 1902 eruption of Mount Pelée, in Martinique.

Domes vary considerably in size from a few tens of meters to several kilometers in diameter. For example, the Puy de Sarcoui in Auvergne, France, is 400 m in diameter and 150 m high, whereas Lassen Peak in California is almost 2 km in diameter and 600 m high. Most domes are composed of rhyolite (obsidian), dacite, phonolite, or trachyte (see Chapter 6 for definitions).

When large quantities of viscous magma are erupted, especially on sloping surfaces, slowly moving lava flows can develop. These are thicker than most flows associated with the types of volcanism discussed previously, with thicknesses of several hundred meters not being uncommon. It will be recalled from Eq. (2.13) that the maximum velocity in a laminar flow is proportional to the square of the thickness and inversely proportional to the viscosity. Hence, with highly viscous lavas it is only the exceedingly thick flow that will travel a significant distance before cooling renders it too viscous to move. The surface of these flows is deformed into ridges, which are elongated transverse to the direction of flow. Although these wrinkles resemble the ropy structure of pahoehoe flows when viewed from a great distance (Fig. 4.34), the wavelength of the ridges is measured in

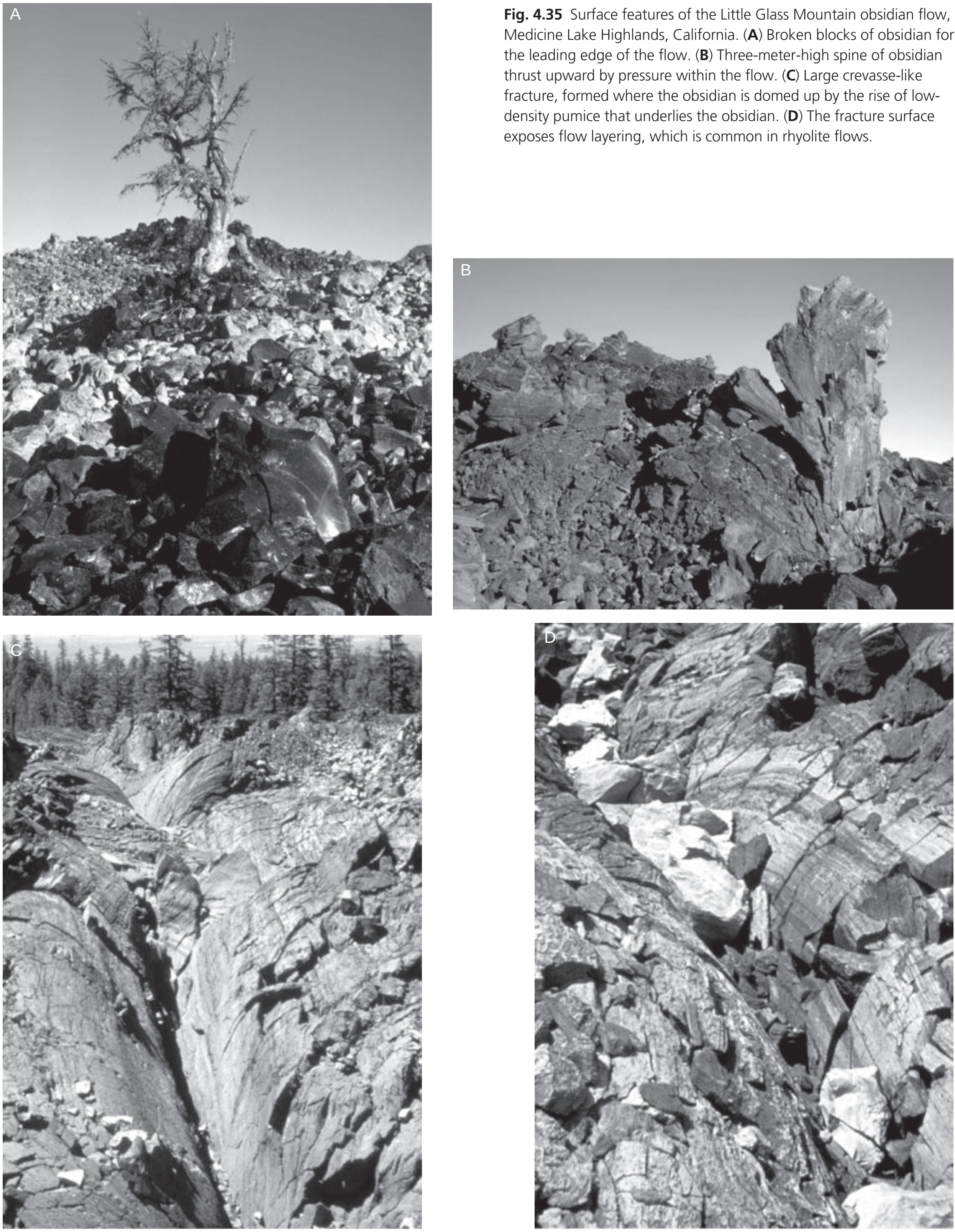


Fig. 4.35 Surface features of the Little Glass Mountain obsidian flow, Medicine Lake Highlands, California. (**A**) Broken blocks of obsidian form the leading edge of the flow. (**B**) Three-meter-high spine of obsidian thrust upward by pressure within the flow. (**C**) Large crevasse-like fracture, formed where the obsidian is domed up by the rise of low-density pumice that underlies the obsidian. (**D**) The fracture surface exposes flow layering, which is common in rhyolite flows.

meters to tens of meters and is not readily visible to an observer on the outcrop. Instead, at this scale the surface of the flow is extremely irregular and consists of broken blocks of obsidian, large crevasse-like gashes, and spines and slabs thrust up from beneath (Fig. 4.35).

Highly viscous flows have much more internal structure than do thinner less viscous ones (Fig. 4.35). An excellent example of such structure is provided by a rhyolite flow in Nevada that was studied in detail by Christiansen and Lipman (1966). This flow, which is up to 250 meters thick, can be traced for 11 km from its leading edge back to a feeder vent. The upper and lower parts of the flow are breccia, consisting of glassy and highly vesicular fragments (pumice) formed from the fracturing of the chilled crust by the continued movement of lava within the flow. The upper breccia is commonly cut by protrusions of viscous lava thrust up from the main central part of the flow, which consists of a strongly foliated rhyolite. The foliation is marked by textural variations resulting from different proportions of glass, spherulites, and microlites (see Chapter 12). When formed, the foliation was vertical and parallel to the walls of the feeder pipe, but it became highly distorted during flow on the surface. In the lower part of the flow, the foliation parallels the base of the flow, but toward the top, it becomes much steeper, with the dip direction being opposite to the flow direction. The foliation also is contorted into recumbent folds that are overturned in the direction of flow.

4.4 PYROCLASTIC DEPOSITS AND CALDERAS

The explosive activity associated with highly viscous magmas produces a variety of pyroclastic deposits that can be classified into three groups based on the way in which fragments are transported from the source to their final site of deposition (Sheridan, 1979; Fisher and Schminke, 1984).

Air-fall deposits are formed from particles that follow a simple ballistic trajectory, the direction and length of which is determined by the eruption velocity from the vent, the wind velocity, and the gravitational terminal velocity of the particle. These deposits, which are well sorted, blanket the surface evenly with little regard for topography. They form plumes, which thin and become finer grained downwind from the source.

Surge deposits are formed from clouds of ash that move out horizontally from a vent at hurricane velocities following a volcanic explosion. Similar rapidly moving clouds have been well documented in association with underground nuclear explosions. The passage of the cloud is short lived and thus only a fraction of the total pyroclastic deposit associated with an eruption is normally formed in this way. Bedding forms such as dunes and antidunes testify to the high velocities of emplacement, as does deposition in the lee of topographic obstructions. These clouds have a high gas-to-solid ratio. It was the switch from ash-fall to surge deposition that resulted in most of the deaths at Pompeii in AD 79 (Sigurdsson *et al*., 1982). The eruption began on August 24 with plinian-type activity. Pliny the Younger, who described the eruption in considerable detail, likened the shape of the erupting column of ash to the umbrella shape of the pines of Rome (Fig. 4.36(A)); it is because of his description that we refer to this type of eruption as *plinian*. At this early stage, the plume would have resembled the plume shown rising from Mount St. Helens in Figure 4.30. By the following morning, sufficient ash had fallen on Pompeii to collapse roofs and most of the 20 000 residents had fled. The lower parts of walls

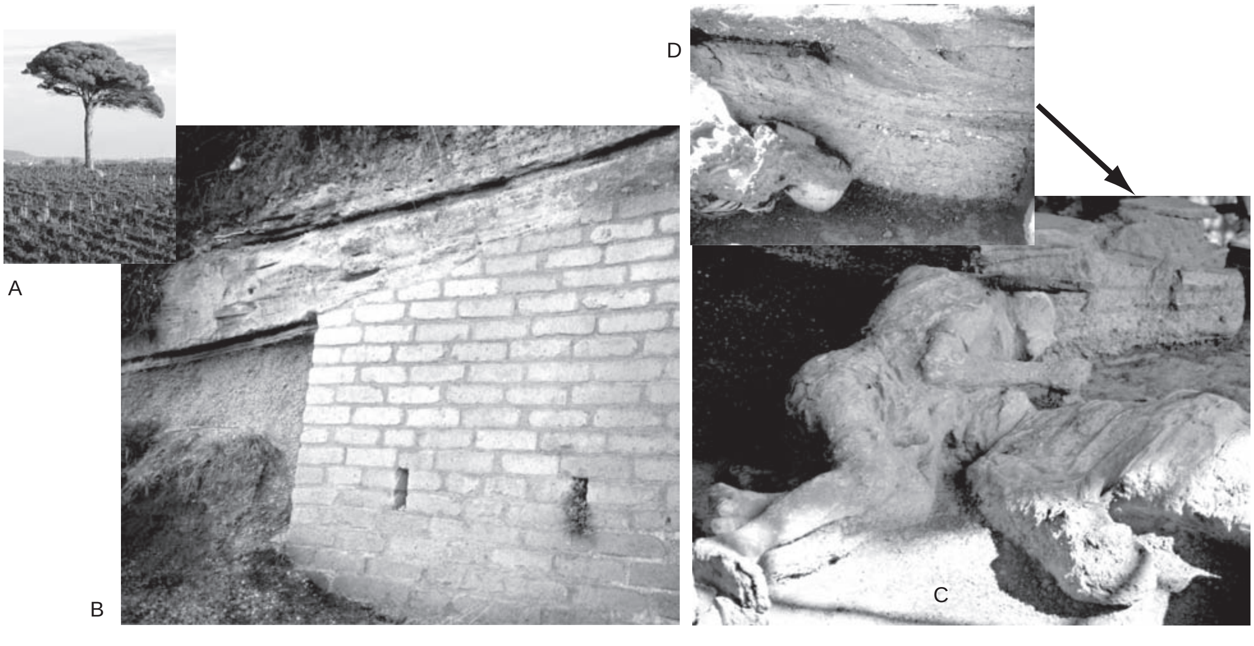


Fig. 4.36 (**A**) Pliny the Younger described the ash plume rising from Vesuvius during the initial stage of the eruption that destroyed Pompeii in AD 79 as resembling the pine trees of Rome (compare with Fig. 4.30). (**B**) During the early plinian stage of eruption, accumulations of ash-fall pumice caused roofs to collapse and buried the lower parts of buildings, as shown here at the left of the brick wall of a Roman tomb. When the plinian ash column collapsed, a high velocity surge deposit sheared off the upper parts of buildings that were not protected by the earlier fall of ash. The upper part of the tomb wall is truncated by the layering of the surge deposit, which contains bricks from nearby buildings. (**C**) Molds of two corpses, a man and a pregnant woman, lying on the ash-fall deposit are covered by bedded surge deposits. (**D**) Dune-like layering in a surge deposit in the lee of a human skeleton.

were buried in several meters of ash-fall pumice (Fig. 4.36 (B)). The plinian column of ash then became unstable and collapsed, sending out surge deposits that traveled rapidly across the countryside destroying anything that protruded above the earlier ash-fall deposit (Fig. 4.36(B)). The upper parts of walls were destroyed and buried along with ~2000 remaining residents in hot surge deposits. Figure 4.36(C) shows the molds of two corpses lying on the surface separating the ash-fall deposit from the overlying surge deposit. Figure 4.36(D) shows the depositional layering of the surge deposit in the lee of a human skull. Vesuvius still poses a potential hazard today. Computer simulations (Dobran *et al*., 1994) show that future eruptions could produce surge deposits that would devastate an area up to 7 km from the volcano, an area in which over a million people now live. Flavio Dobran's computer simulation of how a surge deposit is formed from a collapsing column can be found at http://www.mcli.dist.maricopa.edu/tut/final/vesuvius.html or http://urban.arch.virginia.edu/struct/ACADIA-97/pres-1.6.html.

Ash-flow deposits are formed from suspensions of particles that are carried along by hot rising gases. These suspensions have relatively high densities (about 1000 kg m^{-3}) but very low viscosities. Ash flows have much lower gas-to-solid ratios than do surge deposits. Consequently, they move like fluids, flowing down valleys at high velocities. Indeed, their velocities can be high enough to allow them to "jump" topographic barriers. Thick ash flows pond in topographic depressions; their thickness is therefore variable and determined largely by the underlying topography. They are poorly sorted and can contain particles ranging in size from ash to boulders (Fig. 4.37).

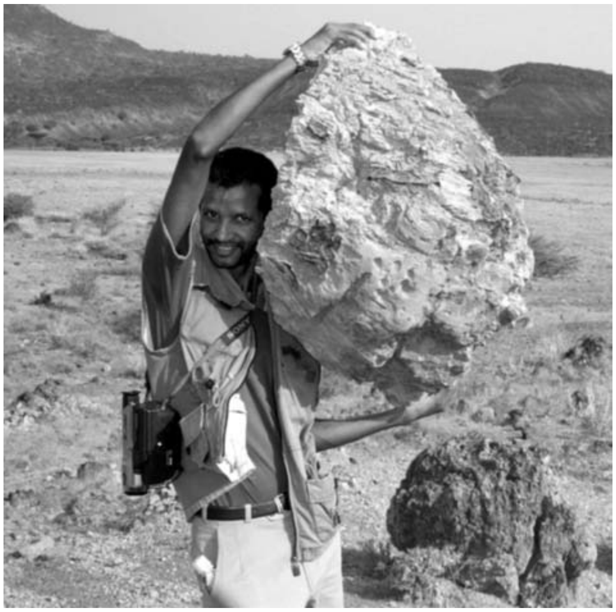


Fig. 4.37 Large fragment of ash-flow tuff containing fragments of rhyolitic pumice from the Dabbahu area of Ethiopia. The low density of the pumice is demonstrated by the geologist Asfawossen Asrat. Ash flows contain a much higher proportion of solids to gas than do surge deposits and as a result commonly contain a wide range of fragment sizes.

Explosive activity, such as that which destroyed Krakatoa in 1883, Monte Somma (Vesuvius) in AD 79, Mount Mazama (now Crater Lake, Oregon) 6600 years ago, or Mount St. Helens (Washington State) in 1980 throws volcanic ash high into the atmosphere, where winds may transport it for a great distance before depositing it as an *air-fall tuff*. Particles may consist of older rocks torn from the walls of the vent by the explosion, or of magma disrupted by expanding gas bubbles. On traveling through the atmosphere, the latter particles cool rapidly, both through radiation and adiabatic expansion of the gas. They are quenched to glass shards with sharp, cuspate boundaries that mark the outline of the previously surrounding gas bubbles (Fig. 4.41(A)). Although several days may be required for the farthest-traveled particles to settle out, deposition of air-fall ash from a single eruption can be considered an essentially synchronous geological event over a wide area. Extensive blankets of volcanic ash therefore provide the geologist with an ideal means of correlating geological events in widely separate areas, as long as individual ash falls can be distinguished from one another. The chemical composition of glass shards and the nature of phenocryst assemblages have both proved useful in "fingerprinting" tuffs. Dating by means of tuffs or tephra is known as *tephrochronology* (Thorarinsson, 1981).

Ash flows were first recognized during the devastating 1902 eruption of Mount Pelée, which destroyed the city of St. Pierre in Martinique. The French geologist Lacroix, who witnessed the eruption from the safety of a boat at sea, described *nuées ardentes* (glowing clouds) descending the flanks of the volcano at high speeds. Studies of ancient volcanic regions have since revealed that ash-flow tuffs are one of the most common modes of occurrence of rhyolite in the geologic record (Chapin and Elston, 1979). Moreover, many ancient ash-flow tuffs are orders of magnitude larger than those seen to erupt in historic times. The mode of emplacement of these large ash flows is still not completely understood.

Before continuing with our discussion of ash-flow tuffs, it is necessary to consider one remaining major volcanic structure, which, unlike those already mentioned, is not a constructional feature; these are the large calderas associated with rhyolitic magmatism. Calderas were mentioned in connection with shield volcanoes, but these are much smaller than those associated with rhyolites. For example, one of the largest shield calderas is that on Mauna Loa with an area of about 9.2 km^2, but this is small compared with a caldera such as that at Valles, New Mexico, which is about 430 km^2, or the Long Valley caldera in California, which has a similar area. These calderas, however, are dwarfed by the one at Yellowstone, which is about 2500 km^2 (Fig. 4.38). These large calderas are not formed by the collapse of the crest of a volcano but result from the foundering of large crustal blocks along ring fractures into large near-surface magma chambers. Volcanic activity that is associated with these structures emanates mostly from the ring fracture.

Collapse of these large calderas cannot take place without the displacement of large volumes of magma, and it is this

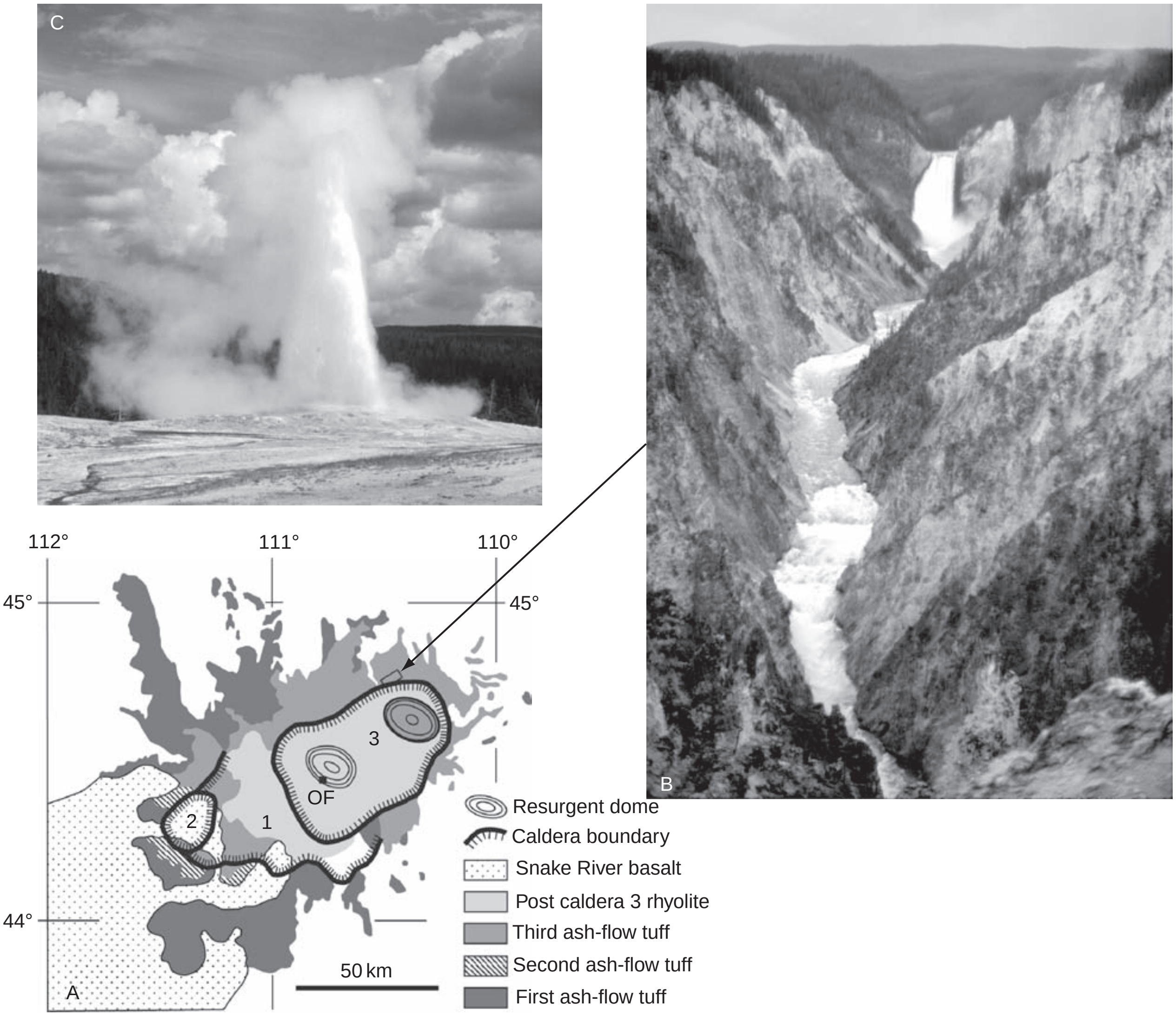


Fig. 4.38 (**A**) The Yellowstone volcanic center, Wyoming, was formed mainly during three successive episodes of caldera collapse and concomitant eruption of ash flows. The first caldera, which is the largest, erupted 2500 km^3 of ash-flow tuff in a single eruption. Note that many of the ash-flow tuffs extend out from the caldera in long tongues as a result of the avalanches of hot ash filling in existing river valleys. Within the third caldera, two resurgent domes have formed, with the Old Faithful geyser (OF) on the western flank of the western dome being driven by the thermal energy from the underlying magma (map simplified from Christiansen, 1979). (**B**) The Grand Canyon of the Yellowstone River is cut into the ash-flow tuff from the third caldera collapse. (**C**) A magma chamber still exists beneath the third caldera and heat from it drives the regular fountaining of Old Faithful geyser.

magma that has so often erupted in the form of ash flows. The volume of erupted ash flow might be expected to correlate directly with the volume displaced by the subsiding caldera block. Unfortunately, the amount of subsidence in a caldera is often difficult to determine. Smith (1979), however, has shown that the volume of an erupted ash flow correlates strongly with the area of the caldera, which is a much more easily determined dimension than its volume. As seen in Figure 4.39, this correlation holds true for calderas ranging in size over several orders of magnitude. The correlation of the erupted volume with the area of subsidence implies that most calderas sink approximately the same amount during an eruption, which from Figure 4.39 is about 0.5 km. No great significance can be placed on this number because there is considerable uncertainty in the values for the volume of erupted material, especially around ancient calderas, where erosion may have removed ash-flow tuffs or later eruptions may have concealed them. Despite this, it is reasonable to conclude that the relatively constant amount of subsidence, regardless of caldera size, indicates that the larger the caldera, the greater is the diameter-to-depth ratio of the subsided block; that is, the subsided block in a large caldera tends to be slab-like, whereas that in a small one is more piston-like.

The fact that many different-sized calderas have undergone similar amounts of subsidence implies that absolute depth of subsidence rather than caldera diameter may be more important in determining the amount of down-drop. This, in turn, suggests that pressure, and its effect on the solubility of volatile constituents in a magma, may be the key factor determining the amount of subsidence.

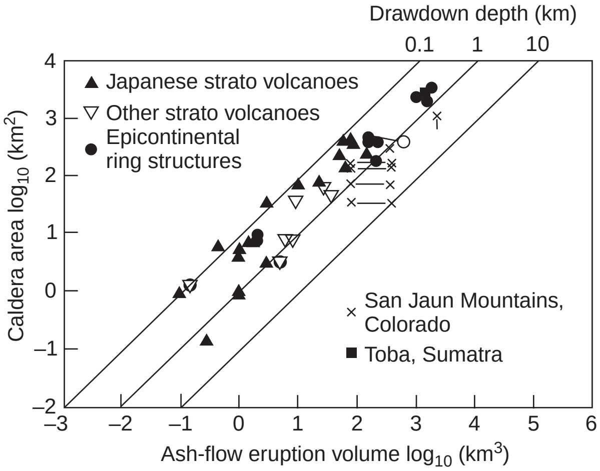


Fig. 4.39 Logarithmic plot of caldera area versus volume of erupted ash flow. The diagonal lines give depth of drawdown in the magma chamber, assuming vertical walls and a flat roof. Most calderas subside about 0.5 km regardless of the area of the caldera. (Drawn from data presented by Smith, 1979.)

By far the largest fraction of magma extruded during caldera subsidence is in the form of ash flows. Ash flows form only from magma that initially contains sufficient volatiles so that on decompression the magma is completely disrupted into a suspension of magma particles in gas. Until this disruption occurs, the rate of caldera subsidence is determined by the rate at which highly viscous magma is able to ascend ring fractures. With the onset of magma disruption the viscosity of rhyolitic magma decreases by at least five orders of magnitude, thus allowing for rapid extrusion of ash flows and rapid subsidence of the caldera (see Problem 4.6). Once started, vesiculation and disruption of the magma will work downward until depths are reached where pressures are great enough to prevent gas exsolution from disrupting the magma. This, then, may be the depth at which significant caldera subsidence ceases. The actual depth would depend on the composition of the magma and its initial volatile content. In Section 14.4 we will see that the tops of rhyolite magma chambers tend to be enriched in volatiles and hence are ripe for ash-flow eruptions.

Enormous volumes of magma have been erupted in the form of ash flows from calderas. Hildreth and Wilson (2007) estimate that approximately 600 km^3 of rhyolitic magma were erupted in about six days from the Long Valley caldera in California to form the Bishop Tuff. Some of the world's most voluminous ash-flow tuffs occur in Yellowstone National Park, in northwestern Wyoming, where there were three major eruptions (Smith and Christiansen, 1980). During the first, which occurred 2.0 million years ago (Perkins and Nash, 2002), a caldera over 75 km in diameter subsided and expelled more than 2500 km^3 of magma in a single ash flow (Fig. 4.38(A)). This volume is equivalent to spreading a layer of magma 200 m thick over an area equivalent to that of the state of Connecticut or the country of Wales. The next eruption, which occurred 0.8 million years later, was much smaller, forming only 280 km^3 of ash flow and a 25-km-diameter caldera at the western end of the first caldera. Magmatic activity then moved eastward, and 0.6 million years later another caldera, which measures 45 × 65 km, formed at the eastern end of the first caldera, and 1000 km^3 of magma was erupted, partly concealing the first caldera. The eastward movement of the magmatic activity, along with the trail of older silicic bodies that resemble islands surrounded by the flood basalts of the Snake River Plain, define the track of the Yellowstone hot spot (Pierce and Morgan, 1992). Post-caldera eruptions of rhyolitic lava flows (not ash flows) have flooded much of the floor of the first and third calderas. Where the Yellowstone River crosses from these harder rhyolites onto the softer ash-flow tuffs, it plunges into the spectacular Grand Canyon (Fig. 4.38(B)). Flood basalts from the Snake River Plain have inundated western parts of the ring structure. Since the last ash-flow eruption 0.6 million years ago, two *resurgent domes* have grown within the youngest crater, indicating that magma is still present at shallow depths. Such domes commonly form during the later stages of the life of a caldera system (Smith and Bailey, 1968). The Old Faithful geyser (Fig. 4.38(C)), located on the western flank of the western dome, and hot springs throughout the area are constant reminders of the proximity of the magma. The Yellowstone Park area remains a potential hazardous volcanic region, and consequently changes in the topography of the caldera floor are constantly monitored using satellite Interferometric Synthetic Aperture Radar (Wicks *et al.*, 1998, 2006).

We are now in a position to return to our discussion of ash-flow tuffs. By far the greatest volumes of these have been erupted from large calderas rather than from central volcanoes. But from either source they flow rapidly, even on gentle slopes, eventually ponding in topographic depressions. They commonly fill in old stream valleys, as can be seen from many of the tongues of ash-flow tuffs emanating from the Yellowstone calderas (Fig. 4.38(A)).

Considerable controversy surrounds the mechanism by which ash flows move. Clearly, gas is the medium in which the particles are suspended, but does this gas emanate from the particles of magma themselves, or is it air simply entrained in an avalanche of hot ash? In a viscous magma, gas can be exsolved only slowly. Thus, even after disruption of magma by expanding bubbles, gas will continue to exsolve from the particles of magma, and this may be sufficient to keep the particles apart and in constant agitation. Suspensions with which we are familiar that have formed in a similar way, but at much lower temperatures, are common aerosol sprays, which engineers refer to as being *fluidized*. Reynolds (1954) was the first to draw attention to the importance of fluidization in petrologic processes, but for a more recent treatment, the reader should refer to the paper by Wilson (1984). He showed, by experiment, that it is difficult to fluidize all particles in a pyroclastic flow because of the wide range in grain size. Also, the fluidization process tends to sort particles, transporting the fine pumice fragments to the top. Pyroclastic flows, however, exhibit little sorting. Some geologists believe that insufficient gas would be liberated from a magma during eruption to keep an ash flow fluidized (McTaggart, 1960). Instead, they think that fluidization results from the heating of air that becomes entrapped beneath the leading edge of an advancing ash flow. Both processes undoubtedly play roles, but the exsolution of gas from the particles of magma would appear to be the more important, at least where ash flows have spread over almost level surfaces for thousands of square kilometers.

A fluidized suspension of lava particles, being relatively dense and yet having extremely low viscosity, is capable of flowing at great velocities down the gentlest slopes, and traveling great distances. The ash flow that descended Mount Pelée on May 8, 1902, traveled a distance of 6 km at an average velocity of 20 m s^{-1} and had velocities as great as 150 m s^{-1} on the steep slopes of the mountain.

In contrast to the blanketing deposits of air-fall tuff from plinian-type eruptions, the distribution of ash-flow tuffs is controlled largely by topography, with flows descending valleys, thinning over topographic rises, and ponding in depressions. Where there have been many such flows, the filling in of topography results in the formation of flat areas or plateaus, similar to those formed by flood basalts. Above a glowing avalanche there is a cloud of ash, which being less dense than the fluidized mass beneath, is not controlled so much by topography and may spread out and blanket large areas with ash. It was such a cloud that actually destroyed the city of St. Pierre, not the ash flow itself, which was diverted by a river valley and entered the sea 3 km north of the city. Unfortunately, an upper stretch of this valley heads directly toward St. Pierre, and it was from here that the cloud above the ash flow was hurled straight at the city.

The temperature of an ash flow is initially that of magma in the vent, but as flow occurs, heat is lost through conduction, convection, and radiation, but also through the internal processes of gas expansion. Boyd (1961), who has analyzed the thermal behavior of ash flows, has shown that if magma is saturated with water prior to eruption, expansion of the gas resulting from the ascent from the source to the surface is approximately 20 °C/wt% H_2O in the initial magma. Thus, magma originally containing 10 wt% H_2O would have its temperature lowered by 200 °C due only to adiabatic expansion of gas (see Section 7.5). Counteracting the cooling is heat liberated by the exsolution of gas (about 11 °C/wt% H_2O exsolved) and by crystallization (about 0.4 J kg^{-1}), and by the dissipation of viscous forces which in a rapidly moving flow may be actually great enough to cause heating (Shaw and Swanson, 1970).

Although the precise thermal history of an ash flow may be difficult to ascertain, field evidence clearly indicates that when these flows come to rest they are hot enough to cause some degree of welding of their constituent glassy particles. These welded ash-flow tuffs are commonly referred to as *ignimbrites*. Some ash-flow tuffs, such as the Bandelier, which was erupted from the giant Valles caldera in the Jemez Mountains of New Mexico (Heiken *et al*., 1990), shows variable degrees of welding (Fig. 4.40(A) and (C)). In contrast, the ash-flow tuff of the Chiricahua National Monument in southern Arizona welded itself together to form a thick massive unit that developed large columnar joints as it cooled (Fig. 4.40(B)). Later weathering along these joints developed remarkable rock spires. During welding, particles of pumice and glass shards can become flattened parallel to the sheet, giving the rock a prominent foliation. Contrast, for instance, the shape of the pumice particles in the ash-flow tuffs in the examples of the Bandelier and Chiricahua shown in Figure 4.40(C) and (D) respectively. Welded ash-flow sheets can undergo considerable compaction during welding, as indicated by the flattening of pumice particles, which on average are spherical prior to compaction (see Problem 4.5). Welded ash-flow tuffs are seen in thin sections to consist of a streaky mixture of glassy and devitrified glassy particles that wrap around crystals and solid rock fragments; this texture is referred to as *eutaxitic*. (Fig. 4.41(B)). This texture contrasts with that of ash-fall tuffs where the glass shards maintain their cuspate shape (Fig. 4.41 (A)). In welded ash-flow tuffs, obsidian fragments are mostly dark and pumice fragments are light colored and when flattened produce a striking flame-like, or *fiamme*, texture (Fig. 4.41(B)). Unlike the fragments in ash-fall deposits, those in ash-flow deposits show little evidence of sorting and can range in size from that of sand to several tens of centimeters. This is to be expected, because most ash flows move in a turbulent manner, at least near their source, and this will continuously mix the particles.

Ash-flow tuffs are extremely common in the geologic record. One of the first to be recognized was that which occupies the Valley of Ten Thousand Smokes, Alaska. It was formed in 1912 by a great eruption in the vicinity of the Katmai volcano, but it was not until several years later that geologists examined the deposit and concluded that it had been formed by an ash flow similar to those seen on Mount Pelée. The great explosion that produced Crater Lake, Oregon, in addition to forming ash-fall tuffs over large parts of the northwestern United States, erupted 33 km^3 of ash flows. The Bishop ash-flow tuff (Fig. 4.41(B)), which has a volume of 600 km^3, erupted 0.76 million years ago in ~6 days from the Long Valley Caldera in eastern California. In some areas, successive eruptions have produced large volumes of ash-flow tuffs, as for example, in the central part of North Island, New Zealand, where they have a total volume in excess of 7000 km^3. Even more extensive deposits formed during the Miocene in Nevada and covered an area of approximately 100 000 km^2. Ash-flow tuffs are, therefore, the only type of volcanic product with volumes approaching those of the great flood-basalt flows.

Intrusive bodies

4.5 GENERAL STATEMENT

The shapes and sizes of intrusive igneous bodies are more difficult to determine than those of volcanic bodies, because

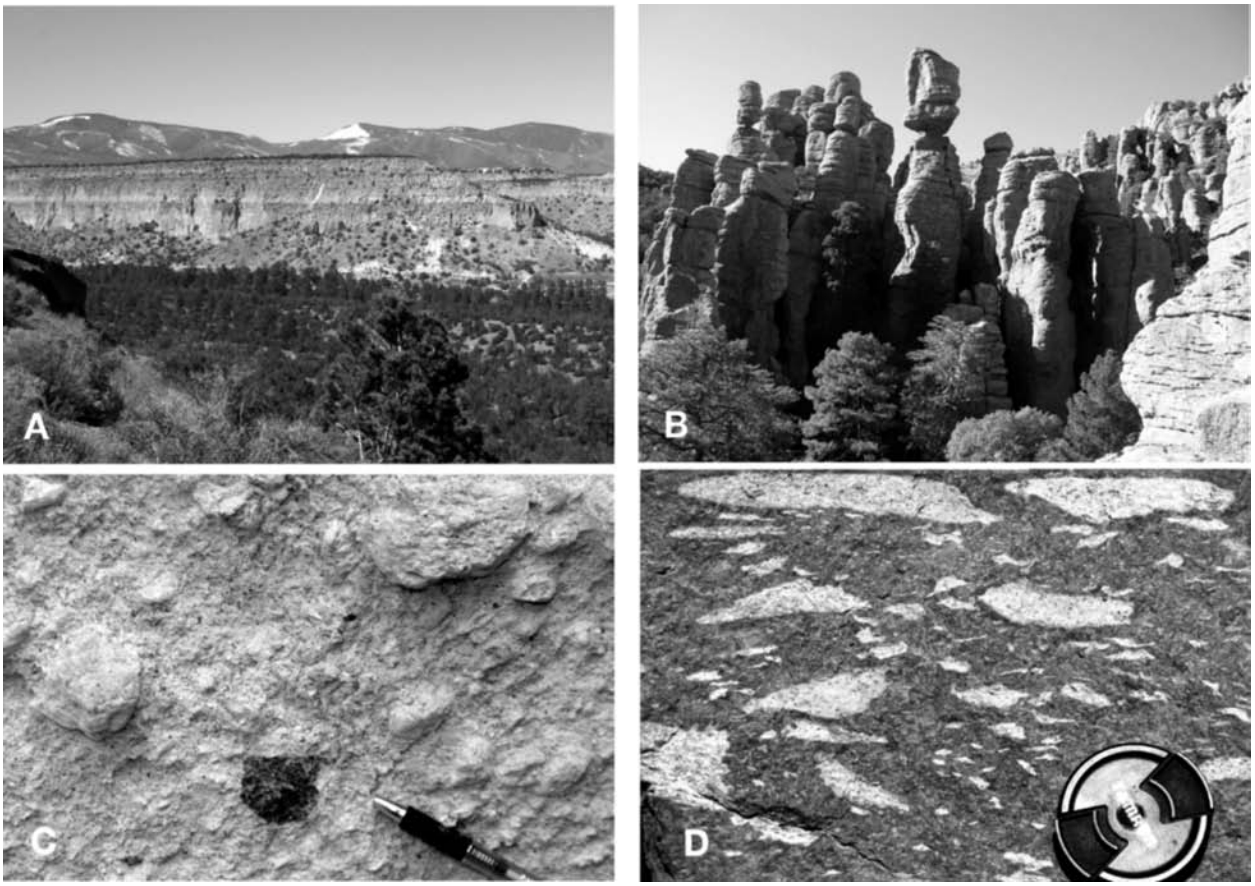


Fig. 4.40 (**A**) The Bandelier ash-flow tuff exposed in cliffs below Los Alamos, New Mexico. The snow-capped Jemez Mountains in the background form the rim of the giant Valles caldera from which the tuff was erupted (Heiken *et al*., 1990). (**B**) Welded ash-flow tuffs in the Chiricahua National Monument, Arizona. On cooling, this welded sheet developed columnar joints, which later weathering has sculpted into remarkable rock columns. (**C**) Pumice fragments in the Bandelier tuff. Note the one particle of dark basalt; these are remarkably common in rhyolitic ash flows and attest to the presence of basaltic magma at depth. (**D**) Particles of light-colored pumice in the Chiricahua ash flow were flattened into ellipses during welding and compaction.

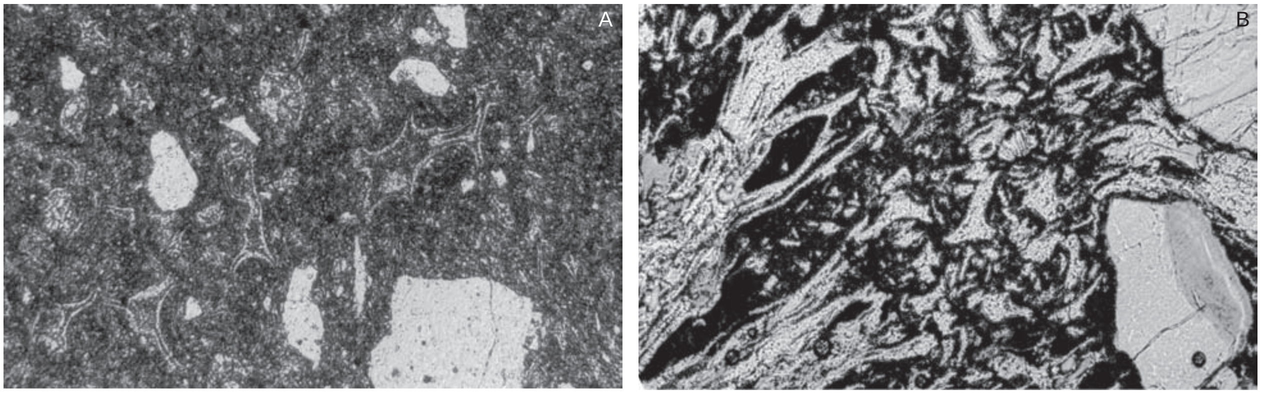


Fig. 4.41 (**A**) Ash-fall deposits typically consist of glass shards and pumice fragments that show no signs of deformation; Los Alamos, New Mexico. (**B**) Similar fragments in ash-flow deposits become flattened, welded together, and wrap around phenocrysts; Bishop tuff, California. Both fields of view are 3.5 mm wide.

the Earth's erosion surface provides only a limited three-dimensional view of these bodies. Geophysical surveys, including gravity, magnetic, electrical, heat flow, and seismic, can impose constraints on our interpretation of the shape of bodies at depth, as long as the physical properties of the igneous rock and country rock are sufficiently different. Structures in rocks surrounding intrusive bodies can also provide information on the form of an intrusion. Finally,

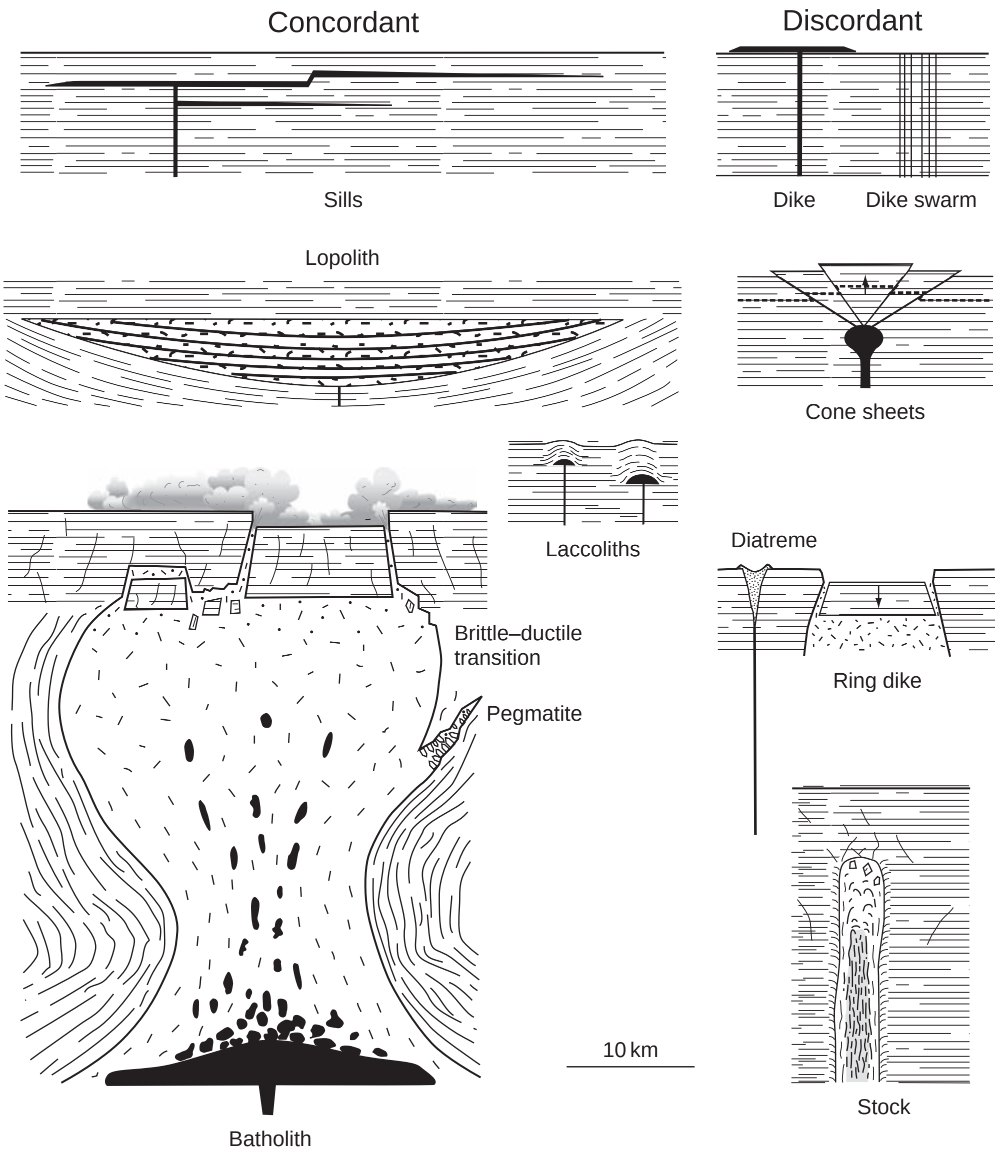


Fig. 4.42 Forms of major intrusive bodies and their approximate sizes. Rocks formed from basaltic magma are dark colored, whereas those formed from rhyolitic magma are light colored. Bodies on the left, in general, are concordant; that is, they parallel the structure of the intruded rocks. Bodies on the right are discordant and cut across the structure of the intruded rocks.

from maps of intrusions that are exposed to different depths by erosion, a generalized picture of the form of intrusive bodies has emerged. Only a relatively small number of common intrusive forms exist (Fig. 4.42).

It is important to remember that the final shape and size of an igneous intrusion need not reflect the size of the magma chamber that produced the body at any given time (see, for example, Gibb and Henderson, 1992). This is particularly true of large intrusions. The rate of supply of magma must be finite and depends on many factors, such as melting in the source region, porous and channeled flow to segregate the magma, and eventually conduit flow to bring the magma to its final resting site. Large intrusions cannot instantly come into existence; they must form over an extended period. Even a body as simple as the Palisades Sill is now recognized to have been inflated to its present 300-m thickness in a series of injections. The Palisades Sill probably never existed as a 300-m-thick body of magma. Similarly, although exposures through ancient mid-ocean ridges indicate large intrusive bodies, seismic surveys across modern ridges indicate that the volume of magma present at any one time is much smaller than these bodies. These bodies represent the sum of the solidification products from smaller magma chambers that were replenished during the life of the chamber. The volume of a volcanic eruption provides at least a lower limit on the volume of near-surface magma chambers. For example, the magma chamber that existed beneath Lakagigar, Iceland, in 1783 was able to supply 14 km^3 of basaltic magma in a period of several months, and the magma chamber beneath Yellowstone was able to supply 2500 km^3 of rhyolitic magma during the first caldera subsidence.

Intrusive bodies are divided into two classes; *discordant bodies* crosscut the structure of the intruded rocks, whereas *concordant bodies* parallel it (Fig. 4.42). Although purely descriptive, this division has some genetic implications. Near the surface of the Earth, where rocks react to stress by fracturing, discordant bodies are common. At depth, where fractures are less abundant due to the ability of rocks under high confining pressures to relieve stresses by flowing, other features, such as compositional layering and schistosity, can provide planes of weakness along which magma intrudes to form concordant bodies.

Unlike volcanic bodies, intrusive ones must displace rocks to make room for themselves. The way in which this occurs largely controls the form taken by the intrusion. There is considerable uncertainty, however, as to precisely how this takes place. Historically, arguments concerning this point have figured prominently in discussions concerning the origin of granite and was one of the reasons some geologists believed that granite was not of igneous origin at all, but was formed by solid-state replacement of other rocks. Today, the room problem remains one of the most perplexing in igneous petrology.

Magma may simply rise into tensional fractures that result from extension of the Earth's crust. Open fractures, however, extend only to shallow depths. Moreover, the width of bodies formed in this way is likely to be small and limited by the rates of tectonic plate motion.

If magma does not have open fractures to enter, it must displace the rock it is to intrude. This can be done by forcing country rock aside. Near the surface of the Earth, this may simply involve the lifting of the overlying rocks. At depth, however, solid rock must be pushed aside, which is likely to occur only if the country rock is rendered plastic by heat from the intrusion. But most magmas do not contain excess heat; thus, this mechanism of making room for magma is severely limited.

Magma may also provide space for intrusion by having blocks fall from the roof of the magma chamber and sink to lower parts, thus effecting an exchange of solid material from above with molten material from below. This process is known as *stoping* because of its similarity to the common mining practice of removing blocks of ore from the roof of underground workings or *stopes*. Stoping, of course, does not provide an explanation for how the initial body of magma made room for itself. Evidence of stoping is found in the margins of many igneous bodies, in particular those formed at high levels in the Earth's crust. Zones of igneous rock containing fragments of the country rock rim many intrusions and form what are known as *igneous breccias* (Fig. 4.76). For stoping to be effective, the country rocks must be denser than the magma and be fractured or capable of being fractured. This second requirement limits stoping to the upper part of the crust, for in the lower part rocks are able to flow slowly under high confining pressures and hence fractures are not likely to exist except during periods of rather rapid tectonic movement. The rate at which magma can intrude by stoping is determined by the availability of fractured rock and the sinking rate of fragments in the magma; this, in turn, depends on the viscosity of the magma, the size of the fragments, and the density contrast between fragments and magma (see Section 14.3).

4.6 VOLCANIC NECKS

The simplest form of intrusive body for which the room problem has a simple solution is the *volcanic neck* (Fig. 4.43). In volcanic regions that have undergone some degree of erosion, volcanic feeder pipes commonly stand out as resistant spines against surrounding softer rocks. Once an eruption starts, the width of a volcanic vent is increased by pieces of the wall rock being broken off and carried away by the erupting lava. Near the surface, pipes tend to be almost cylindrical, but at depth, they can become elliptical or even cloverleaf shaped. At still greater depth, the shape commonly merges into the intersection of several dikes. For example, Shiprock, New Mexico, is an example of a shallow volcanic neck (Fig. 4.43). Radiating from the neck are three vertical dikes. At depth, the intersection of these dikes probably provided the main focus for the flow of magma, which on rising toward the surface eroded the intersection into the near cylindrical section shown at the level of Shiprock. Deep mining operations in diamond mines in South Africa have exposed, in three dimensions, this transition from dike intersection to carrot-shaped volcanic pipe (Fig. 4.59).

4.7 DIKES AND SILLS

Dikes and sills are sheet-like bodies that have great lateral extent relative to their thickness, and their opposing contacts are approximately parallel. Dikes crosscut the structure of the intruded rocks and thus are discordant; sills parallel the intruded structure and thus are concordant (Fig. 4.44). Dikes derive their name from the fact that they are commonly composed of rock that is more resistant to weathering than the rocks they intrude and consequently stand up in relief and resemble a wall (Fig. 4.43(B)).

Dikes and sills range from millimeter-thick films (Fig. 4.45) to massive sheets, hundreds of meters thick and tens of kilometers long (Figs. 4.46, 4.80). In general, sills are thicker than dikes, but dikes commonly occur in such large numbers that their cumulative thickness can be immense. The Palisades Sill, which forms the escarpment on the west bank of the Hudson River opposite New York City, is 300 m thick and can be traced along strike for 80 km. The Whin sill in northern England is only 75 m thick, but it is laterally extensive; it forms the 125-km-long escarpment that provided the Romans with a substantial foundation on which to build Hadrian's Wall. In the southern hemisphere, extensive sills and lava flows are associated with the breakup of Gondwanaland (Hergt *et al.*, 1991). In South Africa, the Karoo sills are very extensive, with individual sheets having areas of as much as 12 000 km^2. Extensive sills of the same age are found in the Dry Valleys region of Antarctica (Fig. 4.46). A useful web site discussing these sills and associated extrusive rocks can be found at www.geokem.com/flood-basalts-1.html.

Dike widths are commonly in the range of meters to tens of meters, with ones such as the 150-m-wide feeder to the

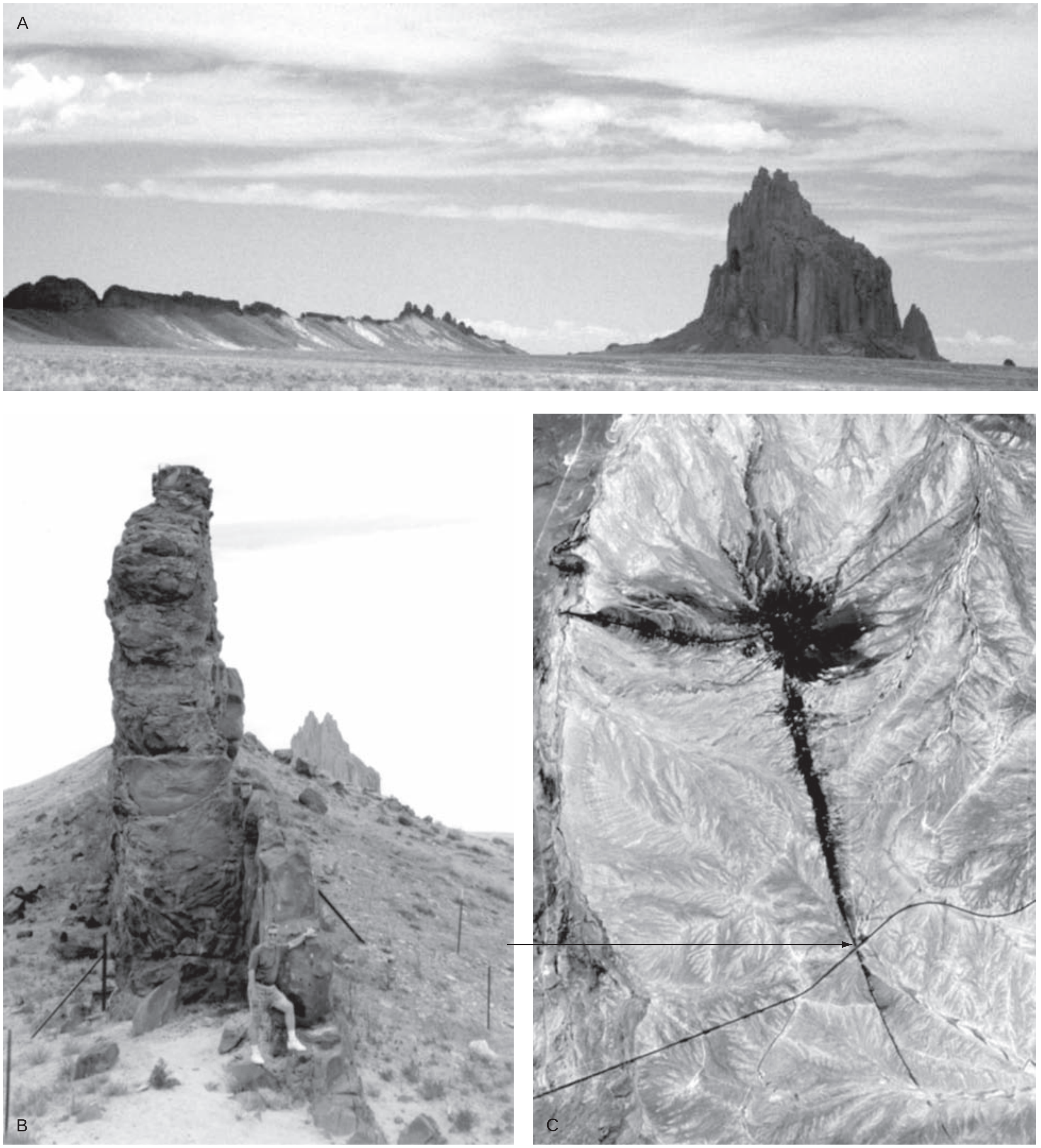


Fig. 4.43 Shiprock volcanic neck and radiating dikes, New Mexico. (**A**) The volcanic neck towers 335 m above the plain. A resistant dike forms a prominent wall extending south from the neck. (**B**) The dike is more resistant than the surrounding rocks and consequently has weathered to form a wall. (**C**) Satellite image showing the volcanic neck and three dikes radiating to the south, west, and northeast. The width of the image is 10 km. Satellite image is from NASA Earth Observatory (www.visibleearth.nasa.gov/17587/shiprock).

large Muskox intrusion in Arctic Canada (Fig. 15.18) being rare. The so-called *Great Dyke of Zimbabwe* (*Rhodesia*), which is 5 km wide and 500 km long, is not actually a dike but is, instead, a long, narrow down-faulted block of igneous rock formed in a previously overlying horizontal sheet-like body. Dikes in Iceland average only 3 to 5 m wide, yet in one 53-km stretch of coastline, Walker (1965) measured approximately 1000 dikes with a total thickness of about 3 km. Dikes may extend laterally for great distances (Fig. 4.80), with lengths in excess of 100 km being common for those with widths of a few tens of meters.

Open fractures exist only near the Earth's surface, so at depth dikes and sills must generate their own pathways. This they do in one of three ways (Fig. 4.47) depending on the

Fig. 4.44 Dikes and sills intrusive into gently dipping Ordovician limestone surrounding the Cretaceous Monteregian intrusion at Montreal, Quebec. The two narrow, dark dikes that slope upward to the left and are cut by the wider vertical dike on the right are both offset by the vertical dike; the narrow dikes must therefore be older. The upper narrow dike is offset also by the upper sill (center of photograph), but the lower narrow dike cuts the lower sill. The lower sill must therefore be one of the earliest intrusive rocks in this exposure. The wider vertical dike on the left is composite; a narrow felsic dike is intruded within it. What can be concluded from the intersection of this composite dike with the two sills? (See Problem 4.12.)

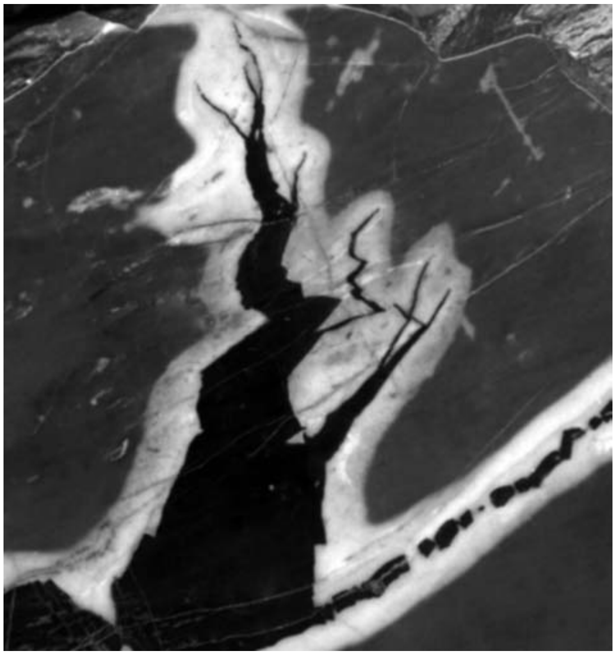


Fig. 4.45 Cretaceous basaltic dike intrusive into grey Ordovician Trenton limestone in the contact aureole of the Mount Royal stock, Montreal, Quebec. Heat from the basalt appears to have converted the gray organic-rich limestone to white marble containing small crystals of graphite. However, the width of the zone of marble is not proportional to the width of the dike, which would be expected if the dike were the source of the heat. The conversion of gray limestone to white marble involved the conversion of organic material to graphite and loss of volatiles. This was probably caused by a flux of hot fluids that passed through the fractures prior to the emplacement of the magma. The magma then sealed the fractures. The dike was emplaced by dilation, as indicated by the fit of opposing sides of the dike. Small fragments of marble were just beginning to separate from the wall to form xenoliths in the dike. The thin lower dike branch was segmented by post intrusive flow of the marble. Width of sample is 8 cm.

properties of the intruded rocks (Pollard, 1973; Shaw, 1980). In each case, the direction of propagation is in a plane normal to the direction of minimum principal stress. In the upper part of the lithosphere where rocks are brittle, excess magma pressures are able to fracture rock. Magma flowing into a fracture concentrates great stress on the tip of the fracture, even when excess pressures are small. The rock is therefore wedged apart by the intruding magma (Figs. 4.47(A), 4.45) in much the same way that fractures can be opened in rock by pressurized water. Indeed, the term *hydraulic fracturing* is commonly used for this type of dike propagation. Magma is far more viscous than water, so the tip of the dike lags behind the tip of the fracture (Fig. 4.48). At greater depth in the Earth where confining pressures are higher, an intrusive sheet may propagate by forcing aside rock along conjugate brittle shear planes, which are oriented at about 35° to the plane of the sheet (Fig. 4.47(B)). This bulldozer-like action produces a more blunt-ended sheet than does the wedging action (Kattenhorn and Watkeys, 1995). In addition, magma may intrude along the shear planes, and fault-bounded blocks of country rock may become incorporated in the sheet. At still greater depths, the mechanism of propagation is essentially the same, except that the shear planes become ductile faults, which make an angle of about 45° with the plane of the sheet (Fig. 4.47(C)). If the intruded rocks become completely plastic, a diapiric dome rather than a sheet-like body will form.

In the zone where sheet propagation is by "hydraulic fracture," the width $2w$ of a sheet and its horizontal length $2b$ (Fig. 3.8) are related to the physical properties of the intruded rock by

$$\frac{2w}{2b} = \frac{2.25P_{\mathrm{ex}}}{\rho_{\mathrm{host}} V_P^2} \tag{4.1}$$

where P_{ex} is the excess pressure of the magma (see Eq. (3.1)), ρ_{host} the density of the host rock, and V_P the compressional seismic wave velocity in the host rock. With increasing depth, V_P increases and thus the ratio $2w/2b$ must decrease. Therefore, we can expect dikes to be longer and narrower at depth and shorter and wider near the surface. Typical values inserted in Eq. (4.1) indicate that near Earth's surface $2w/2b$ has values of 10^{-2} to 10^{-3}, but at depth it has values of 10^{-3} to 10^{-4}. A 10-m-wide dike near the surface of the Earth, for example, might have a length of ~3 km, whereas at depth a 10-m-wide dike would be 30 km long.

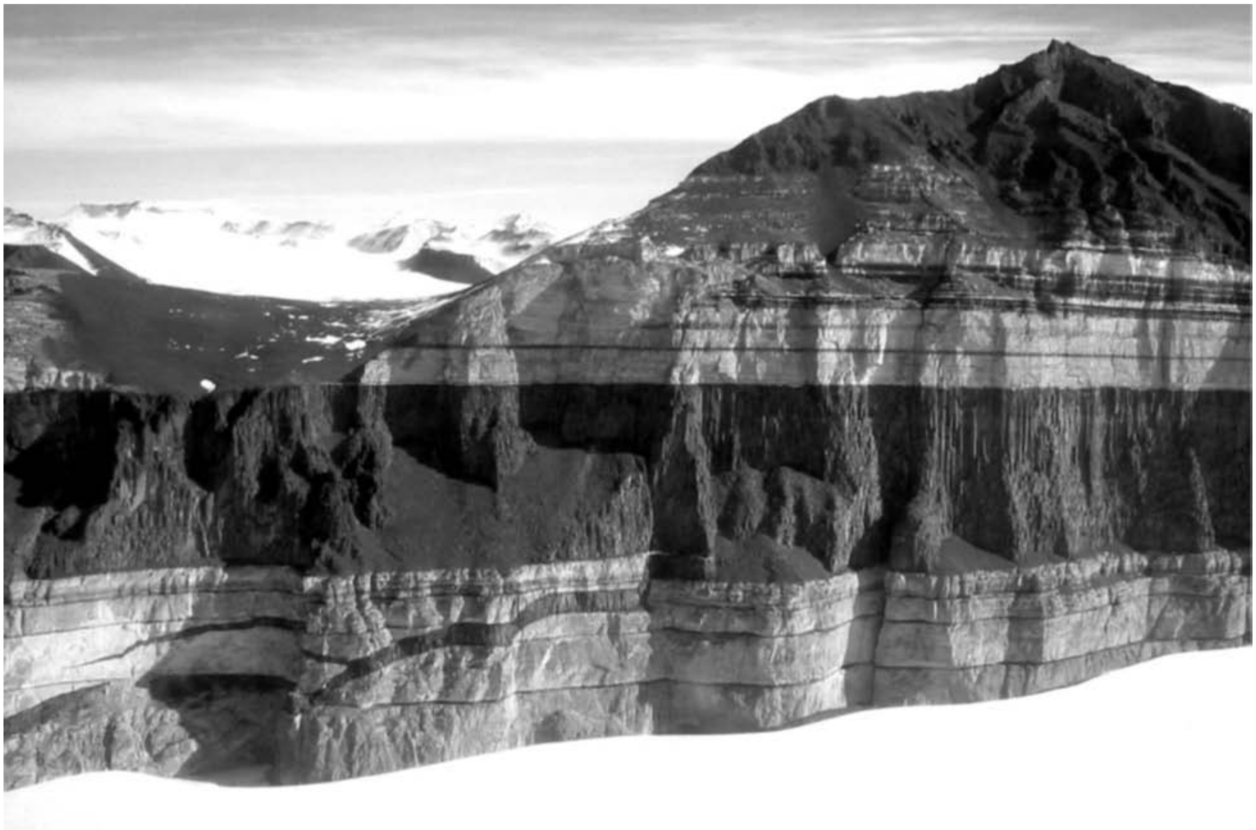


Fig. 4.46 One of the higher Ferrar diabase sills that fed the overlying Jurassic Kirkpatrick basalts (mountain tops) in the Dry Valleys region of Antarctica. The sill intrudes the Beacon Sandstone, which was deposited in rift valleys during the breakup of Gondwanaland. A large slab of sandstone can be seen to have separated from the floor of the main sill and risen vertically, probably as a result of buoyancy (see Fig. 4.50(D) for geometry of emplacement). (Photograph by Jean H. J. Bédard of the Geological Survey of Canada.)

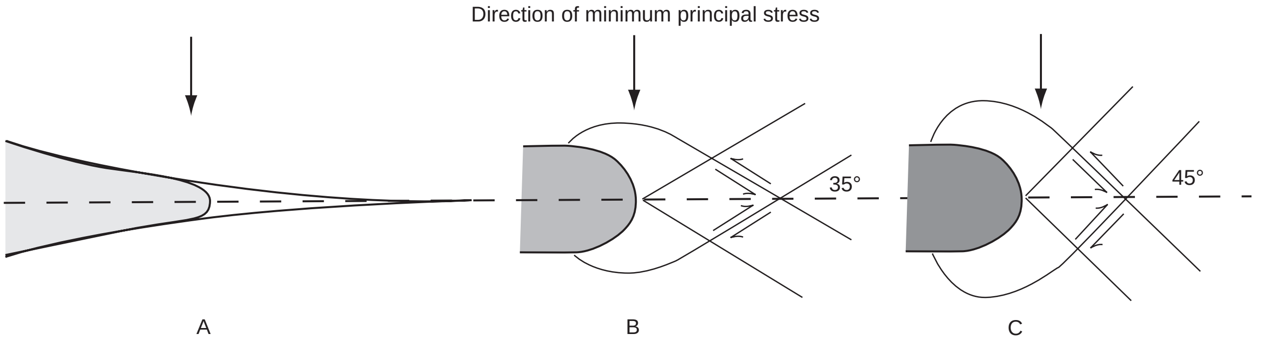


Fig. 4.47 Three mechanisms of dike propagation. (**A**) In the brittle upper crust, excess magma pressure can fracture rock; in such cases, magma lags behind the tip of the fracture. (**B**) At greater depth where confining pressures are greater, magma forces aside the intruded rock along conjugate brittle shear planes. (**C**) At still greater depth, deformation takes place on ductile shear planes. In each of these cases, the direction of propagation is in a plane perpendicular to the direction of minimum principal stress (after Pollard, 1973).

The pattern formed by the 250-km-long Mesozoic Higganum dike set, which crosses Connecticut and Massachusetts, may well be explained in terms of the width-to-length ratio changing as the dike propagated toward the surface, as implied in Figure 4.49. The dike consists of short individual segments that are ~4 km long and up to 60 m wide. These dimensions give a typical near-surface value for $2w/2b$ of about 10^{-2}. The short segments, in turn, form longer groupings extending over distances of 10 km. These longer groups may reflect the presence of 10-km-long dike segments at depth, which would be expected from Eq. (4.1). The longer groups, in turn, may branch from a main fracture in the lower lithosphere that controls the overall trend of the dike set. Two other dikes to the west of the Higganum dike exhibit similar segmentation. In each of these dikes, segments are slightly offset from the next, and the segments tend to overlap. Where they overlap, they thin, so that their combined thickness remains constant. The strikes of the overlapping sections

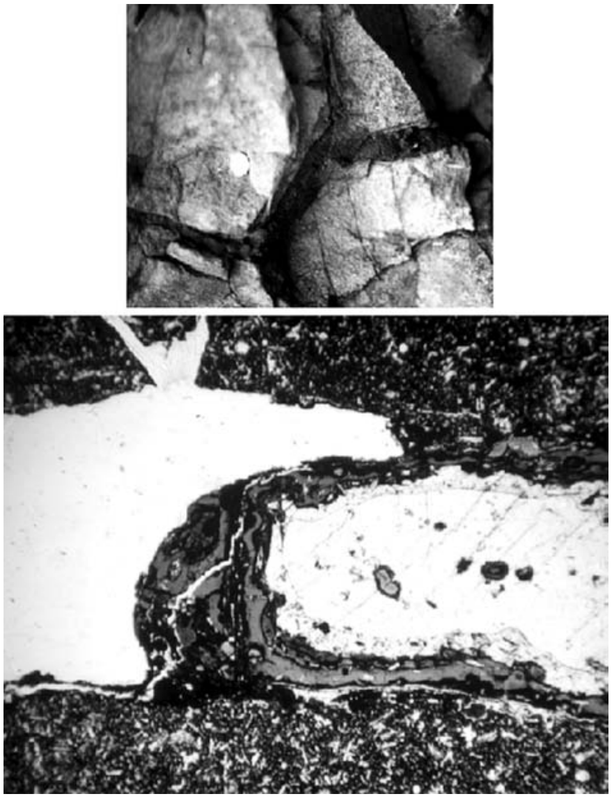


Fig. 4.48 Lobate front of magma that was quenched to a glass as a second pulse of magma advanced into a columnar joint set (inset) in the margin of the Mesozoic Fairhaven diabase dike, Connecticut. The dark rim on the lobe is due to alteration and devitrification of the glass. Width of fracture 5 mm.

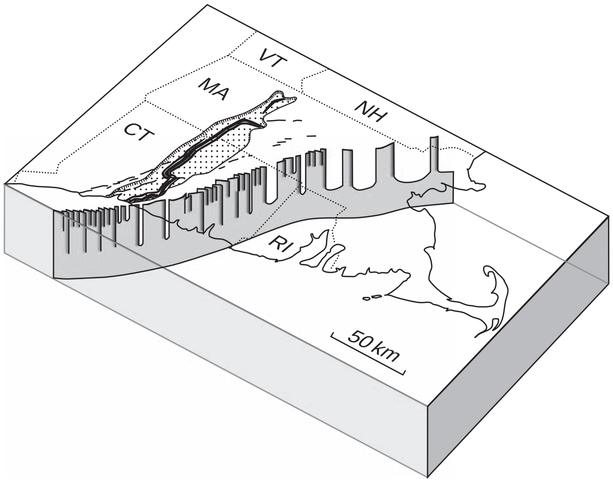


Fig. 4.49 Three en echelon diabase dike sets that traverse southern New England in a northeasterly direction were the feeders to the three flood-basalt units preserved as eastward-dipping flows in the down-faulted Mesozoic Hartford basin (stippled) of Connecticut and Massachusetts (Philpotts and Martello, 1986). The Higganum, the easternmost of these dike sets (shown in 3-D), fed the earliest basalt. The en echelon segments of this dike form two distinct groupings; the shortest segments are ~4 km long, whereas the larger groupings are ~10 km long.

tend to veer toward the other dike, and if the dikes actually meet, they do so at right angles. The 11-km-long Inyo dike of Long Valley, California, shows a similar pattern of segments, which has been interpreted to result from depth variations in tectonic stresses (Reches and Fink, 1988). Dike sets along spreading axes on the ocean floor can exhibit this same geometrical arrangement of dike segments (Lonsdale, 1985). Indeed, this pattern is common in dike sets at all scales and is referred to as *en echelon* (Pollard *et al.*, 1975).

The width of an intrusive sheet, in particular of a feeder dike, can be modified by magmatic erosion. For example, country rocks adjacent to the contact can be melted or blocks ripped off and transported away with the magma. Because flow velocities in dikes are proportional to the width of a dike (Eqs. (3.16) and (3.24)), this erosion is more likely to occur where a dike is already widest. Dikes also solidify more rapidly where they are narrow, which further concentrates the flow in wider parts. Consequently, wider parts of dikes tend to become wider (Delaney and Pollard, 1982). This widening and resultant concentration of flow at a few points is well illustrated by fissure eruptions, which are fed by dikes. The 1959 eruption of Kilauea Iki began with magma erupting along the entire length of an 800-m-long fissure, but within hours the flow became concentrated at a single vent (Richter *et al.*, 1970). Similarly, the huge fissure eruption in 1783 at Lakagigar in Iceland began with lava pouring from a 25-km-long fissure, but soon the flow became concentrated at a few nodes. In addition, the degree of contact metamorphism developed where dikes are thicker is commonly far more intense than can be accounted for simply by the increased thickness of the dike; here the amount of heat liberated by the greater flux of magma is the controlling factor.

There are limits to how thin an intrusive sheet can be. Thin sheets may travel so slowly that heat losses are rapid enough to cause the magma to solidify before it has traveled far. In addition, the yield strength of a Bingham magma may be great enough to halt flow in a thin sheet. The thin dike shown in Figure 4.48 clearly was frozen to a glass as it flowed along a fracture. If magma is to continue flowing, intrusion must be rapid enough to maintain the magma in a liquid or partly liquid state (see discussion of thermal entrance length and Eq. (3.14) in Chapter 3).

An estimate of the minimum conduit dimensions necessary for maintenance of magmatic flow has been calculated by Wilson and Head (1981). Cooling rates and magma yield strengths are both important factors in determining these dimensions. Unfortunately, little is known about the yield strengths of magmas. For magma with a small yield strength, the cooling rate will largely determine the minimum thickness of a sheet. In such cases, the mass flux of magma per unit length of fissure J/L that must pass through a sheet of thickness W to maintain flow is

$$\frac{J}{L} = 1200\eta\left[\left(1+\frac{1.39\times 10^{-4}\; gW^3(\Delta\rho)\rho_{\mathrm{m}}}{\eta^2}\right)^{1/2}-1\right] \quad \mathrm{kg\ m^{-1}\ s^{-1}} \qquad (4.2)$$

where $\Delta\rho$ is the density difference between host rock and magma, ρ_{m} the magma density, and η the magma viscosity. If the flux necessary to maintain flow is determined, the velocity and driving pressure can be calculated. The driving pressure in most cases is determined by the lithostatic pressure on the magma; it therefore has a definite limit, which, in turn, will control the minimum width of the dike (see Problem 4.2). For a more extensive discussion of the thermal entrance length, see Lister and Kerr (1991) and Rubin (1995).

Although dikes and sills are not restricted by definition to having a particular attitude, wide dikes generally dip steeply, and most thick sills are relatively flat lying, provided that they have not later been folded. Also, although dikes and sills can be formed from many different rock types, most thick ones are formed from basaltic magma, which crystallizes to a slightly coarser grain size than in lava flows; this rock is called *diabase* (*dolerite* in British usage).

The grain size and textures of dikes and sills show marked variations across their widths. In the upper part of Earth's crust, where country rock temperatures are relatively low, hot bodies of magma cool rapidly along their margins to produce fine-grained rocks. The grain size increases in a systematic manner toward the center of these bodies (see Chapter 5, Eq. (5.21)), and in the case of diabase sheets, dikes consistently become coarser grained than do sills of comparable thickness formed in the same environment. Deeper-seated bodies cool less rapidly along their margins and hence show less grain-size variation. Dikes and sills commonly contain phenocrysts, which are particularly evident in chilled margins, where they may be aligned by the flow of magma parallel to the contacts. If the early crystallizing minerals are denser than the magma, they may sink and accumulate in the lower parts of sills.

A dike or sill may be the site of more than one injection of magma (Fig. 4.44). Renewed surges of magma prior to the complete solidification of a sheet result in the intrusion of material along the center of the body. If solidification is complete, fractures formed by the cooling and shrinkage of the first sheet can localize the emplacement of later magmas. In the latter case, sharp internal contacts are formed, and with sufficient time for cooling between injections, fine-grained chilled margins can form on the younger rocks. These bodies are described as *multiple* where successive injections are of essentially similar composition and as *composite* where they are of different composition (Fig. 4.44). Composite dikes and sills can contain rocks as different as diabase and granophyre (granite). Such composite dikes are common, for example, around many of the Tertiary igneous complexes in northwestern Scotland and Northern Ireland. In these composite sheet-like intrusions, the granophyre is usually younger than the diabase and typically occurs in the central part of the sheet. Multiple dikes are the main type of intrusive body beneath mid-ocean ridges. In ancient ocean floor that has been obducted onto continents, sheeted dike complexes are found to underlie the pillow lavas that were erupted onto the sea floor. Because successive intrusions occur in the center of earlier dikes, these sheeted dikes commonly exhibit chilled margins only on one side; not until the spreading axis is crossed are the opposing contacts found (see Figure 15.5 for clarification).

Structures associated with the emplacement of dikes and sills generally indicate dilation. In fact, with many small sheets, especially those with irregular outlines, opposing contacts can be seen to fit back together in considerable detail (Fig. 4.45). There are some bodies, however, such as the 120-m-wide Medford dike in Massachusetts or the giant (up to 500-m-wide) Gardar dikes of southern Greenland (Bridgwater and Coe, 1970), whose emplacement can be shown to have involved no dilation; these are thought to have been emplaced by stoping. Evidence for dilation consists of structures in the country rock that, upon intersecting a sheet at an oblique angle, are offset across the body, as illustrated in Figure 4.50(A) and (B). Stoping produces no offset of the country rock structures (Fig. 4.50(C)). Some dikes can be formed by replacement; that is, fluids flowing along a fracture may cause replacement of the surrounding rock. In this case, a sheet-like body may result, but the surrounding rock undergoes no dilation. Such bodies are said to be *metasomatic* (Chapter 21).

An interesting type of dilation occurs in those sills formed by the buoyant rise of the overlying rocks (Fig. 4.46). In these bodies, the dilation must be in a vertical direction,

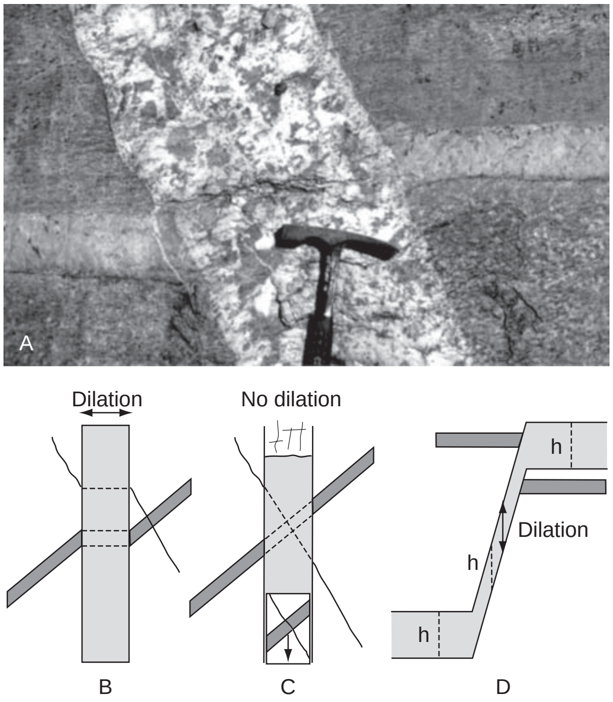


Fig. 4.50 (**A**) Quartz-rich layer in gneiss offset by the dilation caused by the intrusion of a granite pegmatite. (**B**) Simple dilation due to emplacement of magma offsets structures that intersect a dike at angles other than 90°. (**C**) If a dike is emplaced by stoping or replacement, no dilation occurs and structures in surrounding rock line up across the dike. (**D**) If a sill that is emplaced due to buoyancy steps up the section, the vertical thickness *h* is the same everywhere regardless of whether measured in the sills or in the connecting dike.

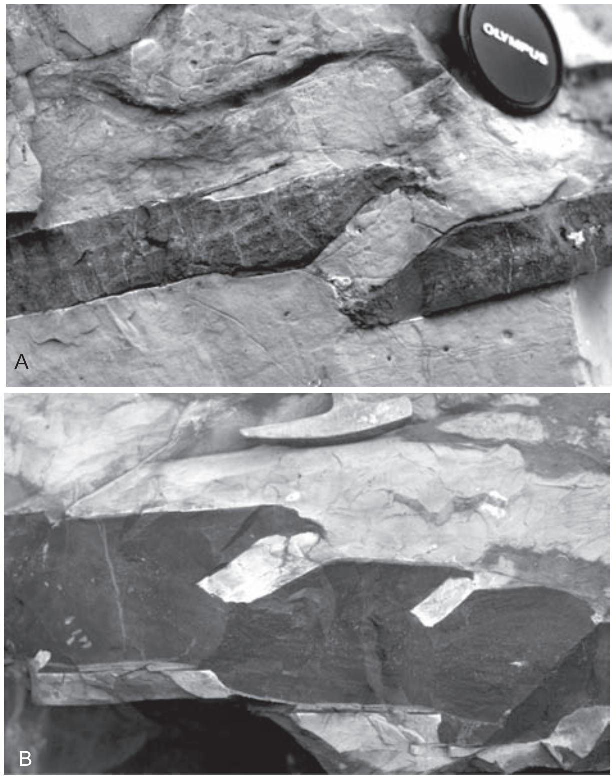


Fig. 4.51 En echelon dikes, bridges, and horns; Mount Royal, Montreal, Quebec. (**A**) Where two en echelon dike segments overlap, the septum of limestone between them was rotated to form a bridge as the dikes widened. Note that the termination of each en echelon segment turns to intersect the other at right angles. This gives the tip of each dike the appearance of the head of an eagle. (**B**) Continued widening of en echelon dikes causes bridges to rupture and form horns extending out from the contacts. Again, note the eagle-head shape of the horn's connection with the wall rock.

regardless of the dip of the sill. Also, if the sill climbs from one stratigraphic level to another, the dike connecting the two parts should also have the same vertical thickness (Fig. 4.50(D)).

En echelon dike sets, such as those that traverse southern New England (Fig. 4.49) are extremely common, but even continuous dikes may initially form as en echelon sets which, on widening, connect to form a single dike. Such dikes commonly contain *bridges* and *horns* (Delaney and Pollard, 1981). Figure 4.51(A) shows two en echelon basaltic dike segments intrusive into the Trenton limestone surrounding the Mount Royal stock in Montreal, Quebec. As the tip of each segment passes the tip of the other segment, the plane of the dike rotates so that it intersects the other dike at right angles. As the dike thickens during intrusion, the piece of country rock between the en echelon segments is rotated to form a *bridge* (Fig. 4.51(A)). As the dike thickens still more, the bridge is eventually broken, and horns are left protruding from the contacts (Fig. 4.51(B)). Such features provide a means of determining the magma flow direction. In the case of the dikes shown in Figure 4.51, magma must have flowed either into or out of the plane of the photograph, because until the bridge ruptured there was no possibility of lateral flow.

In many areas, steeply dipping subparallel dikes occur in vast numbers, forming what are known as *dike swarms*. Almost all of these are composed of diabase and are associated with outpourings of flood basalts. They have been emplaced in zones of crustal extension. Because of the great number of dikes in a swarm, their cumulative thickness represents considerable crustal strain. Many swarms are formed near divergent plate boundaries, with the dikes striking perpendicular to the direction of plate motion. For example, the dikes that traverse southern New England (Fig. 4.49) formed during the early Jurassic parallel to the initial spreading axis of the early central Atlantic Ocean. Similar dikes have continued to form in the vicinity of the Mid-Atlantic Ridge, as evidenced by the intense dike swarms in Iceland. Early spreading of the North Atlantic is recorded in an extensive dike swarm crossing northern England and the west coast of Scotland. Ancient directions of plate motion are presumably recorded by the many different dike swarms found in Precambrian shields (Chapter 15).

Steeply dipping dikes can also occur in *radial* patterns, commonly with a large intrusion or volcanic neck located at their focus (Fig. 4.43). Although the distribution suggests that the dikes emanated from the central body, the rocks in the dikes are commonly of a very different composition and age from those of the central body. Johnson (1970), who has made a detailed analysis of the stresses associated with intrusions, has shown that magma pressures within a vertical pipe-like body will not produce radial fractures. Instead, they will have the opposite effect and cause compression tangential to the chamber walls. However, the association of radial dikes about an intrusion is so common as to warrant an explanation. In some cases, the central body appears to have been localized by a preexisting intersection of dikes, rather than having been the cause of the radial dikes. In others, radial fractures, formed by doming of the roof of an actively rising magma chamber, have localized dikes that were then cut by the main body as it continued to rise. In still other cases, the presence of a body of magma has simply disturbed a regional stress pattern so as to cause tensional fractures to converge on the magma chamber and hence produce radial dikes.

4.8 RING DIKES, CONE SHEETS, AND CALDRON SUBSIDENCE

When a magma chamber intrudes close to Earth's surface, as happens in many volcanic regions, large cylindrically shaped crustal blocks may collapse into the chamber, producing a caldera. Magma forced up along the steeply dipping bounding fracture of the block may issue forth on the surface as volcanic material; that which solidifies beneath the surface forms an annular sheet known as a *ring dike*. These only rarely form a complete ring (Fig. 4.52); most consist of a

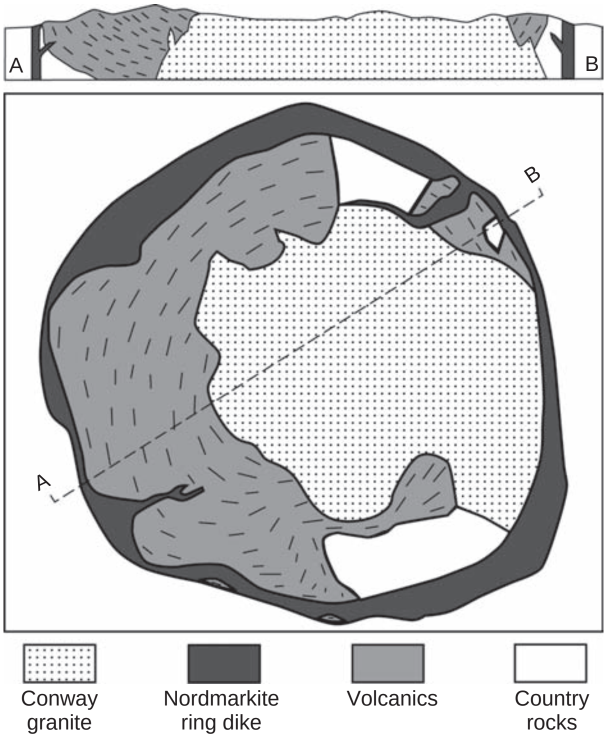


Fig. 4.52 Ossipee Mountains Ring Complex, New Hampshire. The border of the complex is marked by a continuous ring dike of nordmarkite (quartz syenite). Associated but earlier volcanic rocks subsided within the complex. They were intruded by the ring dike and the still younger Conway granite. Note that the granite is restricted to the confines of the ring structure. These Cretaceous age igneous rocks intrude Precambrian and Paleozoic rocks (simplified from Kingsley, 1931).

series of arcuate segments whose curvature nonetheless clearly defines a ring structure.

Ring dikes were first recognized in association with Tertiary volcanic centers of the west coast of Scotland, where they form steeply dipping arcuate bodies up to several hundred meters in width and many kilometers in length (Anderson, 1936). At some of these localities, for example on the Ardnamurchan peninsula (Fig. 4.53), these dikes are concentric about a center that moved laterally with time, producing intersecting ring structures of clearly different ages.

Another type of circular structure, also first recorded from the Tertiary igneous complexes of Scotland, has the form of a cone-shaped dike with its downward-pointing apex coinciding with the same centers as defined by ring dikes (Figs. 4.53 and 4.54). These are aptly named *cone sheets*. They are much narrower than ring dikes, rarely exceeding several meters in width, but they commonly occur in such great numbers that their aggregate thickness can be many hundreds of meters. They dip toward the center of the structure at approximately 45°; however, this angle tends to be steeper for sheets closer to the center. On Ardnamurchan (Fig. 4.53), where sets of intersecting cone sheets attest to the lateral movement of the center of igneous activity, downward projection of the sheets indicates a convergence at a depth of 5 km, which is thought to be the top of the magma chamber from which the cone sheets came (Fig. 4.54).

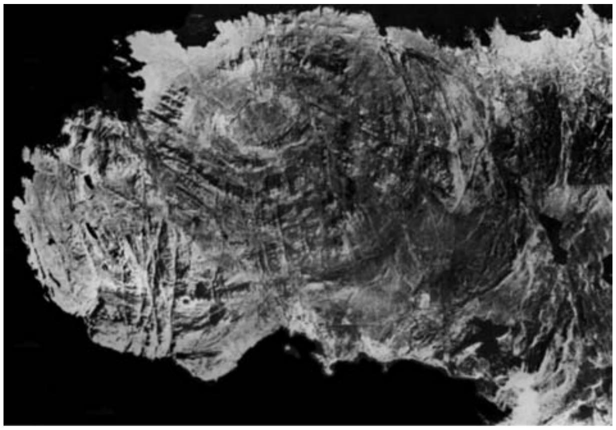


Fig. 4.53 Air photograph of the Tertiary Ardnamurchan ring complex, west coast of Scotland. Igneous activity about three successive centers produced intersecting sets of ring dikes and cone sheets. Only the last two centers are clearly evident at this scale. The youngest center is marked by the circular outcrop pattern of ring dikes. To the southwest of this center, a slightly earlier center is marked by arcuate cone sheets. Width of field is 20 km. (Photo courtesy of Directorate of Colonial Geological Surveys, London.)

Despite their close association in space and time, ring dikes and cone sheets appear strikingly different. Ring dikes are commonly composed of coarse-grained rocks exhibiting textures characteristic of slow cooling. Thus, they resemble plutonic bodies of rock. Cone sheets, on the other hand, are hypabyssal bodies and are typically composed of fine- to medium-grained rocks. Many cone sheets have chilled contacts against other cone sheets, indicating that they were formed in a series of successive injections. Multiple ring dikes, formed from successive injections, are also common, but slower cooling and lack of chilled contacts makes the intrusive sequence more difficult to determine. When there are many dikes the structure is referred to as a *ring complex*.

Ring dikes are formed from a wide variety of rock types, but granitic and syenitic ones (felsic) are the most common. In contrast, most cone sheets are composed of some type of basaltic rock. These differences probably arise from the contrasting forms and modes of intrusion of these bodies. Obviously, a block could sink more easily in a less dense felsic magma than it could in a dense basaltic one. Indeed, many crustal rocks would not sink at all in some basaltic magmas – they would float. Thus, felsic magmas are the most likely ones to form ring dikes, but their high viscosities would necessitate the dikes being moderately wide or the magma having to vesiculate and be disrupted in order to maintain the minimum flow rate necessary for intrusion (see Eqs. (3.16) and (4.2) and Problem 4.6). On the other hand, it is probable that only basaltic magmas, with relatively low viscosities, can flow fast enough in the thin cone sheets to intrude any great distance (see Problem 4.7 and 4.8).

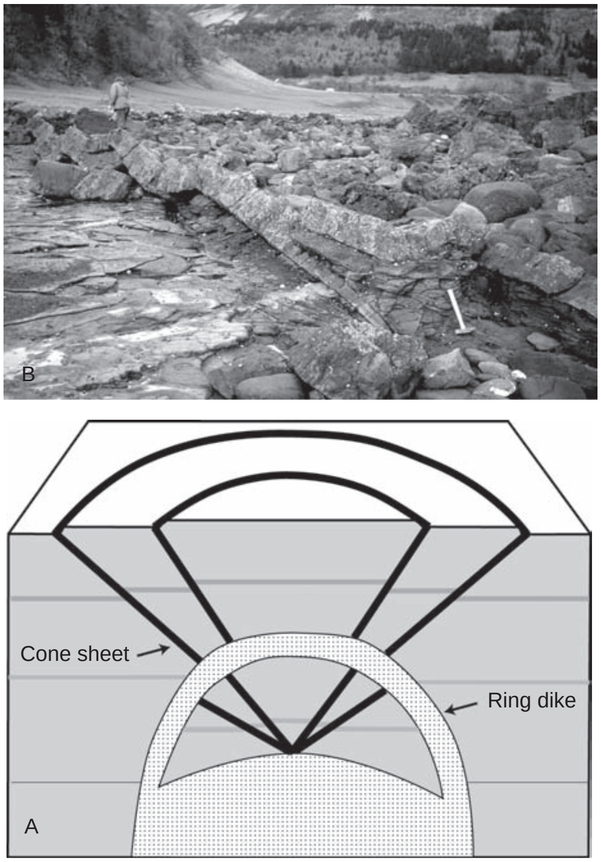


Fig. 4.54 (**A**) Block diagram of cone sheets and a ring dike. Cone sheets result from the filling of conical fractures that develop above a magma chamber in response to increased magma pressure. Stratigraphic markers lying above the cone sheets are lifted by the emplacement of magma. A ring dike forms when a decrease in pressure in the magma chamber removes support from the roof, which then collapses along steeply outward-dipping fractures; magma wells up along the fractures as the central block subsides. Stratigraphic markers within the central block are lowered by the emplacement of the ring dike. (**B**) Cone sheet on the southwest coast of the Isle of Mull, Scotland. Cone sheet dips away to the right toward the center of the Mull igneous complex.

Although the ring complexes of the Tertiary volcanic districts of Scotland are associated with a period of igneous activity that was predominantly basaltic in composition (first the flood basalts, followed by the basalts of the central volcanoes), the majority of ring complexes discovered in other parts of the world are associated with magmas of granitic and syenitic compositions, especially ones in which sodic amphibole and pyroxene have crystallized due to the high contents of alkalis (see Chapter 6). For example, a 1300-km-long belt of Jurassic-age ring complexes that extends in a southerly direction across Niger and Nigeria consists largely of granitic ring dikes, some of which are up to 60 km in diameter (Bowden and Turner, 1974). Important economic deposits of tin and niobium are associated with these complexes. A similar belt of granitic ring complexes, which is of Jurassic to Cretaceous age, occurs in the northeastern United States, extending 400 km in a southerly direction through the White Mountains of New Hampshire. The Ossipee ring dike (Fig. 4.52) is one of the most southerly of these complexes. Syenitic ring complexes of late Paleozoic age are common in several parts of Scandinavia, including those of the Oslo district, Norway, and the huge Khibina and Lovozera complexes of the Kola Peninsula, Soviet Union, which contain important deposits of phosphorus, niobium, titanium, zirconium, and rare earths. In contrast to the ring complexes of the Scottish Tertiary, these more felsic ones generally lack cone sheets, probably because of the viscosities of the magmas involved.

The formation of ring dikes and cone sheets has been analyzed by Anderson (1936), who has shown that a set of conical fractures can develop above a magma chamber if the magma pressure rises above the shear strength of the overlying rock. With a decrease in magma pressure, steeply outward dipping tensional fractures develop along which magma rises to form the ring dikes (Fig. 4.54). Anderson's solution of the stress distribution in the rocks overlying a magma chamber was based on rather special assumptions regarding the shape of the magma chamber (parabolic) and stress distributions along its walls. More realistic assumptions have been used in subsequent analyses (Roberts, 1970), but the general fracture pattern is still essentially that proposed by Anderson. The correlation of cone sheets with fractures developed during periods of pressure buildup within a magma chamber requires that the rock above the cone sheet be lifted by the intrusion of magma. This, in fact, is what is found, and in areas where there are many cone sheets, the central block can be many hundreds of meters above its initial position (Fig. 4.54).

The tension fractures in Anderson's analysis dip steeply outward, allowing the central block to sink easily. Although ring dikes are found with such dips, many dip vertically (Fig. 4.52) or even steeply inward, making it necessary for the subsiding block to adjust its shape, perhaps by faulting or folding, or for the ring dike to intrude by piecemeal stoping; that is, by having small blocks fall from the walls of the dike and sink through the magma. Subsidence of the central block can be demonstrated with most ring dikes, even when deeply eroded. The most common evidence is the presence of volcanic crater filling material found within the ring dike at levels below its original position (see, for example, the cross section of Ossipee ring dike, Fig. 4.52).

The amount of subsidence accompanying the emplacement of ring dikes can be great – hundreds or thousands of meters is not uncommon. The first area in which such large displacements were shown to exist was at Glen Coe, Scotland, where the name *caldron subsidence* was used to describe this phenomenon (Clough *et al.*, 1909). Part of the subsidence can take place along fractures that do not reach the surface of the Earth, in which case a cylindrical block may sink and force magma up along the bounding fracture into the space above the block (Fig. 4.54). The mechanism can be compared with the movement of a piston in a cylinder, and the farther the block sinks, the larger will be the plug-like body of igneous rock that will form above. Many plug-like bodies within ring complexes are formed in this way, as, for

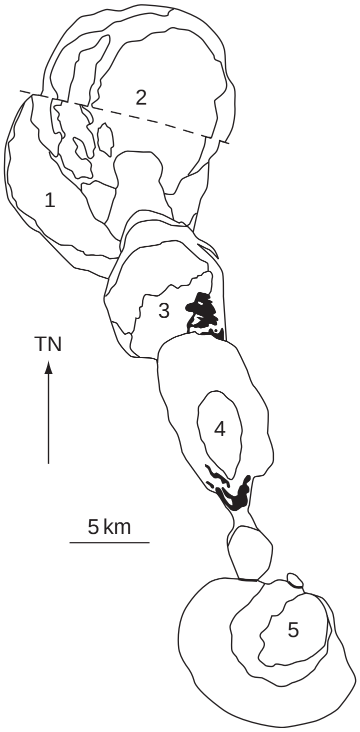


Fig. 4.55 Sara-Fier Complex, northern Nigeria, consists of five ring complexes. With the exception of the second ring complex, the center of igneous activity moved progressively southward with time. Such a linear belt of intrusions could result from the northward migration of a plate over a hot spot. Each of the complexes is composed mostly of granites, with only minor amounts of mafic rocks (black) being exposed in complexes three and four (simplified from Turner, 1963).

example, the Tertiary Mourne granites of Northern Ireland. The body of Conway granite within the Ossipee Mountain complex may also have formed this way, because the granite is completely confined by the ring dike (Fig. 4.52).

Many caldron subsidences that have had long histories, during which the underlying magma chambers have changed their composition and position, consist of complicated patterns of ring dikes of different rock types (Smith and Bailey, 1968). When subsidence occurs more than once, the sinking of the central block commonly results in the formation of successive ring dikes that become progressively younger toward the center, and if subsidence is great enough, the central block may sink so deep as to be no longer exposed. However, there are many exceptions, and ring dikes can be found within, and peripheral to, preexisting ones. Figure 4.55 shows five ring complexes from the Sara-Fier district of Nigeria (Turner, 1963), each of which shows younger ring dikes toward the center, and only in the oldest complex (1) are there volcanic rocks exposed that belong to the original subsiding block. The main center of igneous activity beneath a caldron subsidence may, with time, move laterally far enough to produce a new set of ring fractures, as at Sara-Fier (Fig. 4.55) or Ardnamurchan (Fig. 4.53), but in many cases, the movement is restricted to the confines of the original ring structure, with different parts of the cylindrical block sinking at different times (Fig. 4.38(A)). The Pilanesberg complex of South Africa (Mathias, 1974) provides an example of a cauldron subsidence that had a long history of successive intrusive events (Fig. 4.56), and it is left to the reader, in Problem 4.9, to work out its history.

4.9 DIATREME BRECCIA PIPES

A *diatreme* is a pipe-like body of breccia composed of fragments that are mainly derived from the neighboring country rocks but also include rocks from stratigraphically higher and lower levels; some fragments may even have a mantle origin (Fig. 4.57). The matrix of the breccia consists of finely comminuted fragments, but it may also be of igneous material that typically has low silica but high magnesium and potassium contents (alkaline magma; see Chapter 6). One particularly important type of diatreme is that which contains diamonds, and because of their discovery at Kimberley, South Africa, they are known as *kimberlite pipes*. Much of what we know about diatremes has come from the mining of these pipes. This weight of evidence may bias our ideas on the genesis of diatremes, which may have more than one mode of origin. Reviews of diatremes, especially those with kimberlite affinities, are given by Mitchell (1986) and Sparks *et al*. (2006).

Diatremes are believed to be the feeders to small, shallow volcanic craters known as *maars* (Fig. 4.58). These are explosion craters, which are surrounded by a low rim of coarse fragmental material that grades outward into a thin blanket of air-fall ash. The volume of ejected material is approximately equal to that of the crater, and the ejecta are composed of the rocks in which the crater develops. These ash deposits, however, are small, especially in the case of kimberlite diatremes, because no kimberlitic pyroclastic deposit has been found beyond the immediate vicinity of a crater. Rarely is new, or juvenile, volcanic material erupted, nor are lava flows produced. In short, the entire eruptive event appears to be driven by the explosion of gas. These craters are similar in form to those produced by meteorite impact, which also is an explosive event (Section 15.9). Indeed, there have been numerous arguments over the terrestrial versus extraterrestrial origin of some craters, and it was not until the discovery of diagnostic high-pressure minerals that could have formed only through meteorite impact that the question was resolved.

A subvolcanic rock type that may be related to diatreme breccias is known as *tuffisite*. As the name suggests, this rock resembles a tuff, containing lithic fragments, particles of pumice, broken phenocrysts, shards, and flattened fiamme, but it is of intrusive origin. In contrast with diatremes, tuffisites may form rather irregular bodies with branching veins of breccia invading the country rock. Bodies of breccia that are commonly associated with *porphyry copper deposits* may be related to tuffisites (e.g. Landtwing *et al*., 2002). The evidence for involvement of juvenile magma in most tuffisites is clear, whereas in many diatremes it is lacking. Furthermore,

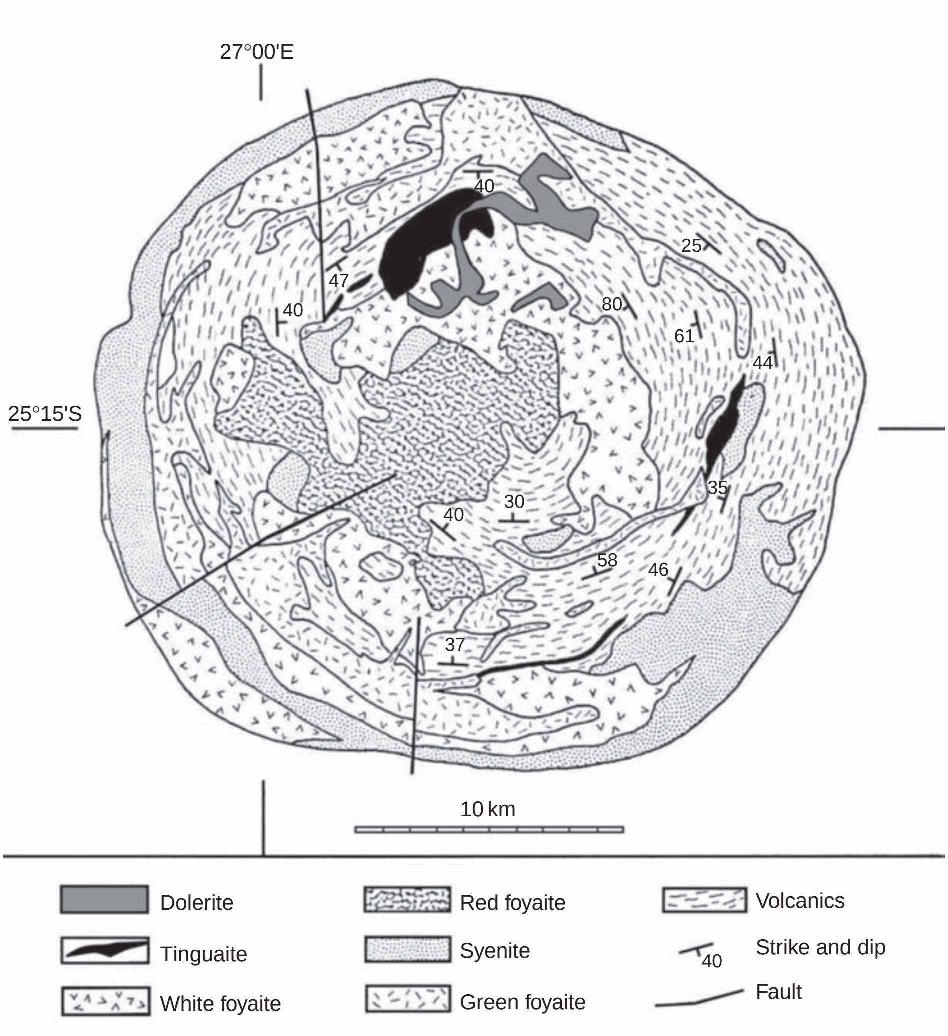


Fig. 4.56 The 1.25-Ga Pilanesberg Complex lies entirely within the 1.95-Ga Bushveld Intrusion, South Africa (Fig. 15.17). For definitions of rock types, see Chapter 6. Intrusive contacts dip steeply except those on the tinguite body, which is a cone sheet. The legend is not in chronological order (see Problem 4.9). (Simplified from Mathias, 1974.)

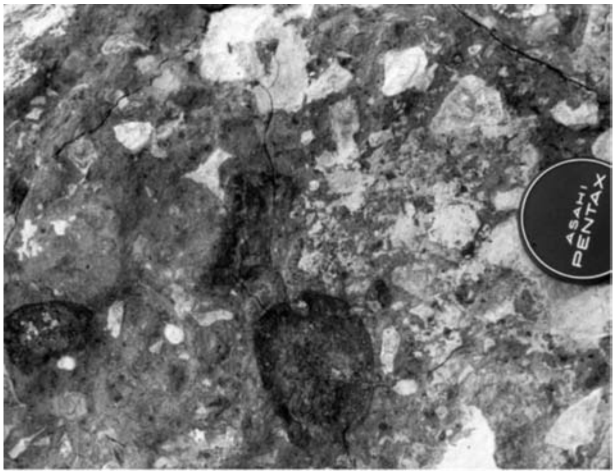


Fig. 4.57 Diatreme breccia from Ile Bizard at the western end of the Cretaceous Monteregian igneous province, Quebec (No. 2 in Fig. 4.74). The rock consists of an assortment of fragments of lower Paleozoic sedimentary rocks, Precambrian basement rocks, and dark mantle-derived nodules composed of olivine, pyroxene, and pyrope-rich garnet. All of the crustal rock fragments are hydrothermally altered and hence are light colored.

Fig. 4.58 A maar volcano formed by gas explosion with no eruption of lava. This maar is located on the east side of the Afar rift system, Ethiopia. The crater, which is deep enough to expose the water table in this arid region, cuts flood-basalt flows and has subsequently been offset by normal faults resulting from the motion of the Arabian plate to the northeast (top left of photograph). To view this crater in its regional setting, go to 11° 58

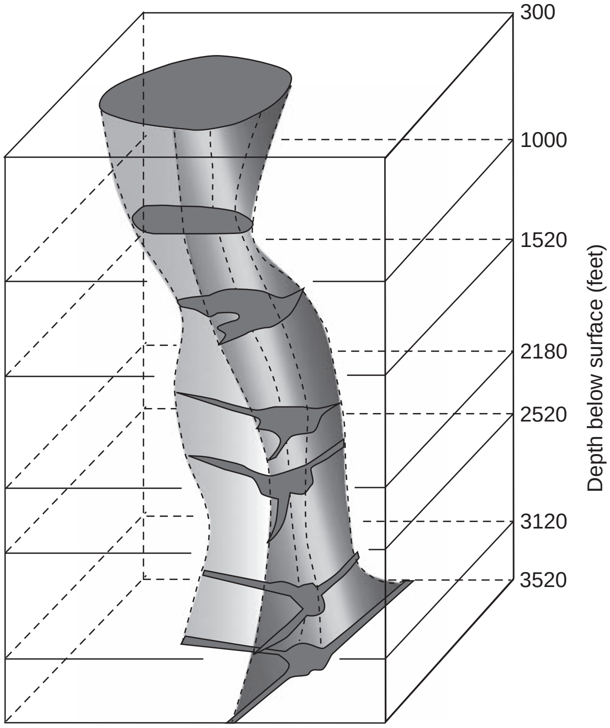


Fig. 4.59 Cross sections through the diatreme pipe of the Kimberley diamond mine, South Africa. This body changes from carrot-shaped near the surface into a number of intersecting dikes at depth (drawn from data given by Williams, 1932).

tuffisites contain highly silicic magmas in contrast with the silica-poor ones of diatremes.

Diatremes are carrot-shaped bodies with nearly circular cross sections. Contacts dip inward at angles as shallow as 60° but 80–85° is more common. Diatremes, consequently, can have only limited vertical extent. The diameter near the surface is at most ~700 m, decreasing to ~100 m at a depth of ~2000 m. Below this is a root zone in which the shape of the pipe becomes irregular, commonly branching or spreading into one or several intersecting dikes (Fig. 4.59). The amount of igneous matrix increases through this zone until massive igneous rocks are encountered in the dikes. Connections with bodies larger than dikes have not been found. The contact of a diatreme with surrounding rocks is sharp, with little evidence of subsidiary veining or deformation, nor is there any thermal metamorphism. The type of country rock also is unimportant in determining the shape of a diatreme, as is clear where a pipe has cut straight through a sequence of layered rocks of highly variable hardness (shale and flood basalt, for example; Williams, 1932). Country rocks are not domed up by a diatreme, nor do they exhibit significant slumping into the pipe. The emplacement of a diatreme resembles a cleanly drilled hole connecting the top of a dike or intersecting dike set with the surface of the Earth, the drilling products being left to fill the hole.

Fragments in diatremes are angular, but some rounding is evident. In some tuffisites, such as that in West Cork, Ireland (Coe, 1966), fragments are not only rounded but some are polished or etched on one side, implying that the orientation of the fragment remained fixed long enough for the surface to be sandblasted by the passage of gas and fine particles. Fragments range from meter-sized blocks down to submicrometer particles with little or no evidence of sorting. Although thermal metamorphism is lacking, most fragments in the breccia are hydrothermally altered, suggesting that at least steam had passed through the breccia (Fig. 4.57). Some diatreme breccia, especially at shallow levels, exhibits crude layering dipping ~35° toward the center of the pipe, which Sparks *et al*. (2006) consider consistent with an avalanche origin.

One of the most striking features of the breccia is the thorough mixing that fragments have undergone. Fragments of mantle-derived rocks can be found juxtaposed against fragments of crustal rocks (Fig. 4.57). This mixing cannot be attributed solely to upward turbulent transport, because the same breccia may contain fragments of country rock that are derived from levels stratigraphically well above the existing level of exposure of the breccia. This is well illustrated by the Cretaceous diatreme breccia at the 1967 World's Fair site in Montreal, Quebec. At its present level of exposure this diatreme cuts Ordovician shales, limestones, and sandstones, rocks that form a large fraction of the fragments, along with a smaller admixture of underlying Precambrian granites and gneisses. But the breccia also contains some blocks of Devonian sandstone which today are exposed nowhere else in the Montreal area. Indeed, it is necessary to travel several hundred kilometers south into New York State to find these rocks. During the Cretaceous, however, these rocks must have covered the Montreal area. Based on stratigraphic thicknesses in New York State, these Devonian blocks must have descended over 1700 m in the diatreme breccia.

Despite the thorough mixing of fragments in diatremes, there may be concentric or vertical zonation in the abundance of fragments from different stratigraphic levels. McGetchin and Ullrich (1973), for example, showed that the abundance of mantle-derived fragments increased upward in the ejecta blanket surrounding a maar on an island off the west coast of Alaska; that is, the abundance of deep-seated fragments increased with time as deeper parts of the diatreme were tapped. Based on the distribution of these fragments, McGetchin and Ullrich calculated eruption velocities of 500 m s^{-1}; that is, supersonic (velocity of sound in air about 330 m s^{-1}).

Although diatremes are clearly formed by gaseous explosions, there is considerable debate over their ultimate origin. What, for example, is the source of gas? How are fragments transported and mixed? Is the presence of magma essential, and if so, need it be of a particular composition? We will examine each of these questions briefly. For a more complete discussion, the reader is referred to chapter 4 in Mitchell (1986).

There are two schools of thought on the origin of the gas. One claims that it is primarily of mantle origin (Wyllie, 1980), rising either as a separate phase or in solution in a magma that exsolves the gas on approaching the surface

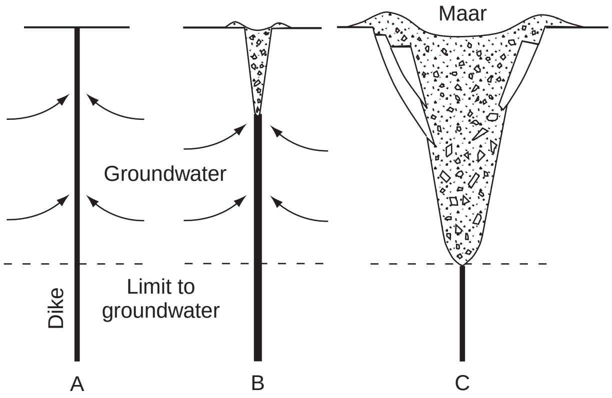


Fig. 4.60 Three successive stages in the development of a hydrovolcanic breccia pipe. (**A**) Groundwater or surface water gains access to a conduit filled with magma. (**B**) Conversion of groundwater to steam causes disruption of the magma and brecciation of the surrounding rock; brecciation commences at the surface and works downward. (**C**) Brecciation of the rock continues to depths where groundwater is still capable of gaining rapid access to the conduit. A shallow crater, known as a maar, forms above the pipe, and it is surrounded by a blanket of ejecta derived largely from the country rocks.

(Skinner and Marsh, 2004; Sparks *et al.*, 2006). The other claims that the gas is derived from groundwater or surface water that comes in contact with, and is heated by, magma (Lorenz, 1975; Sheridan and Wohletz, 1983; Kurszlaukis *et al.*, 1998).

Long-lived accumulations of vapor of sufficient volume to form a diatreme cannot exist in the mantle, because buoyancy would cause it to diffuse upward. Considerable quantities of vapor could, however, be exsolved from an ascending and crystallizing mantle diapir of magma. Sufficient pressure might be generated in this way to propagate a fracture to the Earth's surface and generate a diatreme; successive eruptions might then tap deeper and deeper sources. However, where igneous rocks are exposed in the lower parts of diatremes, they are neither vesicular nor disrupted into shards. They do contain biotite or amphibole, which indicates that the magma did at least contain some dissolved volatiles.

The difficulty in deriving sufficient gas from the mantle led to the idea that surface waters may play a role. Explosive volcanic activity resulting from groundwater or surface water leaking into a magmatic system is well documented from many volcanic regions; it is known as *hydrovolcanism* or *phreatomagmatism* (Sheridan and Wohletz, 1983). Deep craters filled with pyroclastic debris and slumped blocks can form in this way as shown in Figure 4.60. Diatremes are claimed to form preferentially in regions where there are good aquifers, such as permeable sedimentary rocks. Even when formed in regions where such rocks are absent, diatremes tend to be located at the intersection of major fractures that could serve as conduits for groundwater. According to the hydrovolcanic hypothesis, diatremes are formed when groundwater that has been transformed to steam by coming in contact with magma rises along a fracture and explodes through to the surface of the Earth. The brecciation would start at the surface and work downward eventually reaching the magma. Indeed most hydrovolcanoes have craters that contain fragments that in large part are juvenile, but blocks of country rock can slump in from the walls. Most diatremes, in contrast, consist largely of fragments of country rocks. The hydrovolcanic hypothesis does not by itself explain the presence of mantle-derived fragments in some diatremes.

Regardless of source, gas has undoubtedly been involved in the formation of diatremes. The passage of gas through particles brings to mind the fluidization process that was discussed in connection with ash flows (Section 4.4). Following the publication by Reynolds (1954), fluidization gained instant popularity as a hitherto unrecognized geologic process. One of its most convincing applications was in explaining the origin of diatreme breccias. Modeling experiments successfully reproduced many of the features seen in diatremes (Woolsey *et al.*, 1975).

As the velocity of gas passing through a body of close-packed particles is increased, a stage is reached where the gas pressure between the particles becomes high enough to lift and slightly separate the particles. At this stage, the system becomes *fluidized*, and the material changes from being dense and rather rigid to less dense (about 20% less) and extremely fluid. It is this dramatically increased mobility that could explain the thorough mixing of fragments in diatremes. It is important to emphasize that the flow of gas through a fluidized system is not great enough to expel material from the conduit but simply to raise the particles into a fluidized state. This is precisely what is found in most diatremes – the breccia, while thoroughly mixed, is still essentially in its correct stratigraphic position; rarely does a diatreme clear its throat and replace the breccia with magma. Another important feature of fluidization is that it must start at the top and work down because it involves the lifting of particles. Diatremes, then, if formed by fluidization, would have to start at the Earth's surface and work downward along some fracture up which gas was escaping. The gas either could be from a magmatic source or from groundwater heated by a magma.

Fluidization, however, is not as simple to apply to diatremes as was first thought. Industrial fluidization involves the movement of particles of a limited size range. Diatremes contain fragments that vary in size over six orders of magnitude. Experiments using material of variable grain size (Wilson, 1984) indicate that fluidization in such material is complicated. First, there is no single gas velocity at which fluidization of the entire system occurs. The gas flux necessary to lift micrometer-sized particles is very much less than that necessary to lift centimeter-size particles. If gas velocities are increased so as to lift the large particles, large bubbles of gas start ascending through the system; also, finer particles are transported upward, resulting in a sorting of the particles. Sorting is noticeably lacking in most diatremes. In addition, the constant agitation in a fluidized system would be expected to produce more rounding of the fragments than seen in most diatreme breccias.

Finally, there is the question of associated magmatic rocks. Although igneous rocks are not found with all diatremes, they are found with so many that they probably occur in the root zones of all diatremes and would be exposed if erosion went deep enough. More important, most diatremes are associated with a small class of igneous rocks that are referred to as alkaline (see Chapter 6) and particularly to alkaline rocks that are very poor in silica, such as kimberlite. These have associated mantle-derived fragments and diamonds that must come from depths of at least 150 km (Section 15.7). This association indicates that hydrovolcanism, although possibly a contributory factor to brecciation, cannot be the ultimate cause of diatreme activity – it is a phenomenon peculiar to alkaline magmatism. Yet the lack of vesiculation or magma disruption in the root zone of diatremes indicates that if the gas required for brecciation is to have exsolved from the alkaline magma, it must have done so at greater depth so that any signs of vapor-phase separation would be obliterated.

Sparks *et al*. (2006) have proposed a model for diatremes associated with kimberlite magmas that resolves some of the paradoxes associated with these breccias. Although the precise composition of kimberlite magma is difficult to determine because of contamination with many different rock fragments and hydrothermal alteration, kimberlites undoubtedly have extremely low silica contents (20–30 wt%) compared with more common magmas (50–75%). In addition, they are rich in the volatile components CO_2 and H_2O. These factors combine to give the magma a viscosity of 0.1–1 Pa s at 1200–1400 °C and 2–2.5 GPa, the conditions under which kimberlite magmas are thought to be generated, based on included high-pressure minerals (e.g. diamonds). Such a low viscosity in dikes that are typically <1 m wide gives Reynolds numbers (Eq. (3.17)) that are above the critical value for laminar flow. Using Equation (3.24), Sparks *et al*. (2006) estimate turbulent flow velocities of 4–16 m s^{-1} for kimberlite magma in typical dikes. Because many kimberlites contain diamonds, they must rise from a depth of ~200 km. Their thermal entrance length must therefore be at least 200 km. Equation 3.14 for the thermal entrance length for a pipe can be modified to apply to a dike in turbulent flow:

$$L_T \approx \frac{uw^2}{k} \tag{4.3}$$

where u is the average magma velocity in the dike, w is the half width of the dike, and k is thermal diffusivity (5×10^{-7} $m^2\ s^{-1}$). For an average velocity of 4 m s^{-1}, a dike could not be less than 0.3 m wide if the entrance length were 200 km. This agrees well with the observation that kimberlite dikes are typically 0.3 to 1 m wide.

Rapidly rising gas-rich kimberlite magma in a narrow dike would exsolve gas as it approached the surface. With typical gas contents, kimberlite magma can be expected to consist of at least 64% bubbles (the fragmentation threshold; see Section 3.4) at depths of several kilometers. Exsolution of this gas and its expansion increases the velocity of the magma by a factor of four. Gas expansion causes the temperature of the magma to decrease. This, in turn, causes the magma to crystallize, which releases heat and reduces the cooling effect of decompression. Near the surface, however, the mixing of cool rock fragments lowers the temperature of the mixture and explains the lack of thermal metamorphism around kimberlite diatremes.

When a kimberlite dike first breaches the surface, the magma is expected to be at a pressure greater than one atmosphere. As long as the cross-sectional area of the vent is less than some critical value that keeps the magma pressure above one atmosphere, the flow is described as *overpressured* or *choked*, a condition favoring formation of shallow explosion craters. As the vent widens, the pressure of the exiting magma decreases until it becomes one atmosphere and the venting mixture is then *pressure adjusted*. This is analogous to a person changing from coughing or sneezing to breathing normally. When the eruption reaches this stage, the lithostatic pressure on deeper parts of the conduit walls will exceed the magma pressure and rock bursts will occur, especially if aided by groundwater pressure gradients. This widens the conduit and undermines overlying rocks, which can slump into the conduit (Fig. 4.60). Once flow has become pressure adjusted, the gas flow rate is not great enough to eject material from the vent and instead a fluidized bed of material is formed, in which particles become thoroughly mixed. As the supply of magma decreases, the gas flux decreases and only smaller rock particles can be moved by fluidization. Surges of activity would result in renewed fluidization and further wall rock bursts and slumping of earlier deposited material, which could explain some of the complex zonation found in diatremes.

4.10 LACCOLITHS

A *laccolith* is a concordant intrusive body with a flat base and domed roof – that is, a mushroom-shaped body (Figs. 4.42 and 4.61). They form at shallow depths, mostly within 3 km of the Earth's surface and in relatively flat-lying stratified rocks. The emplacement of magma domes the overlying rocks upward in contrast to the strata above and below a sill which are essentially parallel. Laccoliths are thought to be fed

Fig. 4.61 Maverick Mountain, west Texas, is a laccolith of alkali gabbro intrusive into limestone; its topographic expression corresponds almost exactly with its igneous form.

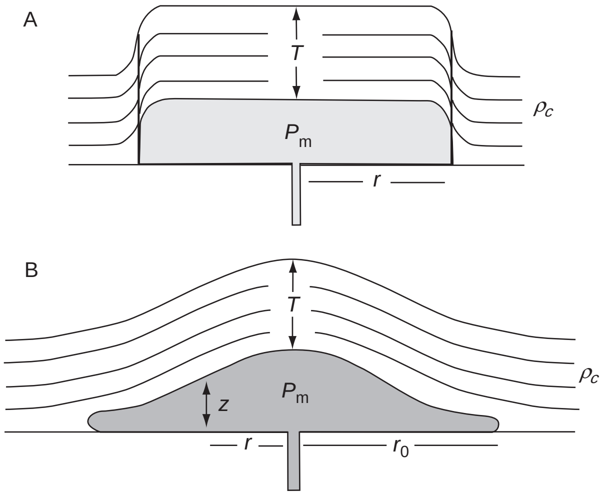


Fig. 4.62 Two possible models for the emplacement of a laccolith. In (**A**) the overlying rocks are lifted largely as a block that moves on a cylindrical ring fracture; the dimensions are those used in Gilbert's (1877) analysis of these structures. In (**B**), the overlying rocks are domed upward by the magma as soon as it extends laterally far enough to supply sufficient leverage on the overlying rocks; the dimensions are those used in Pollard and Johnson's (1973) analysis of these structures. See text for discussion.

by a central conduit, but exposures rarely allow this to be determined. Dikes and lateral feeders may also occur. They are commonly formed from magmas with intermediate silica contents, but examples ranging in composition from basaltic to rhyolitic are known. Because of their thickness, laccoliths contain coarser-grained rocks than do sills.

Laccoliths were first described by Gilbert in 1877 from the Henry Mountains of Utah. His account of the fieldwork, description of the bodies, and insight into their mode of formation make for some of the most fascinating reading in all of geology (see Yochelson, 1980). Gilbert recognized that both sills and laccoliths occur in the Henry Mountains, but sills always had areal extents of less than 1 km^2 whereas laccoliths were greater than 1 km^2. From this he concluded that sills were the forerunners of laccoliths but that they required a minimum area before they exerted sufficient force to dome upward the overlying rocks. He also drew a correlation between the size of a laccolith and its depth of burial at the time of formation, the larger ones forming at greater depth.

Despite the frontier conditions under which Gilbert's fieldwork was done, his analysis of the formation of a laccolith has a distinctly modern flavor, and it emphasizes some of the important forces that must be involved. For the purpose of the analysis, he treated a laccolith as a piston-like body that lifted the overlying rock along a ring fracture, as shown in Figure 4.62(A). He reasoned that the force lifting the roof came from the pressure in the magma, P_m, which acted on the area of the roof of the laccolith, πr^2; that is,

$$\text{lifting force} = \pi r^2 P_m$$

This force would be opposed by the body force resulting from the mass of the overlying rocks. If the overlying rocks have a thickness T and density , then

$$\text{weight of roof rocks} = \pi r^2 T \rho_c g$$

But $T\rho_c g$ is the lithostatic pressure at depth T. We can therefore write

$$\text{weight of roof rocks} = \pi r^2 P_l$$

where P_l is the lithostatic pressure. In lifting the overlying rocks, magma also has to overcome the frictional resistance along the ring fracture resulting from the shear strength τ. This stress acts on the surface area of the ring fracture and gives a resisting shear force of

$$\text{shear force} = 2\pi r T \tau$$

Gilbert argued that for a laccolith to be able to lift the overlying rocks, the lifting force of the magma has to be equal to or greater than the sum of the forces resulting from the mass and shear strength of the overlying rocks; that is,

$$\pi r^2 P_m \geq \pi r^2 P_l + 2\pi r T \tau$$

which, on rearranging, gives

$$(P_m - P_l) r \geq 2T\tau \tag{4.4}$$

This relation indicated to Gilbert that a minimal radius was necessary for a laccolith to overcome the shear strength of the overlying rock. It also accounted for the radius of laccoliths increasing with increasing depth of emplacement.

We cannot expect Eq. (4.4) to give a precise description of the parameters governing laccolith formation. It does not, for example, take account of the force necessary to bend the overlying rocks or the pressure necessary to intrude the magma between the two contacts, regardless of the amount of lifting. A more sophisticated analysis of the forces involved in the formation of these laccoliths has since been done by Johnson (1970) and Pollard (1973), who have also made a careful study of the actual structures in the country rocks at the periphery of these laccoliths.

Johnson and Pollard, like Gilbert, found that sills in the Henry Mountains range in area up to 1 km^2, whereas laccoliths are greater than this; sills are from 0.5 to 10 m thick and laccoliths from 10 to 200 m. Sills do not extend out as thin wedges but terminate abruptly, their blunt ends being surrounded by deformed and sheared country rock as shown in Figure 4.47(B). The periphery of laccoliths is also blunt, with any one of the configurations shown in Figure 4.63 being common. In plan view, the periphery may be smooth but can consist of finger-like extensions, the spacing of which can be explained on the basis of Rayleigh–Taylor instabilities (see Section 3.11) similar to those for the regular spacing of diapirs (Pollard *et al.*, 1975). The diorite, of which the Henry Mountain laccoliths are composed, is porphyritic and may have behaved as a Bingham liquid with a yield strength of about 100 Pa. This may have prevented the magma spreading too far laterally.

The important point of these field relations is that laccoliths are not large sills but distinct bodies formed by a very different type of displacement of the overlying rocks. Sills expand until they have an area of about 1 km^2, at which stage

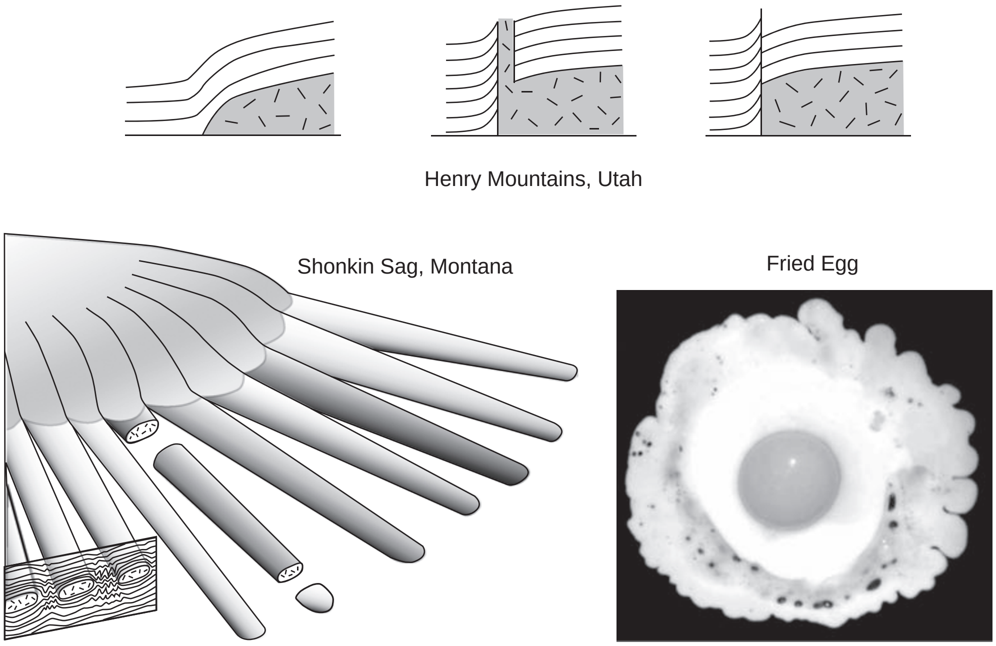


Fig. 4.63 Three different terminations found at the peripheries of laccoliths in the Henry Mountains, Utah (Johnson and Pollard, 1973), and the fingered periphery of the Shonkin Sag laccolith in the Highwood Mountains, Montana (drawn from data presented by Pollard *et al.*, 1975). These fingers result from a Rayleigh–Taylor instability on the advancing edge of the laccolith. A similar instability develops on the advancing edge of a fried egg, but because the egg's white is more viscous than the hot grease into which it advances, the fingers are relatively broader than those on the laccolith, where the magma has a lower viscosity than the intruded country rocks.

continued lateral spreading becomes more difficult than lifting of the roof. Johnson and Pollard's analysis of these structures, combined with experimental modeling, indicates that the intrusion of a sill acts like a lever on the overlying rocks. When a sill is small, the mechanical advantage is also small, because the distance from the point of magma influx to the leading edge of the sill is short. The magma therefore continues to flow laterally as the overlying rocks are simply lifted upward. As the sill expands the leverage increases, until at a critical stage of growth, the rocks at the leading edge and above can be bent rather than simply being lifted, and a laccolith is born.

The extent to which a sill will expand before reaching the critical stage at which a laccolith develops depends not only on the depth of emplacement but also on the physical properties of the magma and the country rocks. A fluid basaltic magma might extend laterally a greater distance than the porphyritic diorite of the Henry Mountains, and a more viscous granitic one might extend less. Also, the more competent the overlying rocks are, the more a sill will have to expand before being able to bend and dome the roof rocks.

Pollard and Johnson (1973) give the following equation relating the height z that the roof of a laccolith will be lifted as a function of distance from its center r, where r_0 is the radius of the body (see Fig. 4.62(B)):

$$z = \frac{3(P_\mathrm{m} - \rho_\mathrm{c} g T)}{16BT^3}\left(r_0^2 - r^2\right)^2 \qquad (4.5)$$

where B is the elastic modulus, which is defined as $E/(1 - \nu^2)$, E being Young's modulus and ν Poisson's ratio (see Johnson, 1970, for further explanation).

It is valuable to inspect Eq. (4.5) qualitatively to gain an appreciation of the factors controlling the shape of a laccolith. The first terms in parentheses give the difference between the pressure of the magma, which tries to lift the overlying block, and the load pressure of the overlying rock, which tries to make the roof sink. The difference, then, is the resultant upward pressure on the roof rocks. The terms in the second set of parentheses define the length of the lever arm; the longer it is, the greater will be the height that the roof can be lifted. Note that this factor, being raised to the fourth power, is extremely important. The denominator expresses the rigidity and thickness of the overlying rocks. Clearly, as either of these factors increase, the height to which the roof can be domed must decrease. Note that the thickness T of the overlying rocks is raised to the third power and thus plays an important role in determining the thickness of the laccolith (see Problem 4.10). The equation, of course, cannot describe completely the shape of a laccolith because, as shown in Figure 4.63, considerable faulting and peripheral vertical dike intrusion accompanies the simple flexing of the roof rocks.

In summary, laccoliths are restricted to shallow depths because room for the magma is provided by the doming and bending of the roof rocks. Their size and form is controlled largely by the stratification and physical properties of the surrounding rocks.

4.11 LOPOLITHS AND LAYERED INTRUSIONS

Lopoliths are large conformable, saucer-shaped intrusions (Figs. 4.42, 15.17, 15.18). Their form superficially resembles

Fig. 4.64 Layered ultrabasic rocks on Hallival in the Tertiary Rum layered intrusion on the northwest coast of Scotland. The prominent steps on the mountain mark alternating ~50-m-thick layers that grade from peridotite (olivine and pyroxene) at the base to more resistant troctolite (plagioclase and olivine) at the top. Changes in the composition of the minerals through the layers indicate that each main layer results from the intrusion of a fresh batch of magma (Emeleus *et al.*, 1996; Emeleus, 1997).

Fig. 4.65 Millimeter- to centimeter-scale layering in gabbroic rocks of the Kiglapait lopolith, Labrador (Morse, 1969). Graded layering is brought out by concentrations of dark-weathering olivine at the base of layers passing up through pyroxene-rich zones to light-colored plagioclase-rich rock at the top of layers.

that of an inverted laccolith, but these two intrusive bodies are quite unrelated. Laccoliths are never more than a few kilometers in diameter; lopoliths may be over a hundred kilometers in diameter. Laccoliths form near the surface of the Earth where they make room for themselves by doming upward the surface rocks; lopoliths are so large that they can affect the entire crust, and room for the intrusion is made as much by bowing down the floor as it is by lifting overlying rocks. Laccoliths can be formed from a variety of magmas; most lopoliths are formed from basaltic or more mafic magmas. The feeder conduit to a lopolith is commonly hidden by the intrusion itself, but large regional basaltic dikes have been found to connect with the base of some bodies (Fig. 15.18). Roof rocks, where present, may include relatively unmetamorphosed flat-lying sedimentary rocks or even volcanic rocks formed from the same magma that produced the lopolith.

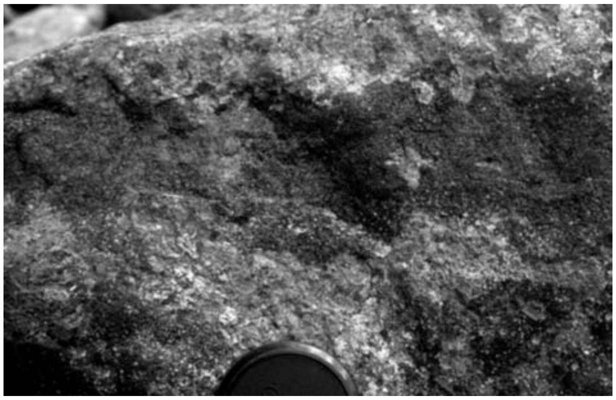


Fig. 4.66 Decimeter-thick layer of chromite in peridotite, Stillwater layered igneous intrusion, Montana (Hess, 1960).

Because of their enormous size, lopoliths cool very slowly, allowing a variety of features to develop that are rarely present or are only poorly developed in small intrusions. The most striking of these is prominent flat or gently dipping layering (Wager and Brown, 1968; Cawthorn, 1996). The layers can vary in thickness from hundreds of meters (Fig. 4.64) down to millimeters (Fig. 4.65) and extend laterally for hundreds of kilometers. The layering, which generally conforms to the saucer-like shape of the body, is due primarily to variations in the abundance and composition of minerals but can also be brought out by preferred orientations of crystals. Some layers can have concentrations of minerals that are great enough to be of economic value. One of the principal sources of Cr, for example, is from layers rich in chromite (Figs. 4.66, 15.14). So striking is the layered appearance of rocks in lopoliths that these intrusions are commonly referred to as *layered intrusions* or layered basic (mafic) intrusions, the term basic meaning that the rocks were formed from basaltic magma (see Chapter 6). Layering does occur in some other types of intrusive bodies but never on the same extensive scale as in lopoliths.

The largest known lopolith is the 2.0-Ga-old Bushveld Complex of South Africa, which has an area of 66 000 km^2 (Fig. 15.17). Over 8 km of layered rocks are exposed between its floor and roof. The complex is lobate in plan, with rocks dipping at shallow angles toward the center of each lobe. The body probably developed by the coalescence of four lopoliths, which may have been initiated by meteorite impact (Section 15.9). Another well-exposed lopolith is the 1.2-Ga-old Muskox intrusion in the Northwest Territories of Canada. It forms an elongate canoe-shaped intrusion with rocks dipping toward its axis at about 30° (Fig. 15.18). Because this body plunges northward, both the floor and roof are exposed.

Fig. 4.67 Rhythmic modal layering in norite (plagioclase + orthopyroxene) of the Bjerkreim–Sokndal layered intrusion, southwest Norway (Wilson *et al*., 1996). Each layer is pyroxene rich at its base (right) and plagioclase rich at the top (left). The layering shows several unconformities that indicate that the top is to the left. The layers in the center of the photograph fill a channel that cut into the underlying layers and then they, in turn, were truncated by the overlying norite.

Fig. 4.68 Block of plagioclase-rich rock (anorthosite) in a layer of norite that overlies a layer of anorthosite, from which the block was probably derived. The Bjerkreim–Sokndal layered intrusion, southwest Norway.

A 150-m-wide, 60-km-long dike feeds into the base of the intrusion; from here the body widens gradually as it extends 50 km northward, reaching a maximum width of 11 km at the point where it disappears beneath the roof rocks. A magnetic anomaly indicates that the intrusion continues northward for at least another 30 km beneath the cover rocks. Within the exposed section, a thickness of 1.8 km of layered rocks constitutes the bulk of the intrusion. The so-called "Great Dyke of Zimbabwe" (Rhodesia), a 530-km-long by 5.6-km-wide body of layered rocks, may have formed in a Muskox-like body that was later downdropped along two faults that parallel the trend of the original body. Many lopoliths are only partially preserved. The 3.2-Ga-old Stillwater Complex in Montana, for example, is only the lower two-thirds of a former lopolith, the upper part having been eroded away or uncomformably overlain by Paleozoic rocks. The remaining 8 km of layered rocks, however, attests to the former existence of a gigantic body. Perhaps one of the most unusual lopoliths is that at Sudbury, Ontario. This 1.7-Ga-old lopolith was emplaced in what is interpreted to have been a huge meteorite crater measuring about 40 km in diameter (Section 15.9). Layered sheet-like bodies are probably present beneath most mid-ocean ridges. Where ancient spreading axes have been obducted onto the continent, layered gabbros are found at the base of the crustal section overlying mantle rocks (Section 15.2).

It is not by design that all of the examples of large lopoliths cited here are of Precambrian age except for the bodies beneath mid-ocean ridges. In the Phanerozoic, only relatively few, small lopoliths have formed. This is one of a number of petrologic facts that indicate that the principle of *uniformitarianism* (the present is the key to the past) may not be completely applicable to all processes throughout geologic time. The restriction of large lopoliths to the Precambrian may indicate that the thermal character of the Earth has changed with time, a point to which we will return in Section 15.8.

Lopoliths are so large that they must take considerable time to form, and as we will see in Chapter 15, their rocks contain abundant petrologic and geochemical evidence for magma chambers that were replenished by multiple pulses of magma. The layering in lopoliths commonly exhibits structures that provide evidence for these multiple pulses. For example, layers are commonly truncated or scoured in much the same way that beds of sand and gravel are scoured in river channels (Fig. 4.67). These scours commonly have an orientation that suggests magma was convecting down the wall and across the floor of the lopolith (Wager and Brown, 1968). Fragments of layers can be ripped up and redeposited in other layers (Fig. 4.68). Aligned crystals within layers also attest to the flow of magma (Fig. 4.69). Layers deposited on sloping surfaces can become unstable and slump and form folds that resemble those formed during soft sediment deformation (Fig. 4.70). Some layers contain exceptionally long crystals of, for example, olivine, plagioclase, or hornblende that grow perpendicular to the layering (Fig. 4.71). During the growth of such minerals, the magma must have been stationary (Wager and Brown, 1968).

Many of the features exhibited by the layers in lopoliths are so similar to those in sedimentary rocks that the layering was originally interpreted to have formed by sedimentation of crystals from magma (Wager and Brown, 1968). Just as gravity plays a role in settling out sedimentary grains, denser crystals were believed to have settled more rapidly than less dense ones in magma. This resulted in the formation of graded layers (Fig. 4.72). Graded layers are particularly well developed in channels (Fig. 4.67). In sedimentary sequences, turbidity currents commonly scour a channel and then deposit a graded layer. By analogy, channels filled with graded layers in lopoliths were interpreted to have formed from dense slurries of crystal-rich magma.

As we will see in Chapter 14, crystal settling in magma is not a simple process. Even if the crystals are denser than the magma, they may not separate if the magma has a yield strength. And in many basaltic magmas, plagioclase crystals are neutrally buoyant and so could not sink under any circumstances. These problems lead to the consideration of other mechanisms that could form layers. Processes involving nucleation and crystal growth mechanisms and diffusion processes have resulted in other possible mechanisms for the formation of layers. Because lopoliths are so large, they solidify slowly. During this extended time, a pile of crystal mush can undergo compaction and the crystals can dissolve and reprecipitate. Many of the textures of the rocks in lopoliths are thought to form by these processes. Figure 4.70 shows layers of gabbroic rocks overlying peridotite that have clearly been deformed by slumping in the Rum layered intrusion of northwest Scotland. If we examine the boundary between the underlying peridotite and the layered gabbros (inset in Fig. 4.70), we see clear evidence that the gabbro has been replaced by the peridotite. The boundary is extremely irregular, yet the layering in the overlying gabbro shows no irregularity. It appears that residual magma

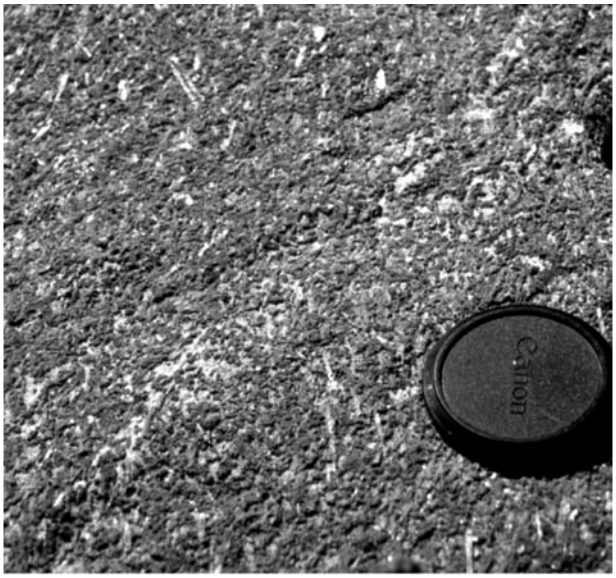


Fig. 4.69 Looking down on a pyroxene-rich layer in the Stillwater Complex, Montana. Elongate pyroxene crystals are preferentially aligned (up and down in photograph), which probably resulted from the flow of magma.

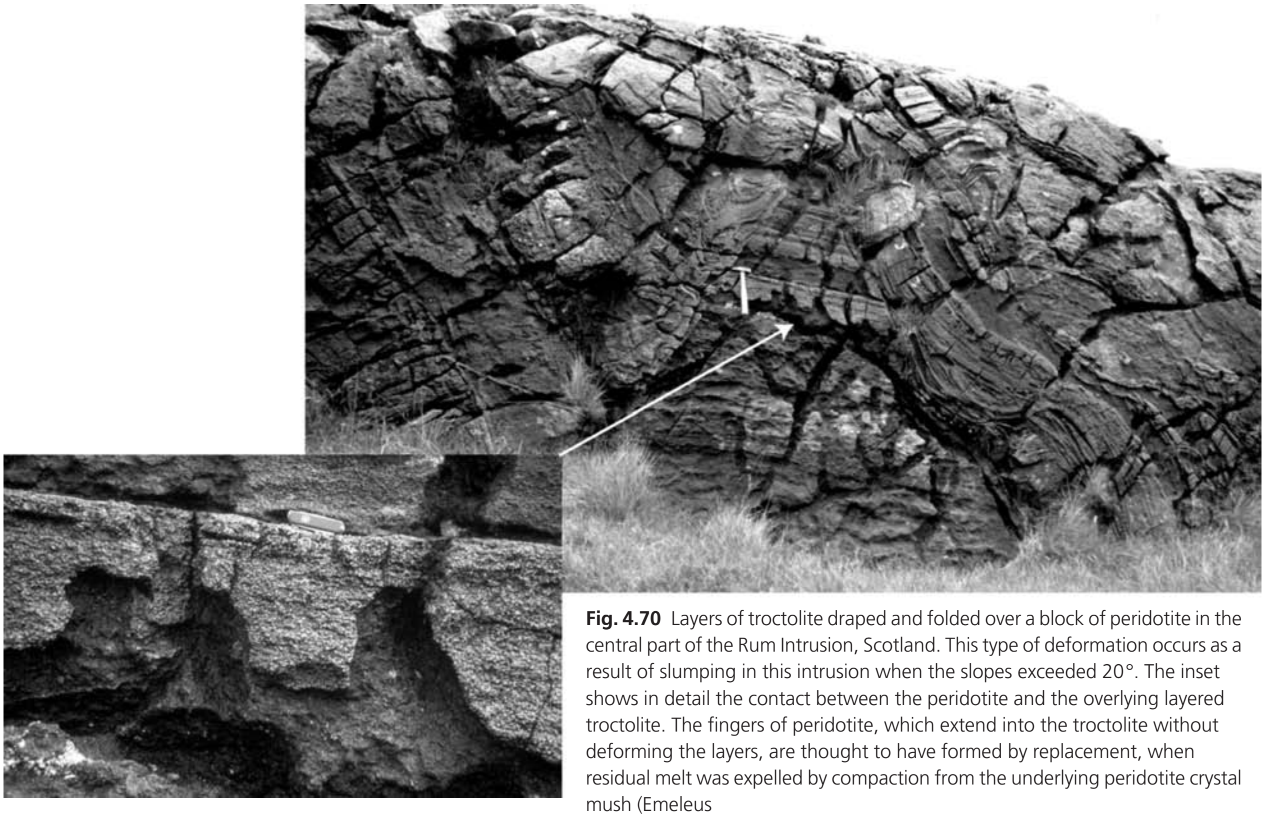


Fig. 4.70 Layers of troctolite draped and folded over a block of peridotite in the central part of the Rum Intrusion, Scotland. This type of deformation occurs as a result of slumping in this intrusion when the slopes exceeded 20°. The inset shows in detail the contact between the peridotite and the overlying layered troctolite. The fingers of peridotite, which extend into the troctolite without deforming the layers, are thought to have formed by replacement, when residual melt was expelled by compaction from the underlying peridotite crystal mush (Emeleus

Fig. 4.71 Decimeter-long dendritic olivine crystals (dark) that grew upward in layers of feldspathic peridotite overlying finer-grained peridotite near Harris Lodge on the Isle of Rum (O'Driscoll *et al*., 2007). Rocks with olivine crystals growing approximately perpendicular to the layering are said to have a *harrisitic* or *crescumulate texture* (Wager *et al*., 1960).

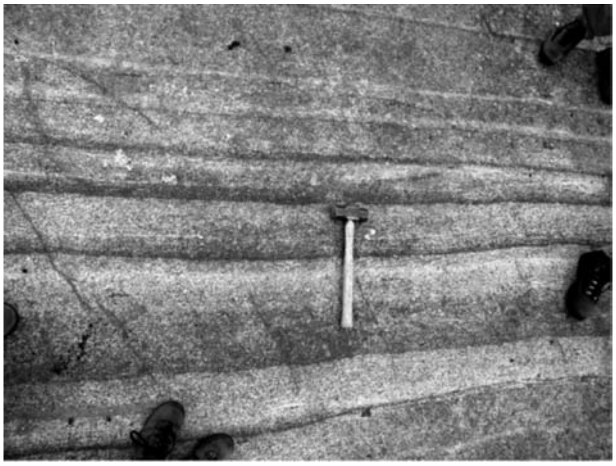


Fig. 4.72 Graded layering in the gabbro of Stavanger, Norway. Each layer grades from pyroxene rich at the base to plagioclase rich at the top. The next layer begins with an abrupt increase in the pyroxene content, and the sequence repeats.

expelled from the underlying magma has risen into the gabbro and replaced it. These problems will be discussed in Chapters 14 and 15.

4.12 STOCKS

A *stock* is defined as a plutonic body having an areal extent of less than 100 km^2, whereas a *batholith* is one whose area is greater than this. The division is arbitrary and thus at first may appear to have little significance. For instance, could a stock be the top of a batholith that had not yet been eroded deeply enough to expose its full extent? If the difference between these bodies depended on the whims of erosion, the definitions would have no more significance than to indicate qualitatively the present area of exposure; this could be done more simply with a statement of the present number of square kilometers exposed. There is, however, a distinct group of intrusions that fit the definition of a stock and whose areal extent does not vary significantly with depth of erosion. This is not to say that the term *stock* has not been used to describe bodies that are the upper tips of a batholith. But we ignore these here and take them up in Section 4.13.

Stocks, as defined above, have two distinct modes of occurrence. One is as bodies of diorite, granodiorite, or granite (see Chapter 6) intruded beneath large calderas. Here they clearly form an integral part of caldron subsidence, providing the source of magma for eruptions of ash-flow tuffs and a medium into which the surface blocks can sink. The other mode of occurrence involves rocks that are referred to as alkaline (see Chapter 6) and which range in composition from peridotite, through gabbro, to nepheline syenite and even granite. These stocks are typically associated with rift valleys on continents.

In Section 4.4, calderas were described as cylindrical blocks that subside along steeply dipping ring fractures into bodies of magma, the subsidence typically being accompanied by the eruption of ash flows. Numerous studies of old, deeply dissected calderas fully substantiate the deep structures surmised from the surface exposures. Furthermore, these studies indicate that most calderas are the sites of multiple collapses, and that the magma chamber into which the surface block sinks makes room for itself by displacing downward the preexisting rocks along the same, or closely related, ring fractures. These relations, which were first worked out by Clough *et al*. (1909) for the caldron subsidence of Glen Coe, Scotland, have since been found with many near-surface magma chambers.

The ring fracture of a caldera extends downward and becomes the wall of the underlying magma chamber. Its steep dip and cylindrical shape produces a plug-like body or stock (Fig. 4.73(A)). The roof of the chamber is the base of the subsided block of the caldera. The floor is formed by rocks that are forced downward into a deeper magma chamber, the floor rocks moving in a manner similar to that of a piston in a cylinder. No room problem exists with this type of emplacement; the floor rocks and deeper magma simply exchange places. This type of emplacement is possible even if no overlying caldera is formed, as shown in Figure 4.73(B). All movement of magma and country rocks is probably in response to buoyant forces that are developed locally; that is, the sinking block must be denser than the ascending magma. While the magma cools and crystallizes, a light fraction may accumulate near the top of the chamber, where it may bow up the roof to form a *resurgent dome* on the floor of the caldera, as happened in Yellowstone (Fig. 4.38). After complete solidification, sections of the intrusive body may sink (it would now be denser) as new magma wells up around the piston to form a new chamber beneath the roof rocks. The subsiding floor rocks may therefore bear testimony to repeated intrusions and subsidences.

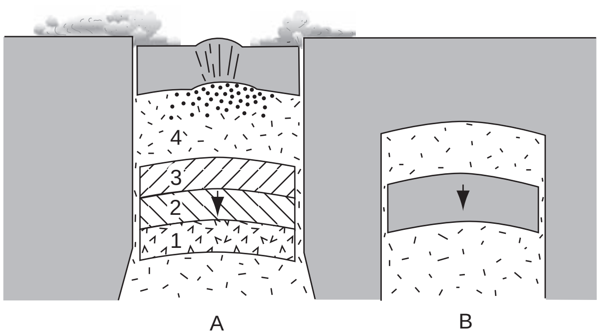


Fig. 4.73 Development of a stock by the sinking of the roof rocks along a ring fracture up which magma rises as the block descends. On the surface, a stock may have an associated caldera (**A**), but it need not break through to the surface at all (**B**).

A stock that forms in this way has certain characteristics. It is circular to oval in plan and has vertical or steeply dipping walls and flat or gently sloping roof and floor. The transition from roof to wall takes place over a distance of meters. Contacts with the wall rocks are strongly discordant. The wall rocks, however, show little sign of deformation and certainly are not bent upward. Some stocks may be surrounded by additional ring fractures along which sheared rocks or dikes may be present. These intrusions are generally composite or multiple, with successive fillings of the magma chamber forming a roughly horizontal stratigraphy as the floor subsides. The rocks are restricted in composition to ones that form from magmas with densities less than those of the intruded rocks. The magmas crystallize to massive rocks except near the feeder ring, where a strong foliation or shearing may be developed.

Stocks composed of alkaline rocks differ from those described above in several important ways, the difference being attributable mainly to the modes of emplacement. First, the magmas involved in alkaline stocks may be denser or lighter than the intruded rock, so buoyancy, at least at the local level, is not important. Many of the rocks are derived from magmas that have come directly from the mantle and which isotopic studies (Section 13.4) reveal have not had a long residence time in the crust. These intrusions rarely show evidence of an obvious floor, either through direct exposure or through geophysical means, indicating that they are of considerable vertical extent. They are strongly discordant. Unlike the stocks associated with caldron subsidence, however, these bodies have contacts with surrounding rocks that are marked by considerable deformation, which may involve plastic flow of the country rocks or intense stoping. The rocks within these stocks commonly exhibit prominent layering or foliation, which dips steeply and is parallel to the contacts. Disruption of the layering attests to vigorous movement of magma during solidification. Multiple injections of strongly contrasting magmas (gabbro-syenite) are also common.

The Cretaceous Monteregian intrusions of southern Quebec provide excellent examples of most of the features characteristic of this type of stock (Philpotts, 1974). These intrusions are located at the intersection of two rift systems, the St. Lawrence and Ottawa valleys, which have been active since the late Proterozoic (Fig. 4.74). A wide range of alkaline rocks, including peridotite, carbonatite, gabbro, nepheline syenite, and granite, form stocks that are circular to elliptical in plan and have vertical contacts (Fig. 4.74). The smallest stock, Mount Johnson, is nearly circular in plan, whereas some of the larger ones are lobate and consist of stocks within stocks, the relative ages being indicated by truncation of layering. Almost all of the stocks exhibit concentric, steeply dipping layering that was built up by deposition or growth of minerals on the walls of the intrusions. This sequence of solidification is evident where channels have been cut in the earlier-deposited layers and then filled in by later layers (Fig. 4.75).

The stocks in the western half of the Monteregian province were emplaced in flat-lying Ordovician sandstone, limestone, and shale of the St. Lawrence Lowlands (Fig. 4.74). These sedimentary rocks provide ideal markers with which to monitor the deformation associated with magma emplacement. No diapiric doming is associated with any of these stocks (Fig. 4.77). Instead, the sedimentary rocks remain flat lying to within a few meters to tens of meters of the contact. Here they begin to dip in toward the intrusion, with the dip increasing toward, and becoming vertical at, the contact (Fig. 4.77). Toward the contact the rocks show increasing signs of plastic deformation and indeed of melting, with refractory layers remaining as brittle material that is segmented by the flow of the surrounding more fusible rock (Fig. 4.76). The contact zone with the igneous rocks of the stock is therefore marked by a zone of breccia with an igneous matrix formed from the partial fusion of the country rocks. Such a liquid is referred to as *rheomorphic* to distinguish it from the magma that rose from depth and brought with it the heat necessary to fuse the country rocks. It will be recalled from Section 3.11 that this zone of melted country rock is precisely what is required in the magmatic ascent model of Marsh (1982).

The stocks in the eastern half of the Montergian province, which are generally larger than those in the western part, were emplaced in the folded Appalachian rocks, so the effects of intrusion on the geometry of the surrounding rocks is not as easily determined. Zones of rheomorphic breccia are, nonetheless, present, and where the country rocks dip steeply and strike into the intrusions at a high angle the details of the contact reveal selective melting of the country rocks (Fig. 4.76). For example, layers of refractory quartzite or slate (which is metamorphosed to corundum-bearing rock at the contact) project into the zones of rheomorphic breccia for several meters before being broken off by the viscous flow of the breccia.

Toward the inside of the rheomorphic breccia zone, fragments of the refractory country rocks become rounded by dissolution in the rheomorphic magma and eventually disappear. This results in a zone, which is a few tens of meters wide, of rather streaky, inhomogeneous rock of granodioritic composition. This granodioritic rock, formed from a mixture of mantle-derived magma and melted crustal rocks, is

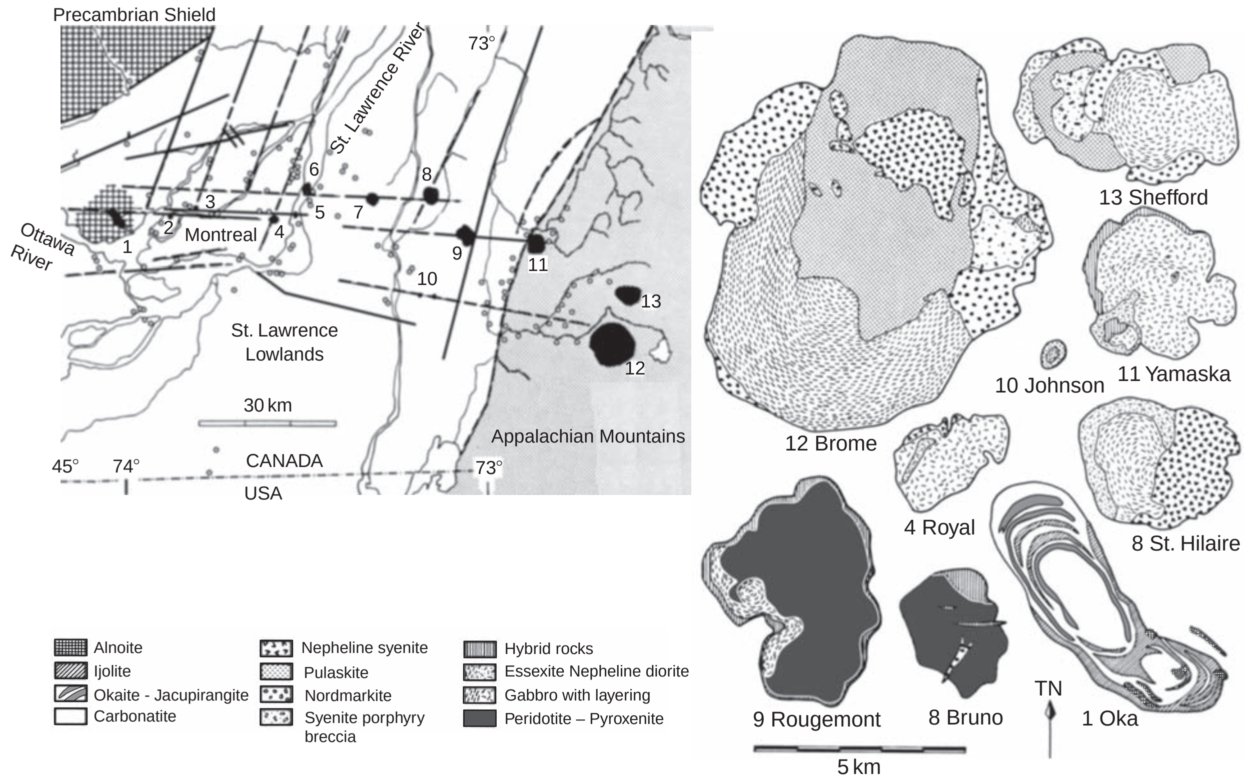


Fig. 4.74 The Cretaceous Monteregian intrusions of southern Quebec form a linear belt of stocks extending eastward from the Ottawa graben across the St. Lawrence Lowlands into the Appalachian Mountains. The important intrusions, which are shown in black, are from west to east: 1 Oka, 2 Ile Bizard, 3 St. Dorothée, 4 Mount Royal, 5 St. Helen's Island, 6 buried intrusion, 7 Mount Bruno, 8 St. Hilaire, 9 Rougemont, 10 Mount Johnson, 11 Yamaska, 12 Brome, 13 Shefford. Each of the main intrusions forms a vertically elongated cylindrical body; larger ones consist of several intrusions that give the bodies cuspate outlines. Most intrusions exhibit compositional layering that parallels the contacts and dips steeply inward. For explanations of rock types, see Chapter 6. (After Philpotts, 1974.)

markedly different in composition from those in the main part of the intrusions that have inherited their chemical characteristics only from a mantle source.

A room problem exists with the emplacement of this type of intrusion. Although the rheomorphic breccia zone indicates that magma was making room for itself by partially melting and stoping country rock (Figs. 4.76, 4.77), it is unclear where the melted country rocks have gone. Mixing with the main body of magma is possible, but the relatively uncontaminated nature of most of the igneous rocks in these stocks indicates that this amount would have to have been small. The melted country rocks would have had a much higher silica content than most of the magmas rising from depth; thus, they would have had a lower density. The rheomorphic magma, therefore, may have been able to rise buoyantly along the margins of the intrusions and collect at the top of the stock. If this happened, the evidence, unfortunately, has been eroded away. A seismic survey in the St. Lawrence River carried out in preparation for construction of the 1967 World's Fair in Montreal, revealed what is undoubtedly a Monteregian intrusion that did not quite rise to the present erosion surface (Fig. 4.77(B)). The seismic survey shows flat-lying shales overlying the intrusion are completely undisturbed by the intrusion. We can conclude therefore that intrusion must be some process of stoping.

Although stocks formed in association with caldron subsidence and with alkaline rocks have similar outcrop patterns, details of the contacts and internal structures indicate two very different modes of emplacement. The caldron-subsidence stock involves passive upwelling of buoyant magma while a cylindrical block of dense rock sinks. The alkaline stock, on the other hand, involves a vigorous thermal event, with heat being brought from considerable depth to high levels in the crust, where it plays a significant role in making room for magmas that can range in composition from ultramafic to felsic. In closing, it should be mentioned that although we have referred to alkaline magmas forming these hot, thermally active stocks, they can also form stocks through the caldron subsidence process, as long as densities are appropriate.

4.13 BATHOLITHS

A *batholith* is a plutonic igneous body that has an areal extent greater than 100 km^2 and was originally thought to extend to considerable depth. In some definitions, batholiths are described as having no floor; although this is not possible, it emphasizes the fact that they do have considerable vertical extent, and their roots may well be hidden. The

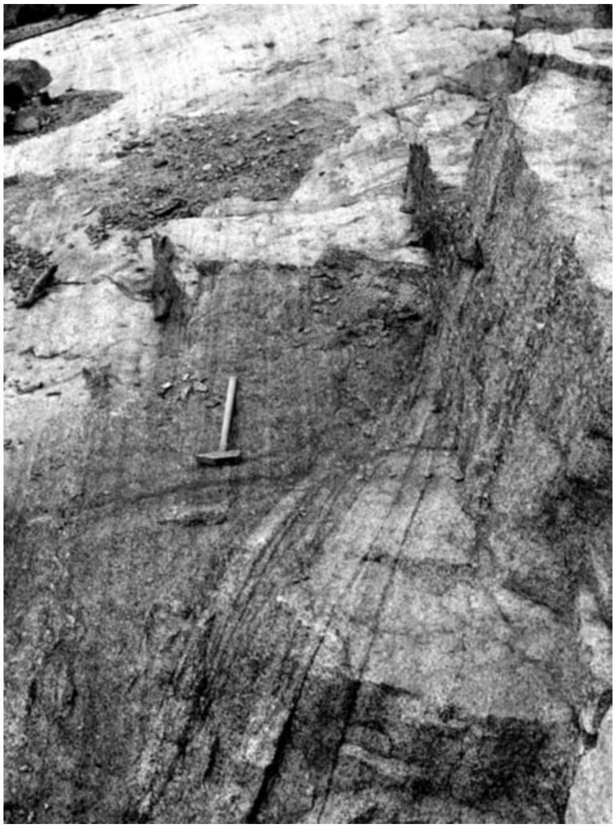


Fig. 4.75 Vertically dipping layers in essexite (nepheline gabbro), Mount Johnson, Quebec. The contact of this nearly cylindrical intrusion (see No. 10 in Fig. 4.74) is to the left of the outcrop shown here. The layers parallel the contact. Flow of magma eroded a channel into the layers of crystals, which were deposited on the wall of the intrusion. Where the younger layers enter the channel they thicken, with the result that the irregularity in the layering caused by the channel is rapidly eliminated following the deposition of several new layers. (After Philpotts, 1968.)

Fig. 4.76 Rheomorphic breccia at the southern contact of the Brome Mountain Monteregian intrusion, Quebec (see No. 12 in Fig. 4.74). Heat from the intrusion of an alkali gabbro magma fused country rocks that have low melting temperatures; the more refractory country rocks, which include highly aluminous rocks (thin dark bed), quartzite (white fragment beneath hammer), and early mafic dikes (dark fragments), were broken and pulled apart by the flow of the newly melted rocks. Rocks in the lower right are vertically dipping metamorphosed country rocks that project into the zone of rheomorphism. Evidence of selective partial melting and flow can be seen by comparing the two layers on the left side of this projection of country rock.

name *batholith* itself comes from the Greek word *bathos*, meaning deep. Batholiths are composed of relatively low-density rocks, ranging in composition from quartz diorite to granite, with granodiorite being the most abundant (see Chapter 6). They are never composed of gabbro or ferromagnesian-rich rocks, but some Proterozoic batholiths are composed of anorthosite, a rock consisting essentially of plagioclase (Section 15.8) and also having a low density relative to surrounding rocks. This restriction in the composition of rocks to ones of low density is evidence that batholiths are emplaced by the buoyant rise of magma through the crust. The low density of the rocks also makes it possible to use gravity surveys to determine the extent of batholiths at depth. Based on such evidence, McCaffrey and Petford (1997) claimed that most granite batholiths do not extend to great depths, but tend to be rather tabular bodies. For example, geophysical surveys over the rapakivi granite batholiths of southern Finland, which have produced large quantities of building and facing stone (Fig. 4.79(E)), show that they cannot be more than 5 to 10 km thick (e.g. Elo and Korja, 1993). At Barre, Vermont, which claims to be the "Granite Capital of the World," two hundred years of quarrying has started to expose the base of the batholith, which has provided the famous Barre Gray granite. The geophysical evidence thus indicates that batholiths have a form similar to that of a molar tooth (Lynn *et al*., 1981).

Many batholiths are extremely large. The Coast Range batholith, which extends from Alaska through the entire length of British Columbia, is 600 km long and 100 to 200 km wide. The Coastal batholith of Peru is 1100 km long and 50 km wide (Fig. 4.78). These batholiths are synorogenic and are elongated parallel to the mountain chain. Postorogenic batholiths are smaller and less elongated, such as the White Mountain batholith in the northern Appalachians of eastern North America, which is elliptical in plan, measuring 45 km by 30 km. Large batholiths are not single intrusive bodies but consist of many smaller batholiths and stocks, each of which may contain distinct rock types. For example, nine different episodes of igneous activity are recognized in the Coast Range batholith in British Columbia, and eight have been found in the Sierra Nevada batholith in California.

Because of their great size, batholiths can be found in many different depth environments, where they exhibit quite different characteristics. Buddington (1959), for example, drew attention to the difference between batholiths emplaced in the *epizone* (shallow), *mesozone* (intermediate), and *catazone* (deep). Although sharp divisions between these zones cannot be drawn, batholiths emplaced near the surface are quite distinct from those emplaced at depth. The main differences result from changes with depth in the behavior of the crust to long-term stresses. In Chapter 1, we saw that the

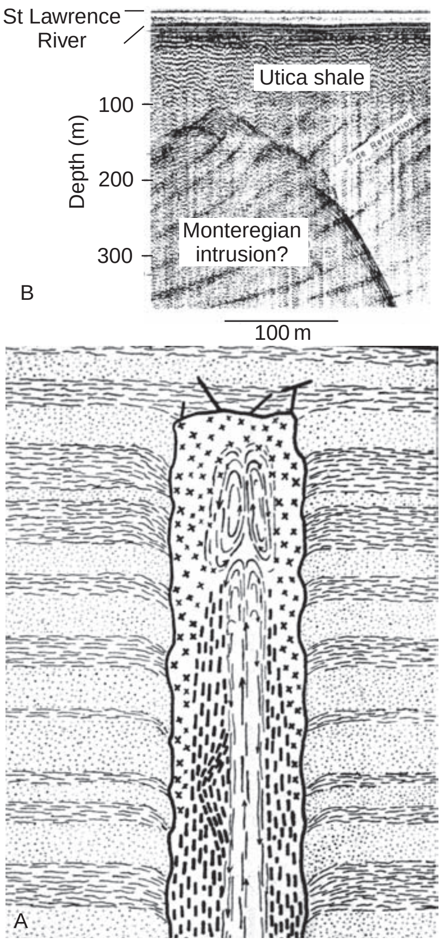


Fig. 4.77 (**A**) Cross section of a typical Monteregian stock intruded into flat-lying Paleozoic sedimentary rocks. Near the stock, country rocks dip inward and become partly melted at the contact, which is marked by a zone of rheomorphic breccia (Fig. 4.76). Within the intrusion, compositional layering parallels the walls. Active convection of the magma (rising in the core and descending along the walls) is indicated by discordances in the layering (Fig. 4.75). Convection was responsible for renewing the heat that was lost to the surroundings by conduction and in melting of country rocks. A separate convecting layer of less dense, felsic magma may have existed in the upper part of the body. (After Philpotts, 1968.) (**B**) A seismic reflection survey in the St. Lawrence River carried out in preparation for the 1967 World's Fair in Montreal revealed what is most likely a Monteregian intrusion (No. 6 in Fig. 4.74), which does not quite intersect the present erosion surface. The seismic record shows that the flat-lying Utica shale is not deformed above this intrusion. Emplacement must therefore have involved stoping.

crust can be divided into an upper brittle zone and a lower ductile zone. Some batholiths are sufficiently large to span both zones, and smaller ones may certainly have passed through both during ascent. Batholith emplacement at great depth involves ductile movements in the surrounding rocks, whereas that at shallow levels involves brittle failure of the country rocks. Because batholiths are so large, their ascent through the crust raises regional isotherms, which in turn causes the brittle–ductile transition to move to shallower depths than normal.

Batholiths emplaced at high levels in the Earth's crust are markedly discordant. Exposures of their roofs are numerous, as are roof pendants and xenoliths. *Roof pendants* are isolated bodies of country rocks within batholiths whose point of attachment to the roof has been removed by erosion. The orientation of structures in roof pendants is consistent with that in country rocks outside the batholith. *Xenoliths* (foreign rock), on the other hand, are unattached blocks of country rock, and they may have any orientation. The abundance of xenoliths near contacts indicates that the surrounding rocks were brittle and that stoping was a major means of magma emplacement. Fine-grained chilled margins are found on the earliest phases of intrusion, but little chilling is seen with later intrusive phases. The rocks are typically unfoliated, and phenocrysts have shapes similar to those found in volcanic rocks. Small gas cavities lined with euhedral crystals (*miarolitic cavities*) are common, but *pegmatites*, veins of extremely coarse-grained granitic rock, are generally lacking.

Batholiths emplaced at great depth in the crust are generally concordant. Contacts with country rocks are steep, and chilled margins and xenoliths are lacking. Large sheets or *screens* of country rocks may occur in the batholith parallel to the contacts. Evidence of small-scale replacement of country rocks at the contact is common. The igneous rocks tend to be foliated and recrystallized. Pegmatites and associated veins of a fine- to medium-grained sugary-appearing rock known as *aplite* are common, but miarolitic cavities are absent.

Batholiths exhibit vertical and lateral compositional variations. These variations may be abrupt, as for example between bodies formed during different episodes, or gradational within a single body. In general, the proportion of quartz diorite is greater at depth than it is near the surface, and successive intrusions become more silicic with time. Within a single intrusive body, rocks tend to be more silicic and hydrous toward the top. Compositional zoning in thick ash-flow tuffs shows that magmas in near-surface chambers can be strongly zoned in composition while still largely liquid (see Section 14.4).

Although rocks of gabbroic composition never form batholiths, inclusions of basaltic composition, ranging in size from centimeters to meters, are extremely common in all rocks found in batholiths regardless of their depth of exposure. These inclusions are blob-like in form and near the floor of batholiths may resemble stacks of basaltic pillows surrounded by granite (Wiebe, 1993). These pillows can even exhibit chilled margins against the surrounding granite (Fig. 4.79(A)). The shape and chilled margins of the blobs indicate that both the basalt and the granite were molten at the same time. In slowly cooled coarse-grained granite, such as the rapakivi granite from Finland (Fig. 4.79(E)), the basalt inclusions are still present but their boundaries are more diffuse. At higher levels in a batholith, the basaltic blobs are smaller and tend to be more dispersed (Fig. 4.79(B)). Even at the highest levels in a batholith, such as in the subvolcanic body on Virgin Gorda in the British Virgin Islands (Fig. 4.79(C)), fist-size blobs are still present. Where granite batholiths have fed ash-flows, blobs of basalt are commonly found amongst the pumice fragments (Fig. 4.79(D)). Basalt inclusions are so common in batholiths that it would appear they

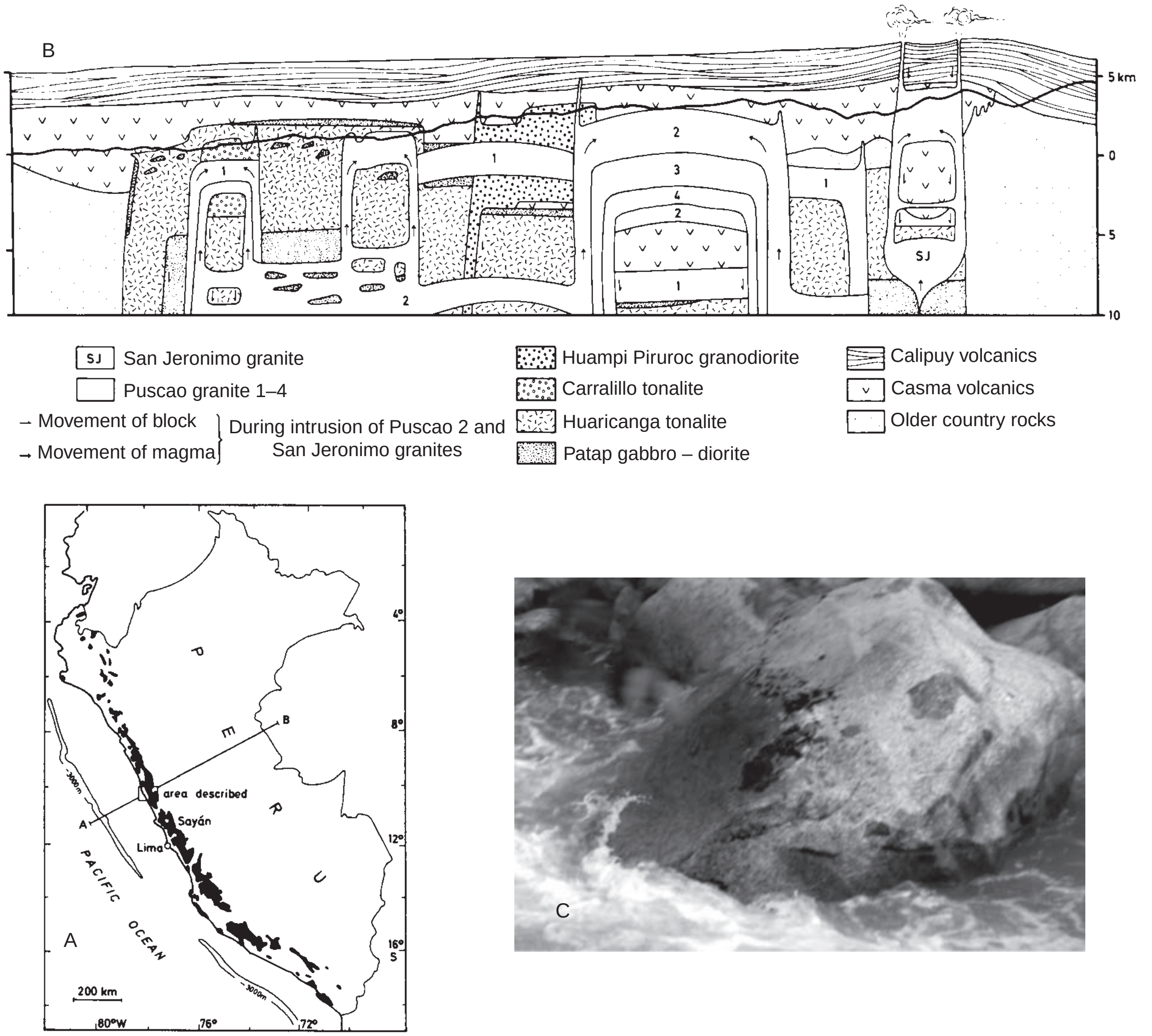


Fig. 4.78 (**A**) Coastal batholith of Peru and simplified cross section (**B**) showing how, at high levels in the crust, the body consists of numerous small plutons that have been emplaced by repeated caldron subsidence into their own volcanic ejecta. (See Problem 4.11.) (After Myers, 1975; reprinted by permission of the Geological Society of America.) (**C**) Blobs of basalt in the silicic rocks of this batholith are a common feature of most batholiths.

are an integral part of the batholith-forming process. The intrusion of basalt at the base of the crust (Fig. 4.42) is probably a necessary step in providing the thermal energy to form many batholiths and to keep them active (Annen *et al*., 2006).

Nowhere is a complete section through a large batholith exposed. Our concept of a batholith must therefore be pieced together with evidence drawn from various sources, including field studies, geophysical surveys, modeling experiments, and theoretical considerations. The following description of the genesis and form of a batholith is therefore purely interpretive but is probably close to the truth.

The first step in the development of a batholith is the formation of an enormous volume of relatively silicic magma. We are not in a position at this stage in the book to discuss this aspect in any detail, but a few general statements can be made. The volume of magma involved indicates that large quantities of heat must be developed in the crust. This may occur by introducing heat either from below in the form of mantle-derived basaltic magma or *in situ* from the decay of radioactive elements. Chemical and isotopic evidence indicates that many batholiths have formed largely through melting of crustal rocks, whereas others have involved a significant contribution from the mantle. It is possible that basaltic magma ascending from the mantle spreads laterally on reaching the top of the lower crust where it encounters rocks of lower density than itself and raises temperatures sufficiently to bring about partial melting of crustal rocks (Annen *et al*., 2006). Some mixing of the newly formed melt with the basaltic magma may occur, but density differences tend to keep the liquids stratified, with the denser basaltic magma remaining beneath. Some of the basaltic inclusions so common in batholithic rocks are probably incorporated at this stage (Fig. 4.79). The principal role of the basalt in this model is to transfer heat from the mantle into

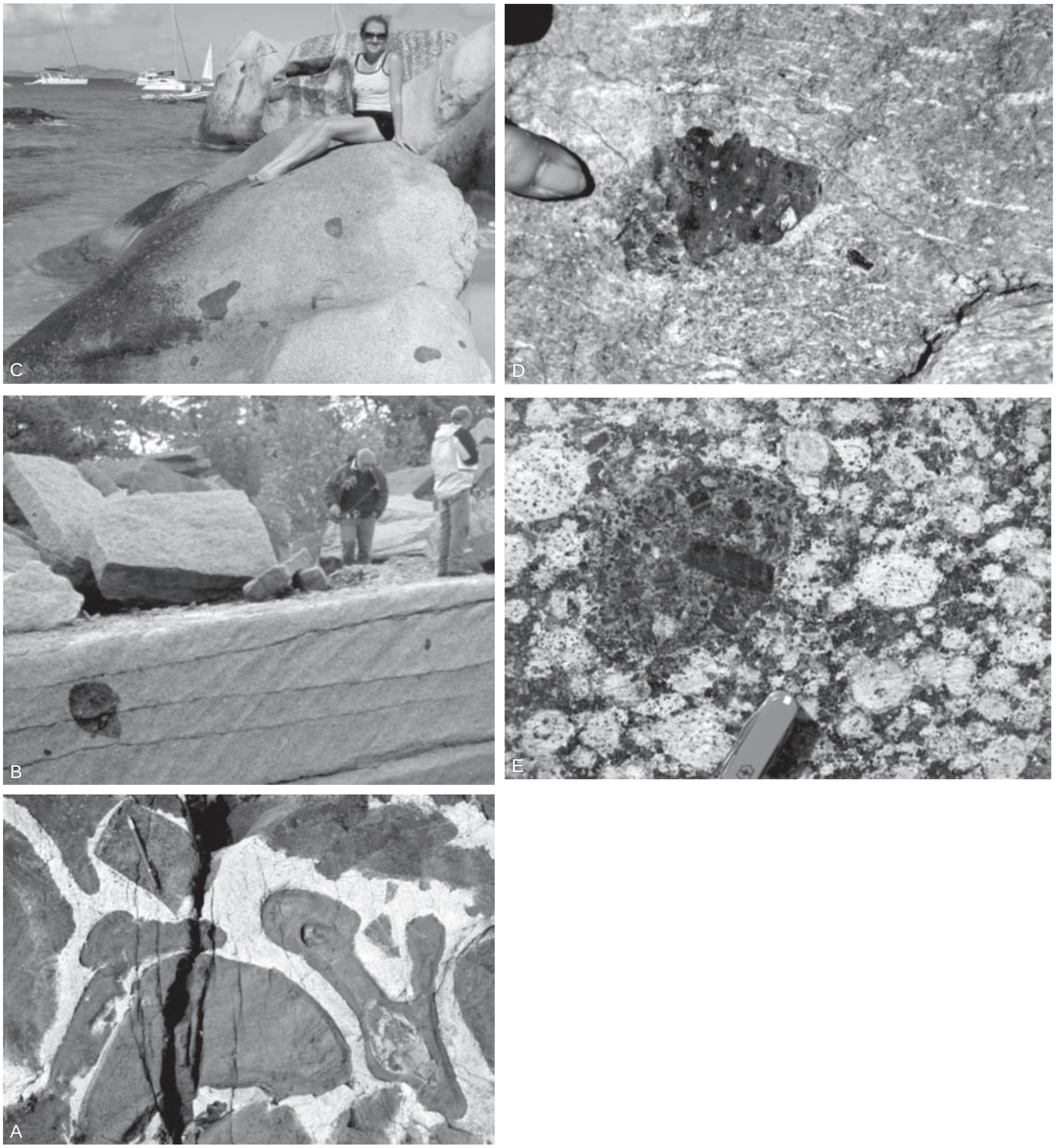


Fig. 4.79 (**A**) Basalt inclusions in granite on Mount Desert Island, Maine (Chapman, 1962). Pillow-like bodies of basalt formed when mafic magma was injected into still molten granitic magma. Rapid quenching of the basalt, viscosity contrasts between the magmas, and even liquid immiscibility may all have played a role in preventing these contrasting magmas from mixing. Note that the body of basalt on which the pencil rests is angular and must be of a different origin from the globular bodies. Note also that the globular bodies have a marginal zone formed from the quenching of the basalt by the lower-temperature granitic magma. The angular basalt fragment lacks this marginal zone. (**B**) Blob of more mafic rock in the Vinalhaven granite, Maine (Wiebe *et al*., 2004). The quarry face shows subhorizontal sheet joints, which are common in massive granite. (**C**) Blobs of basaltic material in the granite of Virgin Gorda, British Virgin Islands. This granite is a very shallow subvolcanic body. (**D**) Vesicular basalt blob and flattened pumice particles in welded ash-flow tuff; Chiricahua National Monument, Arizona. The common occurrence of such inclusions in welded ash flows has been used to indicate that rhyolitic eruptions may be triggered by intrusion of basaltic magma into silicic magma chambers (Eichelberger, 1975). (**E**) Rapakivi granite from southwest Finland; a coarse-grained pink granite containing large ovoid alkali feldspar crystals rimmed by sodic plagioclase. This granite is one of the world's most widely used facing stones for buildings. Rarely can a polished slab of this rock be found that does not contain a dark patch that was originally a basaltic inclusion.

the crust, but some chemical transfer also takes place. Heating of the lower crust to temperatures where partial fusion can occur will take place through radioactive heating if the radioactive crust is greatly thickened during orogenesis (see Problem 1.5). Geophysical surveys in southern Finland indicate that the crust beneath the mid-Proterozoic rapakivi granite batholiths is up to 20 km thinner than beneath the surrounding Fennoscandian shield (Korja *et al*. 2001).

Fig. 4.80 Landsat image of granite batholiths in the Pilbara district of western Australia (Myers, 1993). The lateral spreading of the batholiths has compressed the intruded rocks into tightly downfolded belts between the batholiths. Diabase dikes can be seen crossing the batholiths as narrow, dark, northeasterly trending lines. The dikes are much younger than the batholiths and formed at a time when the crust was brittle. Width of region is approximately 138 km. (Photo courtesy of NASA.)

Radioactive heating would not be a plausible cause for melting in this case. Instead, underplating of the crust with mantle-derived basaltic magma has been proposed as the cause for melting (Haapala and Rämö, 1999). The thinned crust may indicate that dense lower parts delaminated and sank into the mantle, causing hot asthenosphere to rise and replace it. Such a mechanism has been proposed for the formation of the granite batholiths that form the first large cratonic areas in the Archean (Zegers and van Keken, 2001), such as the Pilbara terranes in western Australia (Fig. 4.80).

Regardless of the ultimate source of heat causing fusion of the crust, the melt, once formed, is less dense than the overlying rocks and thus is gravitationally unstable and will tend to rise. How it does this is still open to question. There is no doubt that batholiths form large tabular bodies that can be 5 to 10 km thick and that they can intrude into their own associated volcanic pile. The question is, are batholiths the heads of diapiric plumes or are they fed laterally from a more constricted conduit (McCaffrey and Petford, 1997; Petford and Clemens, 2000)? Petford (2003) has argued that granite magmas rise largely in the liquid state and therefore their textures and structures preserve little or no record of the magma ascent; these features are developed as the magma finally crystallizes. Although the intrusion of a diapir, especially to high levels in the crust, is a thermally difficult process (Marsh, 1982), it would appear to be a mechanism that provides an explanation for the almost omnipresent blobs of basalt in batholiths. The basalt would be intruded near the base of the crust where it causes the melting to form silicic magma and to provide the thermal energy necessary to drive diapiric intrusion. Depending on the velocity of basaltic recharges into the base of this body, basalt may pond at the bottom (Wiebe, 1993; Wiebe *et al*., 2004) or become entrained and commingled in the overlying granitic magma (see Section 14.11 and Fig. 14.35).

If a batholith does form as a diapir, it would rise through the ductile lower crust as a spherical or tear-shaped body that would look very much like the early rising diapirs of oil in the model shown in Figure 3.15. Batholiths are not likely to be preserved in this stage of development, because they would ascend through the ductile zone before stopping. In the deep part of the crust the only record of the passage of an ascending batholith would be a zone of steeply dipping, strongly foliated, and lineated rocks left in the region where the batholith had passed through. Just as narrow tails of oil are left behind the ascending diapirs of oil in the model (Fig. 3.15), a vertically elongated body of granitic rocks might be left behind in this deep zone. Such bodies of granite are found, for example, in the deep crustal rocks exposed in the Grenville province of the Canadian Shield.

After passing through the ductile part of the crust, a diapir's ascent would be slowed by the cooler and more brittle upper crust, and the diapir would start to spread laterally in much the same way that the diapirs of oil do on reaching the top of the model tank (Fig. 3.15). It is at this stage of development that batholiths are commonly observed in deeply eroded sections through the crust (Fig. 4.80).

Because the rise of silicic magma results from a density instability (Rayleigh–Taylor), diapiric batholiths would be expected to exhibit a systematic spacing, just as do the domes of oil in the model shown in Figure 3.15 (see Eq. (3.31)). Many batholiths form regularly spaced domes, which may reflect such a spacing (e.g. Macgregor, 1951). Lateral spreading of the diapirs compresses and displaces downward the intervening country rocks, just as the spreading diapirs of oil displace most of the honey from between them (Fig. 3.15). Zones or screens of strongly flattened, steeply dipping country rocks are present between some closely spaced batholiths and are evidence of the lateral spreading of these bodies. Such zones are evident (Fig. 4.80) between granite batholiths in the Archean Pilbara Craton of Queensland, Australia (Myers, 1993). Lateral spreading may produce an almost continuous belt of magma, with only narrow screens of country rock and root zones marking the original diapirs.

The rise of batholiths above the ductile zone must involve rupturing of the overlying rocks. The distance a body rises through this zone will depend on its initial content of thermal energy and the state of stress in the upper crustal rocks. Heat loss by conduction increases as the temperature of the surrounding rocks decreases toward Earth's surface. Only large bodies of magma are likely to have sufficient heat to allow some fraction to remain liquid and rise to the surface. One

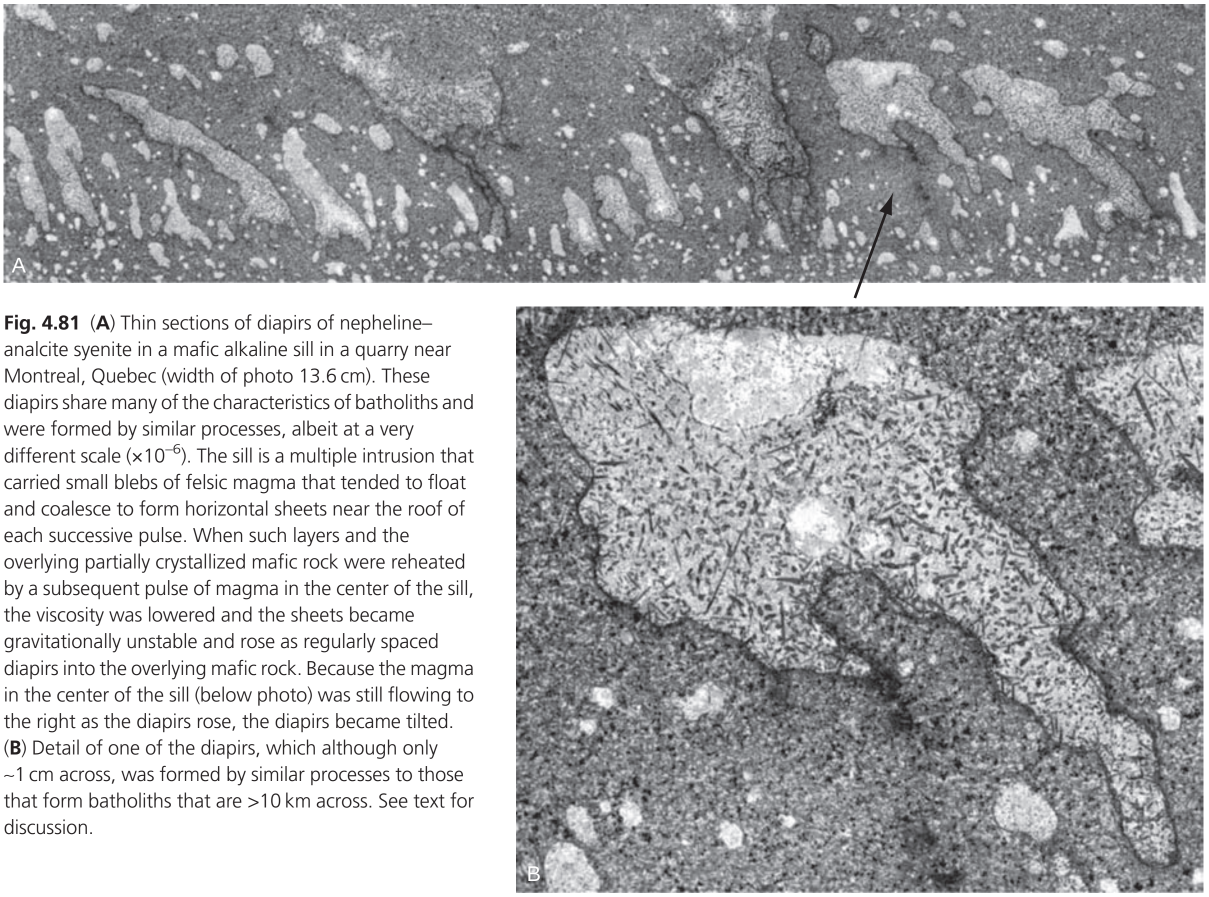


Fig. 4.81 (**A**) Thin sections of diapirs of nepheline–analcite syenite in a mafic alkaline sill in a quarry near Montreal, Quebec (width of photo 13.6 cm). These diapirs share many of the characteristics of batholiths and were formed by similar processes, albeit at a very different scale ($\times 10^{-6}$). The sill is a multiple intrusion that carried small blebs of felsic magma that tended to float and coalesce to form horizontal sheets near the roof of each successive pulse. When such layers and the overlying partially crystallized mafic rock were reheated by a subsequent pulse of magma in the center of the sill, the viscosity was lowered and the sheets became gravitationally unstable and rose as regularly spaced diapirs into the overlying mafic rock. Because the magma in the center of the sill (below photo) was still flowing to the right as the diapirs rose, the diapirs became tilted. (**B**) Detail of one of the diapirs, which although only ~1 cm across, was formed by similar processes to those that form batholiths that are >10 km across. See text for discussion.

exception would be where heat is replenished by the intrusion of new, hot basaltic magma into the base of a batholith (Wiebe, 1993). The common occurrence of basaltic inclusions in batholithic rocks and ash-flow tuffs suggests that this indeed may be a common means of reenergizing a magma chamber (Fig. 4.79). The final ascent towards the surface must be controlled by the ease with which country rocks respond to stoping and cauldron subsidence.

Much is known about the upper parts of batholiths because of the excellent exposures provided in deeply dissected mountain chains. For example, the cross section through the Coastal batholith of Peru (Fig. 4.78) is based on exposures covering more than 4.5 km of relief (Myers, 1975). This batholith exhibits features that are typical of bodies emplaced at this level in the crust. It consists of numerous plutons that have flat roofs and steep walls. The country rocks were not domed up or pushed aside by the intrusion of magma, which appears to have been emplaced by repeated caldron subsidence. The resulting stocks and ring dikes produce a complicated pattern of intrusive relations. The plutons even intrude associated volcanic rocks, some of which have dropped considerable distances during caldron subsidence (see Problem 4.11).

Batholiths probably do not conform to a single static shape. Their enormous size and content of buoyant material (even when solidified) may well cause their form to change continually as they are unroofed and exposed by erosion. Nonetheless, the form outlined here in this genetic description, which can be thought of as resembling a molar tooth with roots extending down into a ductile zone and a cap that is marked by various brittle intrusive relations, is one that many batholiths closely approach. To stress the importance of the relation between process and morphology, we will examine a body of igneous rock that was formed by many of the same processes outlined in the model, but instead of being tens of kilometers across is only 1 cm across (Fig. 4.81).

A flat-lying, meter-thick basaltic sill of Cretaceous age near Montreal, Quebec, carries globules (ocelli) of fine-grained felsic rock, which is interpreted to have formed as an immiscible liquid in the basalt (see Section 14.7). The origin of this felsic liquid need not concern us at this point; it is simply necessary to recognize that two liquids were present and that the felsic one was less dense and so rose and collected to form a sheet approximately 5 cm below the upper contact of the sill. It did not rise all the way to the top, because the cooled basaltic liquid near the contact was too viscous to allow it to pass. Before the rocks solidified, a second injection of fresh, hot basaltic liquid in the center of the sill raised the temperature of the earlier material sufficiently to lower its viscosity. The felsic layer, because of its lower density compared with the overlying basalt, developed a Rayleigh–Taylor instability and regularly spaced diapirs rose into the overlying basalt toward the upper contact, eventually coming to rest where they once again encountered highly viscous, possibly even brittle, material near the upper contact (Fig. 4.81(A)). Such regularly spaced diapirs are analogous to the regularly spaced volcanoes and associated intrusive bodies that rise from the zone of melting above a subduction zone (e.g. the coast of Peru; Fig. 4.78(A)).

One of the diapirs from the sill is shown in detail in Figure 4.81(B). Let us examine the genesis and form of this diapir, keeping in mind the possible analogy to a batholith. First, a pulse of fresh basaltic magma supplies heat that initiates the rise of the diapir. The heat may cause fusion of minerals that have already solidified; it most certainly will lower the viscosities of both the felsic and, more importantly, basaltic liquids. Basalt intruded into the base of Earth's crust will similarly not only cause fusion of crustal rocks but will reduce the viscosity of the ductile zone. The diapir in the sill rose until it reached cooler and significantly more viscous basalt near the upper contact of the sill, whereupon it accumulated and flattened to form a body having the shape of a molar tooth, with a root or tail marking the ascent route of the diapir through the underlying basalt. Because the magma in the sill continued to move slightly as the diapir rose, the structure was tilted. Syntectonic batholiths are similarly tilted by tectonic movements at convergent plate boundaries. The upper surface of the diapir in the sill is quite flat where it abuts against the overlying more rigid basaltic rock. The felsic material stayed molten to temperatures ~200 °C lower than those at which the basaltic rock solidified. Although not shown in Figure 4.81, elsewhere in this sill, fractures allowed the felsic liquid to escape from the diapirs and actually vein the overlying brittle rock, thus mimicking the brittle behavior in the upper zones of batholiths. The diapir in the sill is slightly richer in hornblende crystals (dark) toward its base and richer in feldspar and nepheline toward the top. This is analogous to the upward gradation from quartz diorite to granite found in many batholiths. One patch at the top of the diapir contains almost no ferromagnesian minerals and is composed mainly of alkali feldspar, nepheline, and analcite, which is a hydrous mineral ($NaAlSi_2O_6 \cdot H_2O$). The upper parts of batholiths are similarly enriched in volatile constituents, and accumulations of such material may be responsible for the formation of resurgent domes.

Large batholiths and the small diapir in Figure 4.81(B) have formed as a result of buoyant magma rising slowly through a denser but increasingly more viscous medium following an input of thermal energy. They both effect an upward transfer of low-melting fractions, and in the case of batholiths this process has undoubtedly played a major role in the development of Earth's crust throughout geologic time.

4.14 PROBLEMS

4.1 In historic times, the 1783 fissure eruption of Lakagigar, Iceland, most closely resembles a flood basalt eruption, but it is small by comparison with flood basalts in the geologic record. Between June 9 and July 29 of that year, 9.5 km^3 of magma were erupted from a 10-km-long fissure estimated to have been about 4 m wide at depth. What was the average flow velocity in the feeder dike, and was the flow turbulent or laminar? Use typical values for the physical properties of a basaltic magma.

4.2 The average width of basaltic dikes in Iceland is 4 meters.

(a) If the magma that rises in these dikes has a density of 2.7 Mg m^{-3} and a viscosity of 300 Pa s, and they intrude rocks with a density of 2.97 Mg m^{-3}, what would the average flow velocity need to be to maintain flow?

(b) Would this flow be laminar or turbulent?

(c) Would the density difference be sufficient to provide a buoyant force great enough to produce the required average flow velocity?

4.3 In light of your answers to Problem 4.1 and 4.2, what might you conclude about the 1783 Lakagigar eruption, which was clearly able to maintain flow for almost two months?

4.4 If essentially all the CO_2 in an ascending andesitic magma has exsolved by the time the pressure has dropped to 50 MPa, how much will this gas expand on rising to the surface (pressure at surface = 0.101 MPa) if its temperature is held constant by the thermal mass of the surrounding lava and it behaves as an ideal gas? If the gas at 50 MPa constituted 5 vol% of the ascending magma, what effect would the gas expansion have on the magma on rising to the surface?

4.5 The goal of this problem is to determine how much compaction occurred in the Chiricahua welded ash-flow tuff of southern Arizona illustrated in Figure 4.40 (D), assuming that the pumice particles were initially spherical. It is necessary to determine an average flattening of the particles. This can be done most easily by fitting ellipses to the particles and then obtaining an average ellipse for the sample using the National Institutes of Health (NIH) *ImageJ* program, which can be downloaded free from the Web at http://rsb.info.nih.gov/ij.

First download Figure 4.40(D) from the text's web page (www.cambridge.org/philpotts), and open it in NIH ImageJ. Click on the **straight line tool** in the toolbar, and draw a diameter line across the lens cap. In the pull-down menu bar (pdm) click on **Analyze** and select **Set Scale…** Enter 7.5 for the known diameter of the lens cap in centimeters. Now all measurements will be true to scale. In the pdm, click on **Image** and select **Type ‣ 8-bit** to convert the photograph to an 8-bit grayscale image. Again, from the pdm, click on **Image** and select **Adjust ‣ Brightness/Contrast**. Adjust the contrast and the brightness until the low end of the diagonal line is at ~60 and the upper end at ~250 and click **Apply**. This makes the pumice fragments stand out from the matrix. From **Image** in the pdm, select **Lookup Tables ‣ Invert LUT**. The pumice fragments will now be dark in a light matrix. Again in the pdm, go to **Image** and select **Adjust ‣ Threshold**. Adjust the top bar so that the pumice fragments are mainly red and the matrix mainly gray and click **Apply**. You will now have a strongly contrasted image. The pumice fragments will still contain a few specks of white and the matrix a few specks of black. These can be cleaned up by choosing **Process** in the pdm and selecting **Noise ▸ Despeckle**. Be patient; this takes a while. This process can be repeated (try ~5 times) to get rid of more dots.

You will now have a relatively clean image in which the pumice fragments are mainly black and the matrix white. We now get the program to outline the fragments by going to **Process** in the pdm and selecting **Find Edges**. This will draw a line around each of the pumice fragments. There will be some inclusions in the fragments, which can be removed by going to **Process** in the pdm and selecting **Binary ▸ Fill Holes**. You will now have black pumice fragments. If, on inspection, there is some fragment that clearly was a single fragment but has been separated into two by the image processing, you could join them with the paint tool and repeat the **Find Edges** step.

The next step is to fit an ellipse to each pumice fragment. From **Analyze** in the pdm, select **Set Measurements…** Click the boxes for *Area*, *Center of Mass*, and *Fit Ellipse*, and click **OK**. From the Toolbar select the **wand tool** and click on a pumice fragment; it will become outlined. Then, in the pdm choose **Edit** and select **Selection ▸ Fit Ellipse**. This will draw an ellipse that best fits the outline of the pumice fragment you selected. It is not necessary to draw an ellipse on each fragment; the program will fit such an ellipse and record the information in a data table. With the wand tool, select another fragment, and in the pdm chose **Analyze** and **Measure**. Results table will appear giving the area of the ellipse, the x and y coordinates of its center of mass, its major and minor axes, and the counterclockwise angle the major axis makes from horizontal to the right. With the wand tool, click on another fragment and again select **Measure** from the **Analyze** menu. Each new measurement will appear in the results table. Continue the process until you have measured a sufficient number of fragments to obtain a significant measure of their ellipticity. Do not measure fragments that touch the side of the image, because their full extent is unknown. Scroll over your results and copy them into a Microsoft Excel spreadsheet.

Once the data are in the spreadsheet, they can be sorted according to the area of the ellipses. Calculate the percent compaction, which is given as

$$\%\ \text{Compaction} = 100 - 100\left(\frac{\text{Minor axis}}{\text{Major axis}}\right)$$

The large ellipses indicate compaction of ~80% whereas the smaller ones show less compaction.

Obtaining an average ellipse for the sample is not a simple task. Should the presence of a few large fragments be allowed to bias the mean, or should all ellipses be weighted the same regardless of size? If the welded ash-flow tuff were a completely homogeneous material and the pumice fragments were passive markers, all fragments should indicate the same degree of compaction, in which case all fragments should be weighted equally in the mean. This is done by first normalizing all of the ellipses to a unit area.

An Excel spreadsheet that provides two different ways of obtaining a mean ellipse can be downloaded from the text's web site (www.cambridge.org/philpotts). Using this program, the 30 largest pumice fragments in this image indicates 69% compaction, with the major axis tilted at 177°, when the area of each ellipse is normalized to unity.

4.6 The Loch Ba granophyric (rhyolitic) ring dike on the Island of Mull, Scotland, has an outside diameter of approximately 5.6 km and an inside diameter of 5.5 km.

(a) If intrusion of the ring dike resulted solely from the subsidence of the central block, what would the ratio of the average rate of intrusion to the rate of subsidence have been?

(b) The radius of curvature of the ring dike is sufficiently great that at any point the dike can be treated as a planar sheet. The granophyric magma had a density of 2.4 Mg m^{-3} and a viscosity of 10^7 Pa s. The subsided block consists largely of basaltic rocks with a density of 3.0 Mg m^{-3}. Assuming laminar flow, calculate the average intrusion velocity of the magma relative to the surrounding rocks and the rate of subsidence of the central block. Check to see if the assumption of laminar flow is valid.

(c) If growth of gas bubbles in the ring-dike magma disrupted the magma, and the magma-gas mixture's viscosity was 0.1 Pa s, and its density was 1.8 Mg m^{-3}, what would have been the average rate of intrusion of the magma and the rate of subsidence of the block?

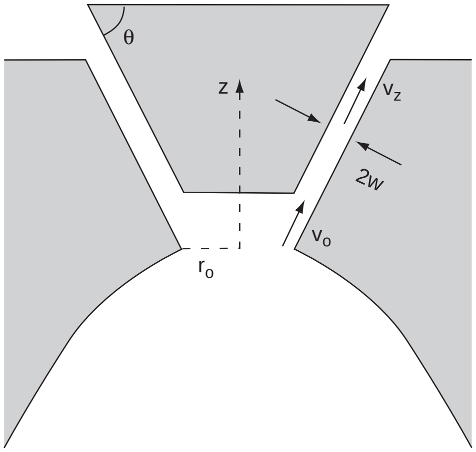


Fig. 4.82 For Problem 4.7.

(d) In light of your answers to parts (b) and (c), what role is magma disruption likely to play in the emplacement of rhyolitic ring dikes?

4.7 (a) For a cone sheet having a width $2w$, a dip of $\theta°$, and an initial radius at the magma reservoir of r_0 (Fig. 4.82), derive an expression for the average magma velocity, $\overline{v}$, at any height z above the source in terms of the initial average velocity, $\overline{v}_0$.

(b) From the equation derived, what might you conclude about the possible dip on cone sheets and the initial radius r_0?

4.8 (a) In Problem 4.2 we saw that an average flow velocity of 10.21 m s^{-1} was necessary to maintain flow in a 4-m-wide basaltic dike in which the magma, with a density of 2.7 Mg m^{-3} and viscosity of 300 Pa s, intrudes basaltic rocks with a density of 2.97 Mg m^{-3}. If a 4-m-wide cone sheet forms with a dip of 30° and an initial radius (r_0) at the top of the magma chamber of 100 m, calculate how high (z) the magma will rise in the cone sheet before its velocity becomes too low to maintain flow. The pressure gradient imposed on the magma at the point of entry into the cone sheet is 3000 Pa m^{-1}. Make use of the relation derived in Problem 4.7.

(b) Repeat the problem for cone sheets that dip at 60° and 20°.

(c) What would be the consequences of the raised block being tilted by the injection of the cone sheet so that the cone sheet would be thicker on one side and narrower on the other? What effect would this have on the distance the magma would rise? What might the resulting outcrop pattern of the cone sheet be on the surface of the Earth?

4.9 By studying the map of the Pilanesberg Complex, South Africa (Fig. 4.56), deduce the sequence of intrusive events that led to the formation of this body.

4.10 (a) Using Eq. (4.5), draw cross sections through two laccoliths (on a single graph), both having a radius of 2 km, one of which is intruded at a depth of 200 m and the other at 500 m into sandstone having a density of 2.7 Mg m^{-3}, Young's modulus of 3×10^{10} Pa, and a Poisson's ratio of 0.1. The pressure on the magma is due solely to buoyancy resulting from a 10-km-long feeder column of magma with a density of 2.4 Mg m^{-3} intruding rocks with an average density of 2.7 Mg m^{-3}.

(b) Repeat the calculation for a laccolith intruded at a depth of 200 m but having a radius of only 1 km.

(c) Repeat part (a) for the case where intrusion is into limestone that has a Young's modulus of 6×10^{10} Pa, a Poisson's ratio of 0.2, and a density of 2.7 Mg m^{-3}.

(d) In light of the results of your calculations, discuss the relative importance of the factors controlling the shape of a laccolith.

4.11 Work out the chronological sequence of igneous rocks shown in the cross section through the Coastal batholith of Peru in Figure 4.78.

4.12 Draw a sketch of Figure 4.44 and work out the chronology of the intrusive events seen in this outcrop. The dark igneous rocks are of various gabbroic compositions, and the light-colored ones are of syenite.

5 Cooling of igneous bodies and other diffusion processes

5.1 INTRODUCTION

Once magma has been intruded or extruded, it begins to cool and crystallize. To cool, the magma must lose heat to its surroundings, which it does mainly by conduction. Conduction can be thought of as the diffusion of heat. The loss of heat from the magma causes solidification and crystallization of the magma and the growth of metamorphic minerals in rocks that are heated by the magma. The growth of crystals in the magma and in the metamorphic rocks first involves the formation of nuclei, and then the components needed for crystal growth must diffuse to these nuclei. This occurs through the liquid in the case of magma and through solid rock or films of grain-boundary fluid in the case of metamorphic rocks. As magma crystallizes, the composition of the residual liquid is continuously changing and the composition of crystals must change their composition if they are to remain in equilibrium with the melt. For early formed phenocrysts to adjust their composition, components need to diffuse in and out of the crystals. As the temperature of metamorphic rocks changes, the minerals must adjust their compositions to remain in equilibrium and this again involves diffusion. Diffusion, consequently, is one of the most important processes in the formation of rocks.

To appreciate diffusion processes, it is necessary to understand some simple mathematical relations. We will develop these relations by discussing the cooling of igneous bodies. The same mathematical relations will then be applied to the diffusion of chemical components.

The rate of heat loss from magmatic bodies plays an important role in many important petrologic processes, such as nucleation and growth of crystals, settling of crystals, magmatic convection, rates of intrusion, and chemical reactions between magmas and phenocrysts or country rocks. The transfer of heat from an igneous body into its surroundings also causes contact metamorphism. On a still larger scale, cooling of newly formed lithosphere as it moves away from oceanic ridges controls the depth to which plates sink into the asthenosphere (and consequently, ocean depths). Eventually, when plates are subducted, the rate at which heat flows back into the descending slab determines the nature of the resulting metamorphic rocks (see Section 1.2).

Heat transfer is a large subject involving considerable mathematics. Only a brief introduction can be given here. Many of the mathematical derivations are complicated, but their results can be applied quite simply to interesting geological problems.

Heat, which is thermal energy, can be thought of as flowing from high- to low-temperature regions. This transfer can take place in three different ways: (1) by *conduction*, in which the thermal energy is transmitted through a substance by the transference of kinetic energy from one atom or molecule to another; (2) by *convection*, in which heat is transferred by the actual bodily movement of hotter material to cooler parts of the system; (3) by *radiation*, where the energy is transferred directly from one place to another by electromagnetic radiation. In many geological situations, conduction is the most important mechanism of heat transfer, but in large bodies of magma or large sections of the mantle, convection can play an important role. Heat loss through radiation is the least important mechanism in the cooling of plutonic masses because of the relative opacity of most rocks and magmas to electromagnetic radiation at the temperatures encountered in the Earth's crust. At higher temperatures, however, such as those deep in the mantle, radiation may play a significant role in the transfer of heat. Extremely felsic magmas, which on rapid cooling form relatively clear glasses, might radiate considerable heat to the margins of an intrusion, giving rise to more rapid cooling than would otherwise be expected from conduction alone. Radiation, of course, plays an important role in the cooling of the surface of lava flows or volcanic ash, and if the lava is in turbulent motion, it can be the most important mechanism for cooling the entire flow (Shaw and Swanson, 1970). In this chapter, we discuss cooling of igneous bodies by conduction and radiation, leaving the topic of convection to Chapter 14.

The literature dealing with heat conduction is extensive. One of the most readable treatises is by Ingersoll, Zobel, and Ingersoll (1954), who develop much of the general theory and give many applications, including geological ones. The most comprehensive treatment of the subject, however, is the standard text *Conduction of Heat in Solids* by Carslaw and Jaeger (1959). But for many geologists the chapter by Jaeger (1968) on the 'Cooling and solidification of igneous rocks', in Volume 2 of *Basalts*, edited by Hess and Poldevaart, is more digestible. Finally, an excellent treatment of many heat transfer problems in geology is given by Turcotte and Schubert (1982).

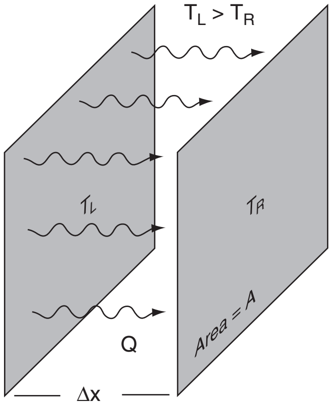


Fig. 5.1 Heat transfer between two parallel plates.

5.2 GENERAL THEORY OF HEAT CONDUCTION

If two parallel plates that differ in temperature by ΔT are separated by a distance Δx (Fig. 5.1), the amount of heat Q transferred from one to the other per area A of the plates is directly proportional to the temperature difference between the plates and the length of time t during which heat flows, and inversely to the distance between the plates. This can be expressed as

$$Q = -KA\frac{\Delta T\, t}{\Delta x} \tag{5.1}$$

The constant of proportionality K is known as the *thermal conductivity*. Although it is not actually a constant (decreases slightly with temperature for most substances), it is treated as such in most heat flow problems. However, if large temperature differences are considered, especially at the scale of the whole Earth or tectonic plates, K should be expressed as a function of temperature (Hofmeister and Criss, 2005). If the plates in Figure 5.1 are moved closer together, the limiting value of $\Delta T/\Delta x$ is $\partial T/\partial x$, which is the *thermal gradient* in the x direction at that point.

If the thermal gradient is taken in the direction of heat flow, it will be negative. The rate of heat flow, which is a positive number, is given by differentiating Eq. (5.1),

$$\frac{\partial Q}{\partial t} = -KA\frac{\partial T}{\partial x} \tag{5.2}$$

This rate of heat flow across a unit area is known as the *heat flux* and is given by

$$J_x = \frac{\partial Q}{\partial t \cdot A} = -K\frac{\partial T}{\partial x} \tag{5.3}$$

This is known as *Fourier's law*, which is similar in its formulation to the expression for fluid flow derived in Chapter 3 (Eq. (3.10)). Values of heat flux from the Earth have already been given in Chapter 1. Typical values range from 60 mW m^{-2} (1 W = 1 J s^{-1}) over stable continents to 335 mW m^{-2} over oceanic ridges (Problem 5.1).

Although thermal conductivity tells us the rate at which heat can be transferred into a substance, it does not indicate the rate at which the temperature of the material will change as a result of this influx of heat. Yet for many petrologic problems the temperature, not the quantity of heat transferred, is the critical factor. For example, temperature determines when a magma will crystallize or a metamorphic reaction will take place, not the amount of heat that flows. The temperature change can be determined only by knowing the heat capacity of the substance into which the heat passes.

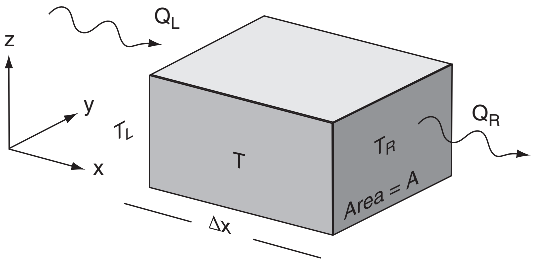


Fig. 5.2 Transfer of heat through a small block.

Heat capacity, C_p, is the amount of heat necessary to raise the temperature of a unit mass of a substance 1 degree at constant pressure. If the density of the material is ρ, the amount of heat necessary to raise the temperature of a unit volume of the substance by 1 degree is $C_p{\cdot}\rho$. Therefore, this quantity of heat must flow into the unit volume for its temperature to be raised 1 degree. The rate of heat flow is given by the thermal conductivity K, which is the quantity of heat that flows in unit time through unit area of a plate of unit thickness having unit temperature difference between its faces. Thus, dividing the value of K by $C_p{\cdot}\rho$ gives the change of temperature produced in the unit volume by the quantity of heat that flows. This is known as the *thermal diffusivity* of a substance and is defined as

$$k \equiv \frac{K}{C_p \cdot \rho} \tag{5.4}$$

This is the property of a material that must be known if we are to determine the change in temperature resulting from an influx of heat. Fortunately for petrologists, most rocks and magmas have almost identical values of k, near 10^{-6} m^2 s^{-1} (Problem 5.2).

Consider a small volume of rock of cross-sectional area A and length Δx through which heat flows only in the x direction (Fig. 5.2). The temperature at the center of this volume is T, and there is a thermal gradient in the x direction of $\partial T/\partial x$. The temperature on the face at the right end of the volume is

$$T_R = T + \tfrac{1}{2}\Delta x\frac{\partial T}{\partial x} \tag{5.5}$$

and at the left end it will be

$$T_L = T - \tfrac{1}{2}\Delta x\frac{\partial T}{\partial x}$$

The rate of heat flow across the right end is, from Eq. (5.2),

$$\frac{\partial Q_{\mathrm{R}}}{\partial t} = -KA\frac{\partial T_{\mathrm{R}}}{\partial x} = -KA\frac{\partial}{\partial x}\left(T + \tfrac{1}{2}\Delta x\left(\frac{\partial T}{\partial x}\right)\right)$$
$$= -\tfrac{1}{2}KA\Delta x\left(\frac{\partial^2 T}{\partial x^2}\right) \tag{5.6}$$

and across the left end is

$$\frac{\partial Q_{\mathrm{L}}}{\partial t} = -KA\frac{\partial T_{\mathrm{L}}}{\partial x} = -KA\frac{\partial}{\partial x}\left(T - \tfrac{1}{2}\Delta x\left(\frac{\partial T}{\partial x}\right)\right)$$
$$= \tfrac{1}{2}KA\Delta x\left(\frac{\partial^2 T}{\partial x^2}\right) \tag{5.7}$$

The difference between the rates of heat flow into the left end and out of the right end of the small volume is the amount of heat gained per unit time by the material in the volume:

$$\frac{\partial Q}{\partial T} = \tfrac{1}{2}KA\Delta x\left(\frac{\partial^2 T}{\partial x^2}\right) - \left(-\tfrac{1}{2}KA\Delta x\left(\frac{\partial^2 T}{\partial x^2}\right)\right)$$
$$= KA\Delta x\left(\frac{\partial^2 T}{\partial x^2}\right) \tag{5.8}$$

The rate of heating of the volume can, on the other hand, be expressed in terms of the heat capacity. Multiplication of the mass of the volume ($\rho \cdot A \cdot \Delta x$) by the heat capacity ($C_{\mathrm{p}}$) gives the amount of heat necessary to raise the temperature of the volume by 1 degree. Multiplication of this amount ($C_{\mathrm{p}} \cdot \alpha \cdot A \cdot \Delta x$) by the rate of temperature rise ($\partial T/\partial t$) gives the rate of heating of the volume, which can then be equated to the expression for this quantity in Eq. (5.8). Therefore,

$$C_{\mathrm{p}} \cdot \rho \cdot A \cdot \Delta x\left(\frac{\partial T}{\partial t}\right) = K \cdot A \cdot \Delta x\left(\frac{\partial^2 T}{\partial x^2}\right) \tag{5.9}$$

and

$$\frac{\partial T}{\partial t} = \frac{K}{C_{\mathrm{p}} \cdot \rho}\left(\frac{\partial^2 T}{\partial x^2}\right) = k\left(\frac{\partial^2 T}{\partial x^2}\right) \tag{5.10}$$

In the general case where heat, instead of flowing in the x direction only, has y and z components as well, the same general relation holds, but the derivatives of the thermal gradients in each of the mutually perpendicular x, y, and z directions must be summed. For the general case, then,

$$\frac{\partial T}{\partial t} = k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) \tag{5.11}$$

This is known as *Fourier's equation*, which states that the change in temperature of a substance with time is given by the product of the thermal diffusivity and the derivative of the thermal gradient with respect to distance. The solution to any problem involving heat flow must satisfy this equation.

The significance of Eq. (5.11) is illustrated in Figure 5.3. The temperature distribution near an igneous contact is shown at some time following intrusion. Below this is the first derivative of this temperature distribution; note that a maximum occurs at the contact. The second derivative, at the bottom of the figure, has a maximum at A and a minimum at B. According to Fourier's equation, the rate of temperature change at any point is determined by the second derivative. Hence, at the particular time represented in Figure 5.3, the country rock at point A is heating the most rapidly, and the magma at point B is cooling the most rapidly. As the temperature gradient changes with time, so do the positions of these maxima. Because the second derivative is zero at the contact, the temperature there remains constant and, as will be shown later, remains constant as long as magma and country rock at some distance from the contact still have their initial temperatures.

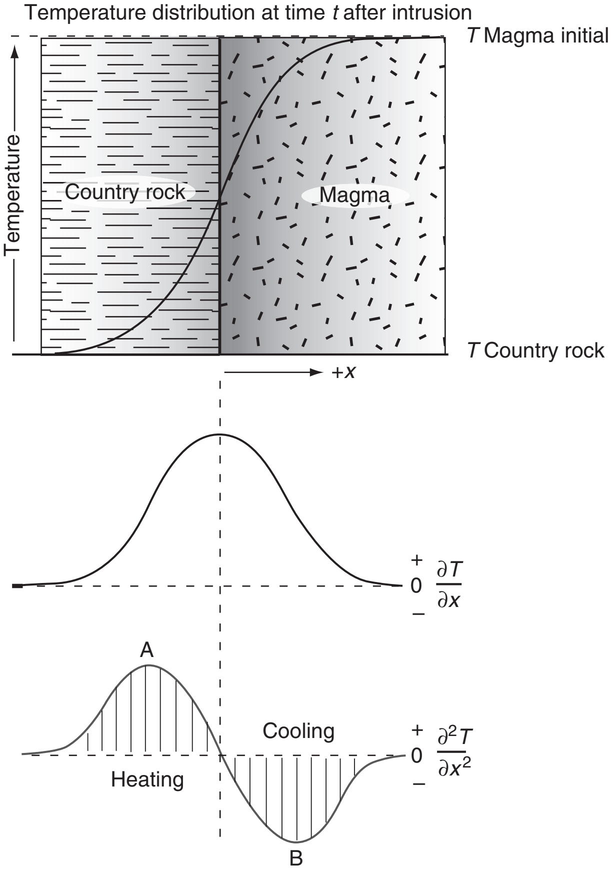


Fig. 5.3 The temperature distribution in the vicinity of an igneous contact some time after intrusion is shown at the top of the diagram. Beneath this are the first and second derivatives of temperature with respect to distance. According to Fourier's equation (Eq. 5.11) the maximum rate of cooling occurs where the second derivative is most negative (B), and the maximum rate of heating occurs where it is most positive (A).

5.3 HEAT CONDUCTION ACROSS A PLANE CONTACT

We will now consider a simple problem, that of magma cooling in the vicinity of a plane contact, such as would occur near the margin of a dike or sill, or perhaps at the margin of a batholith, as long as the contact was not significantly curved. We must first set up the initial and boundary conditions. For

Table 5.1 *The error function:* $\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y e^{-u^2}\,du$

y	erf(y)	y	erf(y)	y	erf(y)
0.00	0.000	0.80	0.742	1.60	0.976
0.10	0.112	0.90	0.797	1.70	0.984
0.20	0.223	1.00	0.843	1.80	0.989
0.30	0.329	1.10	0.880	1.90	0.993
0.40	0.428	1.20	0.910	2.00	0.995
0.50	0.521	1.30	0.934	2.20	0.998
0.60	0.604	1.40	0.952	2.40	0.9993
0.70	0.678	1.50	0.966	2.50	0.9996

The error function for positive values of y can be approximated by

$$\mathrm{erf}(y) = 1 - (a_1 t + a_2 t^2 + a_3 t^3)\exp(-(y^2))$$

where $t = 1/(1 + 0.47047y)$, $a_1 = 0.34802$, $a_2 = -0.09587$, $a_3 = 0.74785$. For negative values of y, use erf($-y$)=$-$erf(y).

The error function is also available in Microsoft Excel. If it is not installed, run the Setup program and install the Analysis ToolPak. Enable the function by using the **Add-Ins** command on the **Tools** menu. The error function is obtained by writing =erf(x).

simplicity in the calculations, the temperature of the country rock is taken to be initially 0 °C, and after intrusion this same temperature occurs at greater and greater distances from the contact. The fact that rocks may initially be at some temperature other than zero is taken into account simply by adjusting the temperature scale up or down as the case may be. The magma is emplaced with a temperature T_0 (adjusted, as previously discussed) and begins to lose heat through the contact to the country rock. The igneous body is, however, of sufficient size that magma with the initial intrusion temperature still remains at some distance from the contact. This distance, of course, increases with time, and eventually no magma with the initial temperature will remain. At this time the boundary conditions are changed, and this poses a different problem, which will be considered later. Mathematically, these boundary conditions can be expressed as follows. If the distance from the contact is denoted by x, negative values being in the country rock and positive ones in the igneous body, then for $x > 0$, T/T_0 tends to the value 1, and for $x < 0$ it tends to 0.

To find the temperature T, at distance x, at time t after intrusion, a solution to the Fourier conduction equation must be found that satisfies the conditions of the problem. This involves the use of Fourier's series and integrals. For the reader with knowledge of this area of mathematics, the solutions presented in the previously cited references are worth consulting. The solution to this problem and to many others concerning heat conduction, and diffusion in general, involves a rather frightening-looking integral known as an *error function* (erf):

$$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}}\int_0^y e^{-u^2}\,du \qquad (5.12)$$

Fortunately, values of the error function are given in most mathematical tables, and for convenience a few are given in Table 5.1 [note that erf(y)= erf(y)].

A derivation of the solution to the problem posed here is given by Carslaw and Jaeger (1959, p. 58). The solution, which is based on the thermal diffusivity k of magma and country rock being the same (a valid assumption), is given by

$$\frac{T}{T_0} = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{x}{2\sqrt{kt}}\right) \qquad (5.13)$$

The reader should verify that this solution does indeed satisfy the initial and boundary conditions of the problem. For example, when $x \gg 0$, the error function tends to the value 1, and hence T/T_0 tends to 1. What does the value of T/T_0 tend to when $x \ll 0$?

This solution reveals an important fact about the maximum possible temperature in the country rock near an intrusion. At the contact, $x = 0$, and therefore $T = 0.5\,T_0$. That is, the temperature at the contact is half the temperature of the magma (suitably adjusted so that the initial country rock temperature is zero) immediately after intrusion, and it remains at this value as long as the boundary conditions are the same. The country rock cannot be heated to a higher temperature no matter how large the body of magma unless there is an additional supply of heat, as could happen with a fresh surge of hot magma or with the emplacement of another intrusion nearby.

Equation (5.13) also allows us to follow the progress of an isotherm, which may correspond to a particular event, such as the temperature of complete solidification of the magma or the temperature at which a particular metamorphic mineral assemblage appears in the country rock. If this temperature is T_1, then T_1/T_0 is a constant and the error function remains constant. Thus, $x = \text{constant} \times \sqrt{kt}$. That is, the distance of the isotherm from the contact is proportional to $\sqrt{t}$. This explains the linear relation discussed in Chapter 2 between the thickness of the crust on the Alae and Kilauea Iki lava lakes and the square root of the cooling time (Fig. 2.3). It has also been used to explain the linear fit of the depth of the ocean as a function of the square root of the age of the ocean floor as it moves away from a mid-ocean ridge (Fig. 1.6).

To illustrate Eq. (5.13), we will determine the time necessary for magma 1 m from the contact of a large dike to cool to 900 °C following intrusion at 1200 °C into country rock at 100 °C. The thermal diffusivity of the magma and country rock is 10^{-6} m^2 s^{-1}. Adjusting the temperature scale so that the country rock is at zero gives $T_0 = 1100$ °C and $T = 800$ °C, and therefore $T/T_0 = 0.727$. This gives a value for the error function of 0.454, which, from Table 5.1, indicates that $x/\left(2\sqrt{kt}\right) = 0.425$. When $x = 1$ m, the cooling time, t, is 16 days. For comparison, magma 2 m from the contact would take 62 days to cool to the same temperature, and at 4 m it would require 258 days.

In these calculations, the latent heat of crystallization of magma or the heat of a metamorphic reaction is not taken into account. When magma crystallizes, heat is liberated, which causes magma to cool more slowly than predicted by Eq. (5.13) (Problem 5.5). This heat is not simple to incorporate

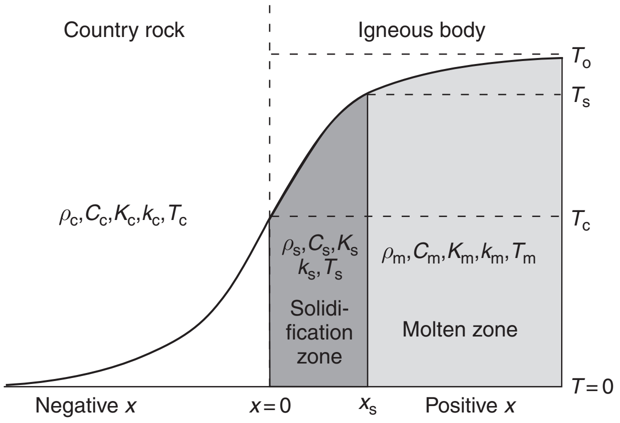


Fig. 5.4 Temperature distribution near an igneous contact where magma solidifies at a specific temperature T_s, resulting in a solidification zone that progresses into the magma body with time. See text for definition of terms and discussion.

into the calculations. By adjusting the initial temperature of the magma upward, however, an amount of heat equivalent to that liberated by crystallization can be introduced. Thus, substituting (T_0+L/C_p) for T_0, where L is the latent heat of crystallization, a closer approach to the rate of cooling is obtained, at least in the country rock. This approximation, unfortunately, gives temperatures in the igneous body that are too high.

There is one relatively simple situation in which heat of crystallization can be taken into account. This involves magma that, on cooling near a plane contact, solidifies at a single definite temperature. In this case, a solidification surface migrates inward as cooling proceeds (Fig. 5.4). Most magmas, in fact, solidify over a temperature range (Section 2.2), but this is commonly sufficiently small (Chapter 10) that the approximation of magma solidifying at one temperature does not introduce large errors.

The conditions are essentially the same as in the previous case; that is, a plane contact between magma and country rock, with distance x from the contact being positive in the magma and negative in the country rock. The initial temperatures of magma and country rock are T_0 and zero respectively, and these continue to exist but at progressively greater distances from the contact as cooling proceeds. To these conditions must, now, be added a zone of solidified magma that extends from the contact ($x=0$) a distance x_s, where x_s is a function only of time. Solidification of the magma occurs once the magma temperature has decreased to T_s. The physical properties – density, heat capacity, thermal conductivity and diffusivity, and temperature – of the materials constituting the three different zones – magma, solid igneous rock, and country rock – are designated with subscripts *m*, *s*, and *c*, respectively, as indicated in Figure 5.4.

The temperature gradient through the three zones following some period of cooling is illustrated in Figure 5.4. Magma with the initial temperature, T_0, must first cool to T_s before solidification occurs. This signifies that the magma is initially superheated by an amount T T, a condition that is not common. Indeed, as indicated in Section 2.2, the almost universal presence of phenocrysts in rapidly quenched magmas indicates that most magmas are not superheated. In the case considered here, therefore, we can make the conditions more realistic by setting $T_s=T_0$, and at the same time simplify the problem.

The initial and boundary conditions set forth in the preceding problem hold true for this case as well. The moving surface of solidification, however, provides additional conditions. There must be continuity of temperature across the solid–liquid interface. Thus, at $x=x_s$, the temperatures of the solid and magma must be equal; that is, $T_s = T_m$. Furthermore, as the interface advances, heat is generated by the liquid–solid phase change. The heat liberated by the crystallization of a certain volume of magma is given by the product of the latent heat and density of the magma, $L\rho_m$. If the solid–liquid interface advances a distance dx_s, a quantity of heat $L\rho_m dx_s$ per unit area is liberated, and this must be removed by conduction. A further condition, then, is that the heat flux leaving the interface must equal the sum of the heat flux entering from the magma and the heat generated by the phase change; that is,

$$K_s\left(\frac{\partial T_s}{\partial x}\right) = K_m\left(\frac{\partial T_m}{\partial x}\right) + L\rho_m\left(\frac{\partial x_s}{\partial t}\right) \tag{5.14}$$

Note that each group of terms in this equation has units of W m^{-2}. If the solidification temperature, T_s, and initial magma temperature, T_0, are equal (no superheat), there will be no temperature gradient in the magma; that is, $\partial T_m/\partial x=0$. The heat flux leaving the interface is, therefore, equal simply to the heat flux generated by the phase change.

The solutions to this problem are given in Carslaw and Jaeger (1959). First, the position of the solid–liquid interface at time t is given by

$$x_s = 2\lambda\sqrt{k_s t} \tag{5.15}$$

where the dimensionless constant λ is obtained from the relation

$$\lambda[\sigma + \mathrm{erf}(\lambda)]e^{\lambda^2} = \frac{C_s T_s}{L\sqrt{\pi}}$$

by iterative numerical calculations (trial and error or using the Solver function in Microsoft Excel). In this equation, $\sigma = K_s k_c^{1/2}/K_c k_s^{1/2}$, with values of σ varying from 0.5 to 3.0 depending on the rock types in contact. Equation (5.15) reveals that, just as in the preceding case where no latent heat of crystallization was considered, the distance of a particular isotherm (in this case the solidification boundary) from the contact is proportional to $\sqrt{t}$ (see Fig. 2.1).

The temperature at the contact, $T_{x=0}$, is given by

$$T_{x=0} = \frac{\sigma T_s}{\sigma + \mathrm{erf}(\lambda)} \tag{5.16}$$

If the thermal properties of both the igneous and country rocks are identical (that is, σ = 1) and because the error

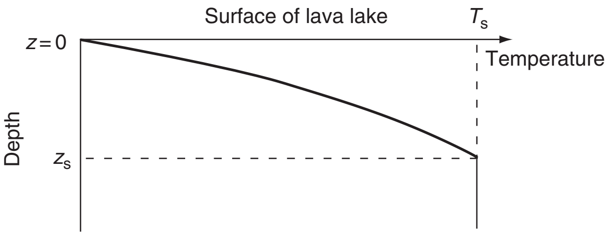


Fig. 5.5 Temperature distribution in the solidified crust of a lava flow or lake according to Eq. (5.19).

function must have a value between zero and unity, the contact temperature is a constant and has a value between T_s and 0.5 T_s. The latter case corresponds to the latent heat of crystallization being zero, and the result is identical to that obtained in the preceding problem, where latent heat was ignored.

Within the solidified region, the temperature is given relative to the contact temperature by the expression

$$T_{x=0}\left[1+\frac{1}{\sigma}\operatorname{erf}\left(\frac{x}{2\sqrt{k_s t}}\right)\right] \tag{5.17}$$

and in the country rock by

$$T_{x=0}\left[1+\operatorname{erf}\left(\frac{x}{2\sqrt{k_c t}}\right)\right] \tag{5.18}$$

The reader should again verify that these solutions do satisfy the initial and boundary conditions.

The cooling and solidification of a lava flow from its upper free surface can be treated in a similar manner (Fig. 5.5). Depth in the lava is measured as positive values of z, with the surface at $z = 0$. The surface temperature is adjusted to be zero, and the initial temperature of the lava is taken to be, as in the preceding problem, its solidification temperature T_s (no superheat). This temperature must also be adjusted by the same amount as the surface temperature. The thermal properties of the liquid and solid lava are taken to be identical. The temperature in the solid zone is then given by

$$T=\frac{T_s \operatorname{erf}\left(\frac{z}{2}\sqrt{kt}\right)}{\operatorname{erf}\lambda} \tag{5.19}$$

The value of the constant λ can be obtained by iterative calculations of the relation

$$\lambda e^{\lambda^2}\operatorname{erf}(\lambda)=\frac{C_s T_s}{L\sqrt{\pi}}$$

where L is the latent heat of crystallization and C_s is the heat capacity of the solid lava. Equation (5.19) again demonstrates that the thickening of the solidification crust is proportional to $\sqrt{t}$ (Problem 5.6).

The boundary conditions in the preceding problems require that there always be magma at its initial intrusion temperature at some distance from the contact. No matter how large the intrusion, this condition eventually breaks down and the center of the body begins to cool. At this point we have different boundary conditions and a new problem. Carslaw and Jaeger (1959) give the following solution for a sheet-like intrusion of thickness $2a$, in which distance, x, is measured from the center of the body;

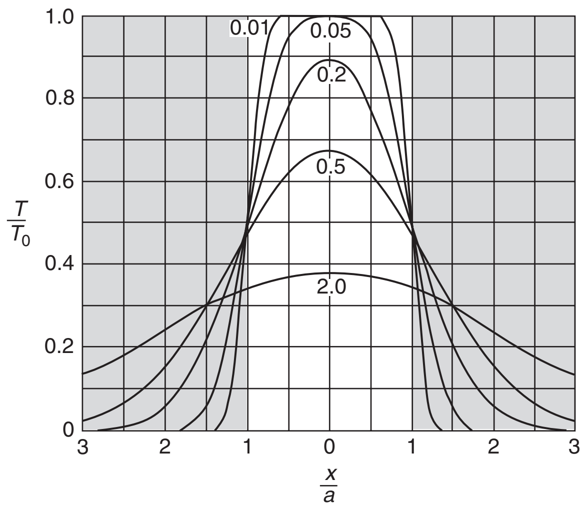


Fig. 5.6 Temperature distribution across a sheet-like intrusion of thickness 2*a* at various times after intrusion. Distance is given by the dimensionless term *x/a*, where *x* is measured from the center of the sheet. Times are given on the curves by the dimensionless term kt/a^2, and temperatures are expressed as fractions of the initial magma temperature T_0 (after Carslaw and Jaeger, 1959)

$$\frac{T}{T_0}=\frac{1}{2}\left[\operatorname{erf}\left(\frac{a-x}{2\sqrt{kt}}\right)+\operatorname{erf}\left(\frac{a+x}{2\sqrt{kt}}\right)\right] \tag{5.20}$$

A graphical representation of this equation is given in Figure 5.6.

The early period of cooling in a sheet is identical to that developed in the first problem, with the temperature of the contact remaining at one-half the initial adjusted magma temperature. As soon as the center of the sheet begins to cool, however, the temperature at the contact begins to fall. The time at which the maximum temperature is attained at any particular point increases with distance from the contact. Therefore, if minerals in contact metamorphic aureoles are formed at the time the maximum temperature is attained at any plane, those in the outer zones would form last. No account is taken of latent heat of crystallization in this solution. Of course, until the solidification surfaces advancing in from both contacts meet at the center, the solutions presented in Eqs. (5.16) to (5.18) can be used. Beyond this time, however, no simple solution involving latent heats of crystallization exists.

Winkler (1949) used Eq. (5.20) to obtain a mean cooling velocity that could then be related to the grain size variation at the margins of sheet-like bodies. This mean velocity is the temperature range through which the mineral crystallizes, divided by the time required to cool this amount. If crystallization begins at s_1T_0 and ends at s_2T_0, where s_1 and s_2 are simply fractions that give the appropriate temperatures in terms of T_0, the mean cooling velocity is given by

$$\frac{s_1 T_0 - s_2 T_0}{t} = \frac{4kT_0(s_1 - s_2)}{x^2\left\{[\operatorname{inverf}(2s_2 - 1)]^{-2} - [\operatorname{inverf}(2s_1 - 1)]^{-2}\right\}} \tag{5.21}$$

where x is the distance from the contact and inverf(z) is the inverse error function, such that if $y = \operatorname{inverf}(z)$, then $\operatorname{erf}(y) = z$. If we assume that grain size is inversely proportional to the mean cooling velocity, then, from Eq. (5.21), the grain size of the rock should be proportional to the square of the distance from the contact (see Chapter 12 for further discussion).

Because most rocks have very similar thermal diffusivities, grain sizes on opposite sides of dikes and sills are expected to show similar variations, even where the country rocks differ from one side to the other. This is not true for the cooling of a lava flow where the upper surface, in contact with air or water, cools much more rapidly than the lower one, which is in contact with rock or soil. In fact, the upper surface can be considered, for practical purposes, as instantly attaining some very low temperature, such as that of the air or water, and remaining at that temperature throughout the cooling of the flow. Equation (5.19) will then describe the rate at which cooling takes place from the upper surface. Cooling from the lower surface occurs more slowly by heat flowing into the underlying rocks; Eq. (5.17) describes this process. The temperature distribution within the flow becomes skewed as cooling proceeds, with the temperature maximum moving from its initial central position toward the base of the flow. From Figure 5.6 it is evident that if the rate of cooling of magma controls crystal grain size, the coarsest rock in a lava flow would be expected to occur below the center, whereas in a sill it would occur at the center, as long as there were no movement of the crystals after they had formed.

5.4 NUMERICAL ANALYSIS

Many heat flow problems are too complicated to have exact, analytical solutions. Nonetheless, useful answers can be obtained by numerical methods, which, in general, involve dividing the region of interest into a number of small volumes or cells. The heat flow through any cell can be analyzed in the same manner as was done earlier in this chapter when we derived Fourier's equation. In numerical analysis, however, instead of allowing the cell to shrink to an infinitesimal volume, as was done in deriving the differential equation, it is maintained at some small but finite volume. The overall thermal behavior of the region can then be determined by summing the effect of all the small cells. As the cells become smaller and more numerous, the results, of course, approach those of the exact solution of the differential equation. In most cases, however, sufficiently accurate results can be obtained with only modest numbers of cells. Because this work is invariably done on a computer, the number of cells used is limited only by the computational time.

Consider, for example, the cooling of an extensive lava lake whose upper surface is kept at ambient air temperature.

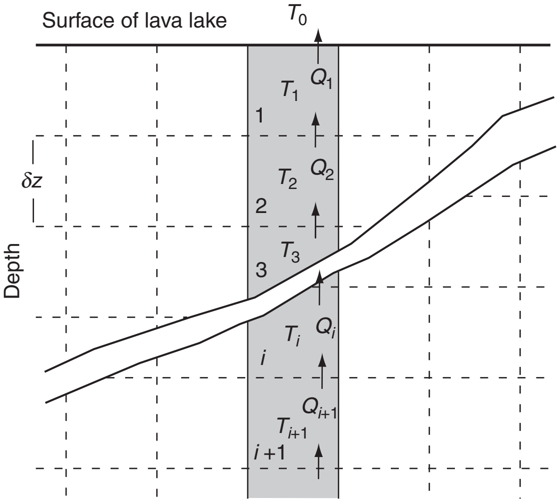


Fig. 5.7. The temperature distribution in a lava lake can be calculated numerically by first dividing the body into many small cells, and then determining the change in temperature of each cell as a function of time. See text for discussion.

This problem is considered by Shaw, Hamilton, and Peck (1977), whose paper should be referred to for details. Because of the lateral extent of the lava lake, heat can be considered to flow only in the vertical, $-z$, direction. The loss of heat around the sides of the lake will not be dealt with here. The lava lake is divided into numerous small cells, such as those illustrated in Figure 5.7. Each cell, which has a cross-sectional area A and vertical dimension δz, is numbered consecutively from the surface downward. Although each cell is shown as being of the same dimensions, this is not a requirement, and indeed computations in many problems can be reduced by allowing the cells to expand in less critical regions. The temperature of each cell, T_i, is taken to be at its center. Physical properties such as thermal conductivity K_i, thermal diffusivity k_i, heat capacity C_i, density ρ_i, and volume V_i can be specified for each cell. The heat flowing from one cell to the next is designated Q_i, and if heat were to flow in directions other than vertical, as occurs around the margins of the lake, there would be heat fluxes across the other faces of the cells.

According to Fourier's law (Eq. (5.1)), the quantity of heat passing from the (i+1)th cell into the ith cell (see Fig. 5.7) is

$$Q_{i+1} = -\frac{[K(T_i - T_{i+1})A\delta t]}{\delta z} \tag{5.22}$$

where δt is the time involved. Should adjoining cells have markedly different thermal conductivities, K would be the average of K_{i+1} and K_i.

To conserve thermal energy, the change in heat content of any cell, δQ_i, must be equal to the heat flux in, Q_{i+1}, minus the heat flux out, Q_i, plus any sources or sinks of heat, Q_i^*, within the cell. Sources, for example, could be latent heats of crystallization, heat generated by radioactive decay, and heat generated by the dissipation of viscous forces, if the magma were moving rapidly. Heat sinks, on the other hand, could include heat of evaporation of rainwater that descends cracks in the cooling crust of the lava lake, or endothermic reactions in a contact metamorphic aureole. The change in heat content of a cell, then, is

$$\delta Q_i = (\delta Q_{i+1} - \delta Q_i) + \delta Q_i^* \tag{5.23}$$

The change in heat content will, of course, result in a change in cell temperature. The heat capacity per cell is given by $C_i \rho_i V_i$, which when divided into the change in heat content, δQ_i, gives the change in temperature, δT_i, of the cell. Thus,

$$\delta T_i = \frac{\delta Q_i}{C_i \rho_i V_i} = \frac{Q_{i+1} - Q_i + Q_i^*}{C_i \rho_i V_i} \tag{5.24}$$

We are now able to apply Eqs. (5.22) and (5.24) to the problem of the cooling lava lake. The initial temperature of the lava, and thus of each cell, is taken to be 1200 °C. Loss of heat through the upper surface to the air, which is at a constant temperature of 25 °C, brings about cooling. Loss of heat would also occur through the base of the lava lake. Although this can be analyzed in precisely the same manner, only the cooling at the upper surface is treated here.

The air can be considered to occupy a cell immediately above the surface of the lava. Following eruption, the difference in temperature between this cell and the underlying one results in a quantity of heat, Q_i, flowing upward out of cell number 1 (Fig. 5.7). According to Eq. (5.22), this quantity is given by

$$Q_1 = -\frac{K(25-1200)A\delta t}{\delta z}$$

The value of Q_1 for any particular cross-sectional area A can be calculated once values of δt and δz have been selected.

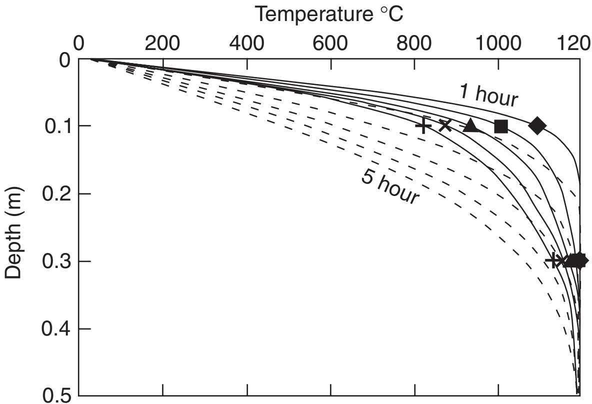
Values of δt and δz are chosen so as to provide an adequate description of the cooling history without using excessive amounts of computational time. If too large a time interval is chosen, however, meaningless results can be obtained. It has been shown (Shaw *et al.*, 1977) that provided the condition

$$\frac{K_i \delta t}{C_i \rho_i (\delta z)^2} \leq 0.25 \tag{5.25}$$

is satisfied, an acceptable or stable result will be obtained. Note that in this equation, $K_i/C_i\rho_i$ is the thermal diffusivity, k_i.

In the problem of the cooling lava lake, where thermal diffusivities of both molten and solidified lava are taken to be 10^{-6} m^2 s^{-1}, values of δz of 0.2 m and δt of 1 h (3600 s) satisfy the condition expressed in Eq. (5.25) [$k_i \delta t/(\delta z)^2 = 0.1$]. The quantity of heat crossing the upper surface of the lava in the first hour is, therefore, given by

$$Q_1 = -\frac{K(25-1200)A\ 3600}{0.2}\ \mathrm{J}$$



Cell	Depth	Temperature °C after time *t*				
		Hours				
#	(m)	1	2	3	4	5
0	0	25	25	25	25	25
1	0.1	1094	1008	936	875	824
2	0.3	1200	1190	1175	1156	1134
3	0.5	1200	1200	1199	1197	1194
4	0.7	1200	1200	1200	1200	1200
5	0.9	1200	1200	1200	1200	1200

Fig. 5.8 Temperatures near the surface of a lava lake at hourly intervals during the first five hours of cooling. Solid lines are calculated using the numerical method involving Eqs. (5.22) and (5.24), whereas the dashed lines are calculated using the exact solution of Eq. (5.13).

The quantities of heat transferred between each of the other cells would be calculated in a similar way, but at least during the first time interval, the lack of temperature differences between any of the other cells results in no heat being transferred across their boundaries. For example,

$$Q_2 = -\frac{K(1200-1200)A\ 3600}{0.2} = 0\ \mathrm{J}$$

Once heat fluxes across all cell boundaries are determined, the temperature changes in each cell can be calculated from Eq. (5.24). Let us assume for the moment that there are no sources or sinks of heat, which is the same assumption made in deriving the exact analytical solution to the differential equation (5.13). Equation (5.24) then reduces to

$$\delta T_i = \frac{k\delta t}{(\delta z)^2}(T_{i+1} - 2T_i + T_{i-1}) \tag{5.26}$$

which for cell number 1 gives a temperature lowering of 106 degrees, from 1200 °C to 1094 °C. Because no heat is transferred between the other cells during this first interval, their temperatures remain unchanged at 1200 °C. The new temperatures for each cell provide the input data necessary to calculate the heat fluxes during the next time interval, which in turn provide the new temperatures for each cell as cooling proceeds. The results for the uppermost five cells of the lava lake for the first five hours are presented in Figure 5.8. This type of calculation is ideally suited for a spreadsheet (see Problem 5.9, which allows you to compare the numerical and exact solutions).

The temperature gradients derived by the numerical calculation (solid lines in Fig. 5.8) have similar forms to those derived from the exact solution of the differential equation (dashed lines). With the exception of the first cell, temperatures calculated by the two methods are in reasonable agreement. Those in the first cell are considerably higher than they should be. This results from taking the air temperature as if it were for a cell immediately above the surface of the lava. Because temperatures are taken as being at the center of cells, the air temperature of 25 °C is consequently displaced 10 cm (1/2 δz) above its actual position. In this region of steep temperature gradients, this small shift causes a serious error in calculated temperatures, at least during the initial stages of cooling. At later stages, the error is drastically reduced. Of course, the error can be reduced and eliminated entirely by decreasing the size of the cells near the upper surface.

With sufficiently small cells and intervals of time, a numerical analysis can approach as closely as necessary an exact solution. The value of the numerical method, however, lies in the ease with which complexities can be incorporated. Heats of crystallization can be introduced in cells in a manner that accounts for the fraction of crystals formed at any particular temperature. Heats of metamorphic reactions in a contact aureole can be inserted at appropriate temperatures and locations. Furthermore, cooling studies need not be restricted to simple geometrical igneous bodies, for any irregular shape can be described by a three-dimensional configuration of small cells.

We will discuss how the latent heat of crystallization can be incorporated into the numerical solution for the cooling of a lava lake. If we are to include the Q_i^* term and the thermal diffusivity in all cells is taken to be the same, Eq. (5.24) reduces to

$$\delta T_i = \frac{k\delta t}{(\delta z)^2}(T_{i+1} + T_{i-1} - 2T_i) + \frac{Q_i^*}{C_i \rho_i V_i} \tag{5.27}$$

The quantity of heat Q_i^* depends on the amount of crystallization that takes place, which in turn depends on $-\delta T$. The latent heat of crystallization (L), which is typically ~400 kJ kg^{-1}, is released over the temperature range through which the magma solidifies (typically ~200 °C). The temperature at which crystallization begins is known as the *liquidus*, and the temperature at which the last liquid solidifies is known as the *solidus*. The amount of heat given out during any small temperature change, $-\delta T$, will be some fraction of the total latent heat of crystallization, which is determined by how much crystallization is caused by that temperature drop. In Chapter 10, we will see in detail how crystallization is related to temperature. Figure 5.9 shows three possible relations between temperature and degree of crystallization. In many magmas the degree of crystallization is nearly a linear function of temperature, and we will assume that is the case here. If we let the total temperature interval between the liquidus and solidus be ΔT (e.g. 200 °C in Fig. 5.9), the fraction of latent heat of crystallization liberated during any temperature interval $-\delta T$ will be $-\delta T_i \cdot L/\Delta T$, or if we express it in terms of per volume, $\delta T \cdot L \cdot \quad /\Delta T$. Therefore, as long as the magma is between the liquidus and solidus, we can express the amount of heat liberated during a temperature change of $-\delta T$ as

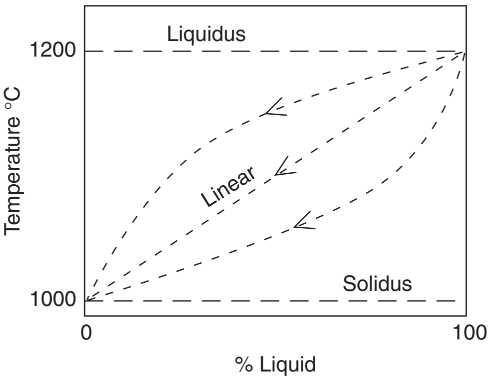


Fig. 5.9 Graph showing three different ways in which the percentage of liquid could decrease as a function of temperature as magma crystallizes. Many magmas follow closely the linear trend.

$$Q_i^* = -\frac{\delta T_i \cdot L \cdot \rho_i \cdot V_i}{\Delta T} \tag{5.28}$$

Using this expression, we can now write Eq. (5.27) as follows:

$$\delta T_i = \frac{k\delta t}{(\delta z)^2}(T_{i+1} + T_{i-1} - 2T_i) + \left(-\frac{\delta T_i \cdot L \cdot \rho_i \cdot V_i}{C_i \cdot \rho_i \cdot V_i \cdot \Delta T}\right)$$

which simplifies to

$$\delta T_i = \frac{k\delta t}{(\delta z)^2}(T_{i+1} + T_{i-1} - 2T_i)\left(\frac{C_i \cdot \Delta T}{C_i \cdot \Delta T + L}\right) \tag{5.29}$$

When calculating cooling temperatures where the latent heat of crystallization is taken into account it is necessary to use **If** statements. If the temperature of the *i*th cell is between the liquidus and the solidus then the temperature change will be given by Eq. (5.29), whereas if the temperature is outside this range there is no latent heat of crystallization involved, and the temperature change will be given by Eq. (5.26). For example, if a magma starts out at a liquidus temperature of 1200 °C (no superheat – the common situation) and the solidus is 1000 °C, then the **If** statement could be formulated as follows:

$$= \text{if}(T_i > 1000,\ \text{Eq. (5.29)},\ \text{Eq. (5.26)})$$

which is to say, if $T_i > 1000$, then use Eq. (5.29), else use Eq. (5.26) to calculate the temperature change (see Problem 5.12). If the magma began cooling from a temperature above the liquidus, the **If** statement could be expanded to include that possibility.

Another widely used numerical solution to heat conduction problems is the implicit finite-difference technique of Crank and Nicolson (1947). This solution can be used not only for simple heat conduction problems but also for problems involving heat sources and sinks and movement of the mass under consideration. Discussion of the latter complexities is left to Section 22.4; here we consider only the

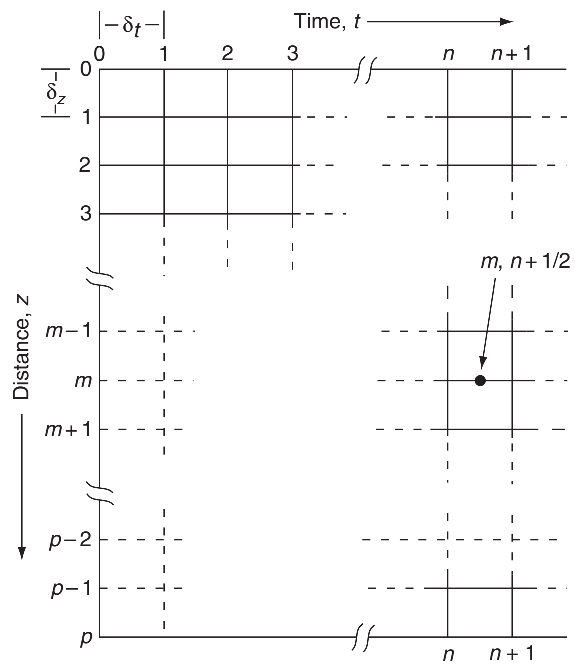


Fig. 5.10 Grid of time versus distance used in deriving the Crank–Nicolson (1947) finite-difference technique of numerical analysis of heat conduction problems. See text for discussion.

numerical solution of Fourier's equation (Eq. (5.11)) for the conduction of heat in only the z direction, which can be written as

$$\frac{\partial T}{\partial t} = k\left(\frac{\partial^2 T}{\partial z^2}\right) \tag{5.30}$$

Temperature in Eq. (5.26) is both a spatial (z) and temporal (t) function. We can represent these two variables on a grid (Fig. 5.10), with the number of grid increments of δz and δt being designated by m and n, respectively. To evaluate Eq. (5.30) numerically, the temperature T at any point m,n must be expressed in terms of the finite differences in δz and δt of the grid.

Given a temperature at any point in the grid, such as $T_{m,n}$ (Fig. 5.10), it is possible to express the temperature at any adjoining point $T_{m+1,n}$ by a Taylor series:

$$T_{m+1,n} = T_{m,n} + \frac{\partial T(\delta z)}{\partial z} + \frac{1}{2!}\frac{\partial^2 T(\delta z)^2}{\partial z^2} + \frac{1}{3!}\frac{\partial^3 T(\delta z)^3}{\partial z^3} + \ldots \tag{5.31}$$

Because points $T_{m+1,n}$ and $T_{m,n}$ are close, higher-order terms need not be included. We can also write the temperature of a point on the other side of m,n as

$$T_{m-1,n} = T_{m,n} + \frac{\partial T(-\delta z)}{\partial z} + \frac{1}{2!}\frac{\partial^2 T(-\delta z)^2}{\partial z^2} + \frac{1}{3!}\frac{\partial^3 T(-\delta z)^3}{\partial z^3} + \ldots \tag{5.32}$$

Adding Eqs. (5.31) and (5.32) together and rearranging gives

$$T_{m+1,n} + T_{m-1,n} - 2T_{m,n} = \left(\frac{\partial^2 T}{\partial z^2}\right)(\delta z^2)$$

or

$$\frac{\partial^2 T}{\partial z^2} = \frac{T_{m+1,n} + T_{m-1,n} - 2T_{m,n}}{(\delta z)^2} \tag{5.33}$$

Equation (5.33) expresses the second derivative in Fourier's equation in terms of the temperatures at several points in the grid and the finite difference δz between them.

Next we express the first term in Fourier's equation as a finite difference ratio by considering the temperatures at points m,n and $m,n{+}1$ on either side of a point $m,n+\frac{1}{2}$ (Fig. 5.10):

$$T_{m,n} = T_{m,n+1/2} + \tfrac{1}{2}(-\delta t)\left(\frac{\partial T}{\partial t}\right) \tag{5.34}$$

and

$$T_{m,n+1} = T_{m,n+1/2} + \tfrac{1}{2}(\delta t)\left(\frac{\partial T}{\partial t}\right) \tag{5.35}$$

Subtracting Eq. (5.34) from Eq. (5.35) and rearranging gives

$$\frac{\partial T}{\partial t} = \frac{T_{m,n+1} - T_{m,n}}{\delta t} \tag{5.36}$$

The second derivative of T with respect to z at this same point, $m,n+\frac{1}{2}$, can be obtained using Eq. (5.33) from the mean of the values determined for points $(m{-}1,n){-}(m,n){-}(m{+}1,n)$ and $(m{-}1,n{+}1){-}(m,n{+}1){-}(m{+}1,n{+}1)$; that is,

$$\frac{\partial^2 T_{m,n+1/2}}{\partial z^2} = \frac{T_{m+1,n} + T_{m-1,n} - 2T_{m,n} + T_{m+1,n+1} + T_{m-1,n+1} - 2T_{m,n+1}}{2(\delta z)^2} \tag{5.37}$$

Substituting Eqs. (5.37) and (5.36) into Eq. (5.30) gives

$$\frac{T_{m,n+1} - T_{m,n}}{\delta t} = \frac{k}{2(\delta z)^2}\big(T_{m+1,n} + T_{m-1,n} - 2T_{m,n} + T_{m+1,n+1} + T_{m-1,n+1} - 2T_{m,n+1}\big)$$

which on rearranging gives

$$T_{m,n+1}\left[1 + \frac{k\delta t}{(\delta z)^2}\right] = \frac{k\delta t}{2(\delta z)^2}\left(T_{m+1,n} + T_{m-1,n} + T_{m+1,n+1} + T_{m-1,n+1}\right) + T_{m,n}\left[1 - \frac{k\delta t}{(\delta z)^2}\right] \tag{5.38}$$

Equation (5.38) can be simplified further if we select values of δt and δz so that $k\delta t/(\delta z)^2 = 1$. Equation (5.38) then becomes

$$T_{m,n+1} = \tfrac{1}{4}\big(T_{m+1,n} + T_{m-1,n} + T_{m+1,n+1} + T_{m-1,n+1}\big) \tag{5.39}$$

Equation (5.39) expresses the temperature at a point in terms of four temperatures at points that differ from it by one finite difference in space and one finite difference in time; this temperature can be thought of as the mean of the four other temperatures.

To use Eq. (5.39) it is necessary to specify the initial and boundary conditions. For example, what was the initial temperature distribution, and what were the temperature gradients across the boundaries at any time *t*? These conditions would be given as follows:

$$\begin{aligned} &T = f(z) && \text{at } t = 0 && \text{for} \quad 0 < z < 1 \\ &\frac{\partial T}{\partial z} = H_0(T) && \text{at } z = 0 && \text{for} \quad t \geq 0 \\ &\frac{\partial T}{\partial z} = H_p(T) && \text{at } z = p && \text{for} \quad t \geq 0 \end{aligned}$$

That is, at $t = 0$ there is a temperature distribution which is a function of z; at any time there is a temperature gradient across the boundary at $z = 0$ which is a function H_0 of T; and at the other boundary, where $z = p$, there is a gradient that is a function H_p of T. Once these conditions are specified, Eq. (5.39) can be solved by trial and error, as will become clear in the example to follow.

Before using Eq. (5.39), however, it is necessary to examine how we can evaluate the temperatures along the top and bottom of the grid in Figure 5.10, that is, for values of $z = 0$ and $z = p$. Clearly, Eq. (5.39) cannot be used because two of the necessary temperatures would fall outside the grid and are thus undefined except in terms of the functions H_0 and H_p that define the boundary conditions. Crank and Nicolson (1947) express these temperatures as follows:

$$T_{0,n+1} = \tfrac{1}{2}\left(T_{1,n} + T_{1,n+1}\right) + \tfrac{1}{8}H_0\left(T_{0,n} + T_{0,n+1}\right) \tag{5.40}$$

and

$$T_{p,n+1} = \tfrac{1}{2}\left(T_{p-1,n} + T_{p-1,n+1}\right) + \tfrac{1}{8}H_p\left(T_{p,n} + T_{p,n+1}\right) \tag{5.41}$$

Of course, if the boundary temperatures are constant, these values can simply be entered into the grid.

Let us now apply Eqs. (5.39) to (5.41) to the problem considered in the previous numerical analysis, that of the cooling of a laterally extensive basaltic lava lake. The initial and boundary conditions are easily defined. At $t = 0$, the temperature at all depths in the lake is 1200 °C, except on the surface ($z = 0$), where it is instantly 25 °C, and it remains at this temperature while the lava at depth continues to cool. During the period of interest, lava with a temperature of 1200 °C continues to exist at depth in the lake. We will ignore the heat of crystallization.

The first step in the calculation is to select values of δt and δz so that $k\delta t/(\delta z)^2 = 1$. We can, for example, choose a time interval of half an hour (1800 s), in which case, $z = 0.0424$ m if $k = 10^{-6}$ m^2 s^{-1}. We can then construct a grid, as shown in Table 5.2, using these values in a spreadsheet. The first column in the table gives the depth in meters in the lava lake. The next column gives the initial temperatures, which are 1200 °C except at $z = 0$, where it is 25 °C. The temperature at $z = 0$ remains at 25 °C at all times; this value can therefore be entered in the first row of the table for all times. The remainder of the columns shows the temperatures in the lava lake following successive half-hour intervals (only the hourly values are shown in Table 5.2). These temperatures are calculated using Eq. (5.39), starting with the first time increment. In the spreadsheet, Eq. (5.39) is entered in each cell of the table except for cells at the greatest depth where Eq. (5.41) is entered. Because the equations make use of temperatures in cells that are yet to be calculated, the calculation is iterative and continues as long as the calculated temperatures are significantly different from the previously calculated ones. In Microsoft Excel the **Iteration** function needs to be set in the Preferences under Calculation.

Table 5.2 *Cooling of lava lake*

		Temperature (°C)				
		Time (h) of cooling after formation of lava lake				
Depth (m)	T_0	1	2	3	4	5
0	25	25	25	25	25	25
0.042	1200	473	349	292	257	233
0.085	1200	810	634	537	475	431
0.127	1200	1043	859	745	667	610
0.17	1200	1144	1013	907	827	764
0.212	1200	1181	1107	1023	951	890
0.254	1200	1194	1158	1100	1042	988
0.297	1200	1198	1182	1147	1104	1061
0.339	1200	1199	1193	1174	1145	1112
0.382	1200	1200	1197	1188	1170	1147
0.424	1200	1200	1199	1194	1184	1169
0.466	1200	1200	1200	1198	1192	1183
0.509	1200	1200	1200	1199	1196	1191
0.551	1200	1200	1200	1200	1198	1195
0.594	1200	1200	1200	1200	1199	1198
0.636	1200	1200	1200	1200	1200	1199

A plot of temperature distributions in the lava lake at hourly intervals during the first five hours is shown in Figure 5.11. At all times, calculated temperatures agree within a few degrees of those obtained with the exact solution of the differential equation (Problem 5.8). The Crank–Nicolson numerical method therefore provides a reliable means of solving heat flow problems. It can, for example, be used to solve any of the problems introduced in the earlier part of this chapter (see Problem 5.10 and 5.11). In Chapter 22 we see how it is used to include heat sources and mass movement of the body of rock being investigated.

5.5 COOLING BY RADIATION

Cooling by radiation from a surface is very much more rapid than heat conduction through a body. As a result, many lava flows rapidly cool and form a crust on their upper surface, while beneath, lava may remain molten and continue

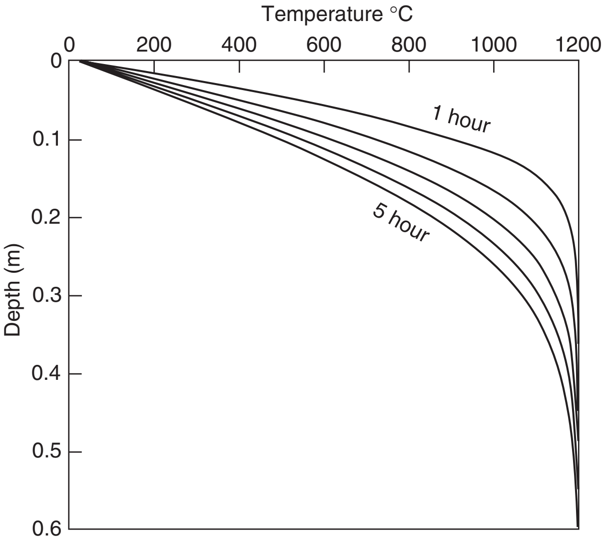


Fig. 5.11 Numerical solutions to the cooling of a lava lake using the method of Crank and Nicolson (1947). The five curves indicate the temperature distributions at hourly intervals during the first five hours of cooling of the lake. Calculated temperatures differ by no more than a few degrees from the exact solution of the differential equation (Eq. (5.13)). See text for discussion.

flowing for considerable time. The radiant heat emitted in time t from a perfectly radiating surface of area A, commonly referred to as *blackbody radiation*, is given by the *Stefan–Boltzmann law*,

$$Q = \sigma A T^4 t \tag{5.42}$$

where σ, the Stefan–Boltzmann constant, has a value of 5.6696×10^{-8} W m^{-2} K^{-4} and T is the absolute temperature. For all natural substances, which are not perfect radiators, the law is expressed as

$$Q = \varepsilon \sigma A T^4 t \tag{5.43}$$

where ε, the *emissivity*, must be measured for the particular material at the temperature of interest. Few data are available on values of ε for molten lavas, but an approximate value of 0.5 can be used (Shaw and Swanson, 1970) without introducing serious discrepancy, because the quantity of radiated heat is far more dependent on temperature.

Many basalts are extruded at approximately 1200 °C (1473 K); hence, they radiate heat at a rate of 133 kW m^{-2}. It is instructive to consider what thermal gradient would be necessary in the lava to supply, by conduction, this amount of heat to the surface for radiation. This can be determined from Eq. (5.3) by assuming the reasonable value of 2 W m^{-1} °C^{-1} for the thermal conductivity, in which case $\partial T/\partial x = J/K = 133/2 = 665$ °C cm^{-1}. Clearly, as soon as cooling takes place to a depth of 2 cm, the thermal gradient would not be great enough to cause sufficient heat to be supplied through conduction to keep up with the radiant cooling of the surface. For this reason the crust on a lava lake or flow may be cool enough to walk on despite the presence of molten material a few centimeters beneath.

In Chapter 4, ash flows and perhaps some flood basalts were said to travel turbulently, in which case heat lost from their upper surfaces by radiation is rapidly replaced by convection of hot material from within the flow onto the surface. The turbulence tends to keep the flow at the same temperature throughout and ensures that radiation from the upper surface is the main process of cooling (Boyd, 1961; Shaw and Swanson, 1970; Danes, 1972).

The cooling rate of a flow which is so turbulent that its temperature is essentially constant throughout, as might be the case for a rapidly moving ash flow, can be evaluated by expressing the Stefan–Boltzmann law in differential form,

$$-\frac{\mathrm{d}Q}{\mathrm{d}t} = \varepsilon \sigma A T^4 \tag{5.44}$$

The negative sign is placed before the derivative to indicate that the heat is being given up by the flow. As heat is lost, the temperature falls by an amount determined by the heat capacity of the cooling material. Because $\mathrm{d}Q/\mathrm{d}T = C_\mathrm{p}\rho V$, where V is the volume of the material and ρ its density, $C_\mathrm{p}\rho V dT$ can be substituted for dQ in Eq. (5.44), yielding on rearranging,

$$-\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\varepsilon \sigma A T^4}{C_\mathrm{p}\rho V} \tag{5.45}$$

Equation (5.45) gives the rate of change of temperature with time. Integration of this equation yields the total temperature change in a given interval of time. To do this, we multiply by dt and divide by T^4 to separate variables,

$$-\frac{\mathrm{d}T}{T^4} = \frac{\varepsilon \sigma A}{C_\mathrm{p}\rho V}\mathrm{d}t$$

If the emissivity and heat capacity remain constant over the temperature range of interest, we can write

$$-\int_{T_0}^{T} \frac{\mathrm{d}T}{T^4} = \frac{\varepsilon \sigma A}{C_\mathrm{p}\rho V}\int_0^t \mathrm{d}t$$

Which upon integration gives

$$\frac{1}{3T^3} - \frac{1}{3T_0^3} = \frac{\varepsilon \sigma A t}{C_\mathrm{p}\rho V}$$

or

$$T = \left(\frac{3\varepsilon \sigma A t}{C_\mathrm{p}\rho V} + \frac{1}{T_0^3}\right)^{-1/3} \tag{5.46}$$

Equation (5.46) gives the temperature to which cooling takes place in t seconds from an initial temperature T_0. If for purposes of illustration, we take the initial temperature of 1 g of basalt to be 1200 °C (1473 K), the emissivity to be 0.5, and the heat capacity to be 0.8 kJ kg^{-1} K^{-1}, then the temperature after 1 h (3600 seconds) would decrease to 300 K or 27 °C due to radiation from a surface of 1 cm^2.

The cooling rate of a flow, as determined from Eq. (5.46), is probably too high, because turbulence, unless extreme, is unlikely to bring hot material from within the flow at a sufficient rate to replenish the heat radiated from the surface. Also, the equation does not take into account any other sources of heat within the flow, such as latent heat of crystallization or conversion of gravitational energy through the dissipation of viscous forces. The first of these can be taken into account approximately by increasing the value of the heat capacity by an amount $L/\Delta T$, where L is the latent heat of fusion and ΔT the temperature range through which the heat is liberated. Heating from the dissipation of viscous forces is usually negligible in slowly moving flows, but in rapidly moving ash it may cause the temperature to rise some tens of degrees (Shaw, 1969).

5.6 DIFFUSION

At the beginning of our discussion of heat transfer, we stated that heat flows from regions of high temperature to regions of low temperature, never the reverse. This is one of a number of simple observations that give a direction to natural processes. In Chapter 7, we will see how such observations are used to formulated thermodynamic laws. Another simple observation is that material tends to move from regions of high concentration to regions of low concentration (there are some exceptions, but we will ignore them for now). The smell of a good meal cooking on the stove, for example, soon permeates the entire kitchen, or if you put cream in your coffee, the liquids start mixing even if you don't stir them. This movement of material at the atomic or molecular scale due to thermal motion is known as *diffusion*, in contrast with the movement of larger volumes of material in response to pressure gradients, which is known as *advection*. When you stir cream into coffee, you are moving large volumes of coffee and cream by advection, but eventually diffusion produces the homogeneous liquid. Because heat and matter both move from regions where the parameter that measures their intensity (temperature and concentration, respectively) is high to where it is low, it is not surprising that the equations describing the diffusion of heat and of matter are similar (Brady, 1995).

Even though a material may be at rest, its atoms are in thermal motion. Although this motion is random, in the presence of a concentration gradient it results in a net transfer of atoms down the gradient. This motion is known as *diffusion*.

In a gas, thermal motion causes atoms to travel in a zigzag path as a result of collisions with other atoms. The distance traveled between collisions, which is known as the *mean free path* (λ), is inversely proportional to the diameter (d) of the atoms and to their numbers (n); that is,

$$\lambda = \frac{1}{\pi d^2 n\sqrt{2}} \tag{5.47}$$

Because the value of n decreases with increasing temperature, the mean free path increases with increasing temperature. Increased pressure, on the other hand, increases the number of atoms in a given volume, and the mean free path decreases. At atmospheric pressure and room temperature, the mean free path of an atom in gas is ~1 μm.

The distance traveled by atoms in solids is much less than that in gases and is normally restricted to jumps from one site in the structure to another; the individual jumps are therefore ~0.3 nm. Crystals are not perfect, regular structures; most contain large numbers of defects, which produce regions where the atoms are not as closely packed or where vacancies may actually exist in the structure. Most diffusion involves these defects. For example, if a crystal structure contains vacancies, diffusion occurs when an atom or ion jumps from an adjoining site into the vacancy. At the same time, the vacancy moves to the site from which the atom jumped. This mechanism can therefore be considered as the diffusion of vacancies through the structure. Because grain boundaries have many vacancies and other defects, diffusion along grain boundaries is more rapid than diffusion through crystals (Brady, 1983). Some diffusion may not involve defects. For example, diffusion might simply involve the exchange in position of two atoms. This would obviously be energetically more difficult than the vacancy transfer mechanism because the atoms would have to squeeze past each other.

Because thermal vibrations are random, the mean free path of an atom in a gas or the jump direction in a crystal can be in any direction. But when a concentration gradient exists, statistically more jumps take place down the gradient than up, resulting in a net transfer of atoms down the gradient. Consider a zoned crystal with a compositional gradient in the x direction for some diffusing species of atom of $\partial c/\partial x$, where c is measured in number of atoms per unit volume (Fig. 5.12). Visualize the crystal being divided into thin slices that are Δx thick, where Δx is the length of individual jumps of the diffusing atoms. On either side of the shaded slice (Fig. 5.12), we have concentrations of c_1 and c_2, where $c_1 > c_2$. If the jump frequency is f, then in a short time interval dt, the number of jumps at the high-concentration end will be $c_1 A \Delta x f \mathrm{d}t$, where A is the area of a slice measured normal to x. Of these jumps, however, as many are likely to be to the

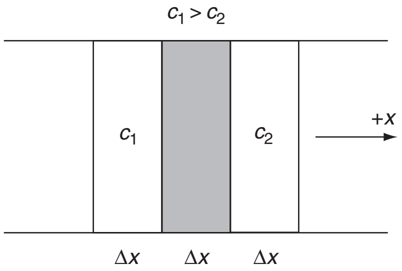


Fig. 5.12 A substance through which there is a compositional gradient can be thought of as consisting of small slabs, Δx thick, each with its own composition. Random jumps of atoms from one slab into the next are just as likely to occur to the left as they are to the right. If the concentration of a diffusing atom in a slab is c_1 and that in another slab is c_2, where $c_1 > c_2$, more jumps will occur from the high-concentration end into an intervening slab than will occur from the low-concentration end. This results in a net transfer of material down the concentration gradient as a result of diffusion.

right as to the left, and therefore the number of jumps down the gradient will be $^1/_2(c_1 A\Delta x f \mathrm{d}t)$. Jumps may occur in directions other than along the x-axis, but these need not concern us because only jumps in the x direction (+ or −) can contribute to diffusion in that direction. The number of jumps that take place up gradient from the low concentration end will be $^1/_2(c_2 A\Delta x f \mathrm{d}t)$. The net flux of atoms passing the unit area in the short time interval in the $+x$ direction is

$$J_x = {}^1\!/_2(c_1\Delta x f) - {}^1\!/_2(c_2\Delta x f) = {}^1\!/_2\Delta x f(c_1 - c_2) \tag{5.48}$$

Because Δx is very small, $c_1 - c_2 = -\Delta x\ (\partial c/\partial x)$. Thus

$$J_x = -\frac{1}{2}\Delta x^2 f\left(\frac{\partial c}{\partial x}\right) \tag{5.49}$$

The group of terms ($^1/_2\ \Delta x^2\ f$) is defined as the *diffusion coefficient*, D, which has units of m^2 s^{-1}. A compilation of diffusion coefficients in minerals, glasses, and melts is given in Brady (1995). Equation (5.49) can then be expressed as

$$J_x = -D\frac{\partial c}{\partial x} \tag{5.50}$$

This is known as *Fick's first law* of diffusion, and it shows that the diffusion flux is proportional to the concentration gradient. In Chapter 9, we will define a function, μ, known as the *chemical potential* (Eq. (9.2)) which is a function of composition (see Eqs. (9.26) and (9.34)). Because chemical potential is a function of concentration, the diffusion flux can also be expressed in terms of the chemical potential gradient; that is,

$$J_x = -D\frac{\partial \mu}{\partial x} \tag{5.51}$$

in which case the diffusion coefficient has units of mol^2 (m s J)$^{-1}$ (Brady, 1995).

Fick's first law is identical in its formulation to Fourier's law for heat flux (Eq. (5.3)) and to the Hagen–Poiseuille law for fluid flux (Eq. (3.10)). Consequently, the solutions to heat flow problems derived earlier in this chapter can be used for diffusion problems simply by replacing the temperature gradient with the concentration gradient and the thermal diffusivity with the diffusion coefficient. For example, to determine the rate of change of composition as a result of diffusion, we can follow a parallel derivation to that used for Fourier's equation (Eq. (5.11)) for the change of temperature with time. The resulting equation for diffusion down a gradient in the x direction is

$$\frac{\partial c}{\partial t} = D\left(\frac{\partial^2 c}{\partial x^2}\right) \tag{5.52}$$

which is known as *Fick's second law*.

In comparing the transport of heat and matter, it is interesting to note that values of thermal diffusivity in most geological materials are about 10^{-6} m^2 s^{-1}, whereas the diffusion coefficient of most components in magmas is 10^{-10} to 10^{-11} ^{2} $^{-1}$, and in most minerals at metamorphic temperatures is less than 10^{-20} m^2 s^{-1} (Fig. 5.13). Even diffusion along grain boundaries, which may be several orders of magnitude greater than that through crystals, is still very much slower than the diffusion of heat.

The jump of an atom from one position to another in a crystal structure is resisted by the surrounding atoms; this is illustrated in Figure 5.14. An atom must therefore have extra energy in order to make the jump, and only atoms with this energy are able to diffuse. Diffusion is therefore referred to as an *activated process*; that is, the atoms must possess an *activation energy* before they can jump. Like other activated processes, diffusion obeys an *Arrhenius*-type relation;

$$D = D_0 \exp\left(-\frac{E_D}{RT}\right) \tag{5.53}$$

where D_0 is a constant, E_D an activation energy, R the gas constant, and T the absolute temperature. The preexponential term, D_0, can be thought of as the diffusion coefficient at infinitely high temperature, and the exponential term gives the fraction of the attempted jumps that have sufficient energy to complete the jump. A plot of the log of the diffusion coefficient versus $1/T$ gives a straight line from whose slope we can obtain the activation energy (Fig. 15.13). An extensive list of values of D_0 and E_D for silicate minerals, glasses and melts is given by Brady (1995).

When an atom moves from one site to another, there is a momentary increase in volume (Fig. 5.14 (B). We might suspect therefore that increased pressure would decrease diffusion rates. If the momentary increase in volume is defined as an activation volume (ΔV_a), the effect of pressure on the diffusion coefficient at constant temperature can be expressed as

$$D = D_P \exp\left(-\frac{P\Delta V_\mathrm{a}}{RT}\right) \tag{5.54}$$

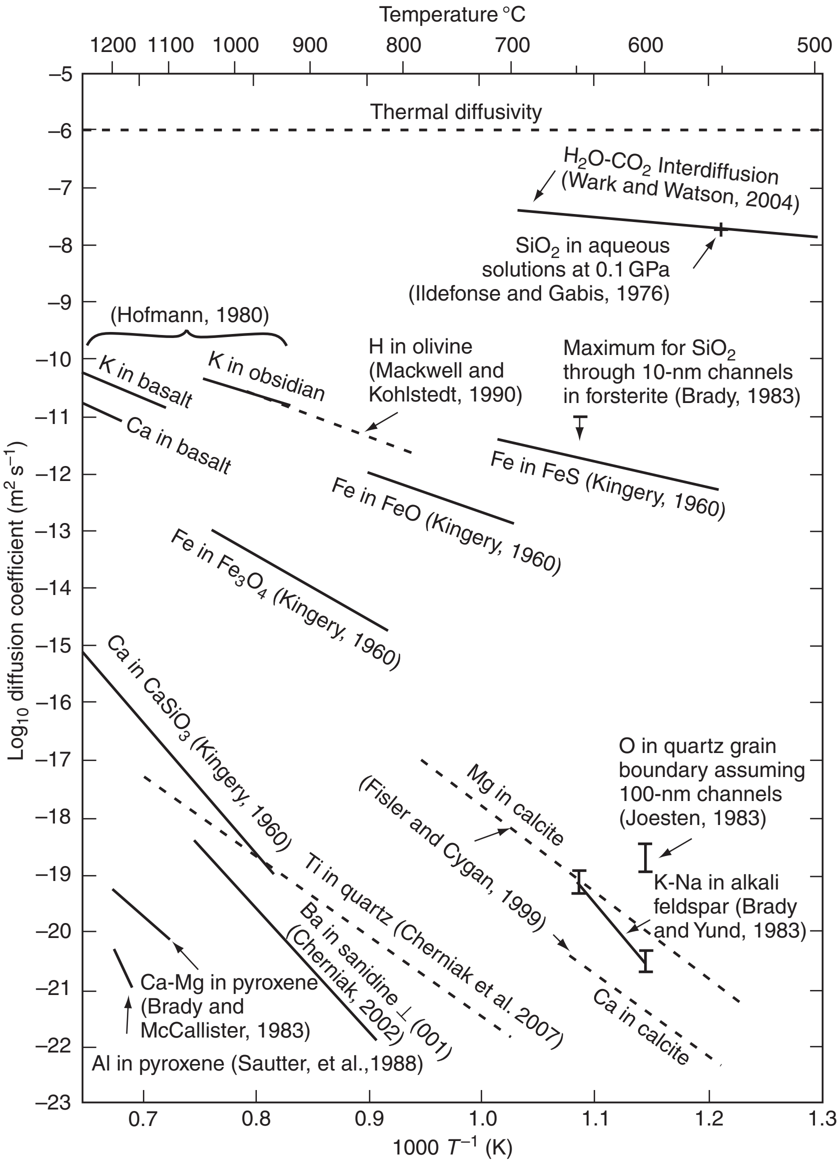
where D_P is a constant. The effect of pressure on diffusion is very small compared with that of temperature. During periods of deformation, however, directed stresses may induce defects in crystals, which in turn increase diffusion rates.

Fick's second law, as expressed in Eq. (5.52), applies only when the diffusion coefficient is constant. In many systems, D is compositionally dependent, in which case Eq. (5.52) becomes

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left[D\left(\frac{\partial c}{\partial x}\right)\right] \tag{5.55}$$

In magmas and minerals with complex compositions, the diffusion coefficient of one component would be affected by the concentration of all other constituents. The analysis of diffusion in multicomponent materials is therefore complex. One instance where the diffusion coefficient is unaffected by composition is *self-diffusion*; that is, the diffusion of a substance through itself. This is normally measured experimentally by introducing and tracking a radioactive isotope of the element in question. Self-diffusion is important to the migration of grain boundaries and the establishment of equilibrium textures in metamorphic rocks (Section 12.6).

Fig. 5.13 Plot of $\log_{10}$ of the diffusion coefficients for a variety of atoms in some minerals, solutions, and magmas versus 1000 T^{-1} (K). For comparison, the typical value of thermal diffusivity in rocks and magmas is included.



In many crystals, diffusion of one ion must be balanced by a counter diffusion of another ion in order to maintain electrical neutrality. The diffusion of these ions cannot, therefore, be treated as independent processes. In such cases, an *interdiffusion coefficient* is defined, which is the effective diffusion for the coupled constituents. Brady and Yund (1983), for example, define a K–Na interdiffusion coefficient for alkali feldspars. The flux of K in an alkali feldspar is then given by

$$J_{\mathrm{K}} = -D_{\mathrm{K-Na}} \frac{\partial c_{\mathrm{K}}}{\partial x} \qquad (5.56)$$

where the interdiffusion coefficient is ~10^{-21} $m^2\ s^{-1}$ at 600 °C (Fig. 5.13).

We will now examine a few simple diffusion problems in order to show that the solutions used for heat flow are equally applicable to chemical diffusion. The ubiquitous occurrence of basaltic blobs in siliceous volcanic rocks is commonly interpreted to indicate that siliceous volcanism is triggered by the intrusion of basaltic magma into the base of a siliceous magma chamber. Morgan *et al.* (2006) have used barium (Ba) diffusion gradients in sanidine crystals from the AD 79 eruption of Vesuvius that buried Pompeii (Fig. 4.36) to show that the eruption was probably caused by an influx of more mafic magma ~20 years prior to the eruption. Historic records show that a major earthquake occurred 17 years prior to the eruption, and this could have been caused by the same magma chamber recharge.

Barium is able to substitute mainly for K in the feldspar structure (celsian). The amount that enters the feldspar depends on the concentration of Ba in the magma. The phonolitic magma (see Chapter 6) that initially lay beneath Vesuvius had extremely low concentrations of Ba, and as a result, sanidine phenocrysts grew with essentially no Ba. When a surge of more mafic higher-temperature magma entered the chamber, the margins of the phenocrysts were rounded by resorbtion. Later, when sanidine began to

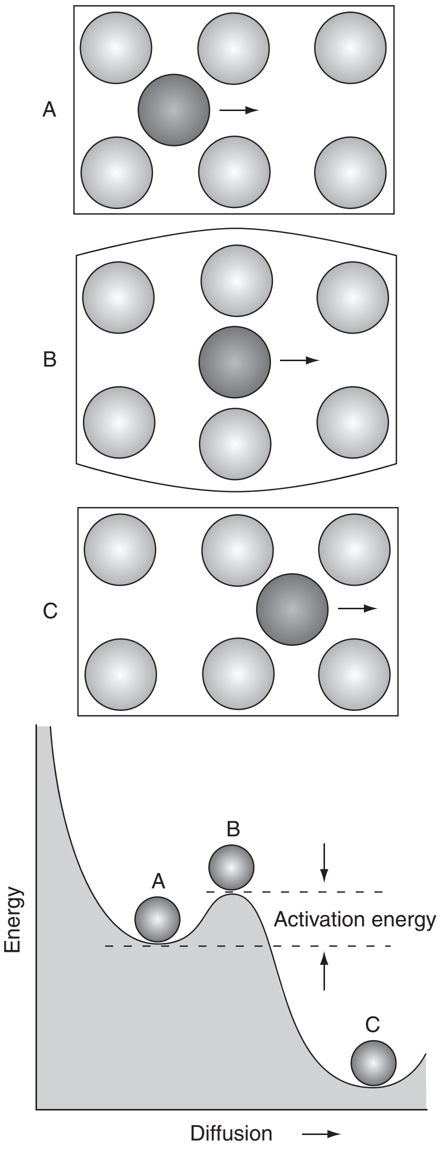


Fig. 5.14 Diffusion of an atom through a structure is an activated process; that is, when an atom moves from one position to another, it passes through an intermediate stage where the energy of the system is momentarily increased. This increase is known as the activation energy. In this example, the atom forces apart the structure of the crystal. The magnitude of the activation energy depends on the actual mechanism by which the atoms move. The number of atoms possessing sufficient energy to make the move is increased by raising temperatures. In the gravitational analogy below, the ball at A would like to decrease its potential energy by moving downhill to C, but to do so it must get over the small hump at B. The energy needed to get over this hump is the activation energy.

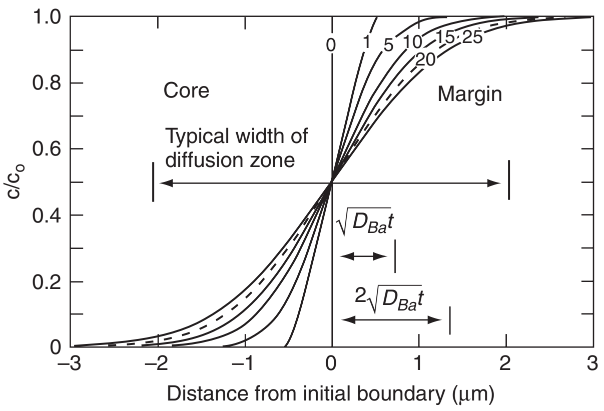


Fig. 5.15 Diffusion gradients of Ba across a planar boundary between the core of a sanidine crystal that contains no barium and a barium-rich rim at 0, 1, 5, 10, 15, 20, and 25 years after having been heated to 870 °C where $\log_{10}D_{\mathrm{Ba}} = -21.2$. The Ba content of the sanidine is expressed as a fraction relative to the initial concentration in the rim (c_0). The gradients satisfy Eq. (5.57). The width of typical diffusion boundaries in sanidine phenocrysts in the AD 79 pumice that buried Pompeii is ~4 μm, which corresponds to a diffusion time of ~20 years (Morgan *et al.*, 2006). The distances $\sqrt{D_{\mathrm{Ba}}t}$ and $2\sqrt{D_{\mathrm{Ba}}t}$ for 20 years of diffusion are shown for reference.

crystallize again, the magma had been enriched in barium, and consequently sanidine with several percent of the celsian component formed rims around the resorbed cores. The concentration gradient across the margin–core boundary initially formed a step function, with the concentration dropping abruptly from, for example, c_{Ba} in the margin to zero in the core (profile 0 in Fig. 5.15). This steep gradient caused the Ba to diffuse into the core as K diffused out into the margin. This process would have continued until eruption, when the magma and the diffusion process would have been quenched. From the measured Ba gradient and the known diffusion coefficient for Ba through feldspar, Morgan *et al.* (2006) were able to estimate the duration of this diffusion process.

To solve this problem we have simply to recognize its similarity to the first heat flow problem that we considered, that of the cooling of magma near a plane contact (Eq. (5.13)). We replace temperature with composition and thermal diffusivity with the Ba diffusion coefficient. The concentration of Ba, c_{Ba}, can then be expressed as a function of time and distance x (measured from the margin–core boundary) as follows

$$\frac{c_{\mathrm{Ba}}}{c^0_{\mathrm{Ba}}} = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{x}{2\sqrt{D_{\mathrm{Ba}}t}}\right) \tag{5.57}$$

where c^0_{Ba} is the initial concentration of Ba in the margin of the sanidine crystals. Had the core of the crystal contained some Ba, all concentrations would have been adjusted to reduce the core concentration to zero. This is equivalent to our having adjusted the temperature scale in the cooling problem to make the country rock temperature zero.

To use Eq. (5.57), we must know the temperature, because the diffusion coefficient of Ba in sanidine (Cherniak, 2002) is strongly dependent on temperature (Fig. 5.13). From heating experiments on melt inclusions in the sanidine crystals, Morgan *et al.* (2006) estimate that the temperature was ~870 °C. Cherniak (2002) gives the temperature dependence of the diffusion coefficient of Ba in sanidine as

$$D_{\mathrm{Ba}} = 0.0071\exp\left(-\frac{416\,000}{RT}\right)\mathrm{m^2\ s^{-1}} \tag{5.58}$$

from which it follows that, at 870 °C (1143 K), $\log_{10}D_{\mathrm{Ba}} = -21.2$. After substituting this value into Eq. (5.57), we can calculate Ba concentration gradients across the margin boundary of the sanidine phenocrysts for various times, *t*, and compare these with the measured gradients (Fig 5.15).

The typical width of the Ba gradient in going from the core to the marginal zone of the sanidine phenocrysts is ~4 μm. Equation (5.57) indicates that following one year the diffusional zone would be only 1 μm wide (Fig. 5.15). Following 20 years, however, this zone would have grown to ~4 μm (dashed line in Fig. 5.15). Morgan *et al.* (2006), consequently, conclude that the magma chamber beneath Vesuvius was recharged with hotter, more Ba-rich magma approximately 20 years prior to the devastating eruption of AD 79.

It turns out that we can estimate how long it will take for "significant" diffusion to occur without even having to solve Eq. (5.57). Consider the diffusive loss of Ba from the crystal margin. We can define the *characteristic length scale for this diffusion* as the distance into the margin where the Ba concentration is midway between its minimum and maximum values. The minimum concentration is $0.5c^0_{\mathrm{Ba}}$, the value at the contact with the core, and the maximum is c^0_{Ba} (cf. Fig. 5.15). The concentration midway between these values is simply $0.75c^0_{\mathrm{Ba}}$. Substitution of $0.75c^0_{\mathrm{Ba}}$ for c_{Ba} in Eq. (5.57) yields

$$0.5 = \operatorname{erf}\left(\frac{x}{2\sqrt{D_{\mathrm{Ba}}t}}\right) \tag{5.59}$$

where it is understood that x is the distance at which $c_{\mathrm{Ba}} = 0.75c^0_{\mathrm{Ba}}$. From Table 5.1, we see that erf (0.5) is very nearly 0.5. It follows that the term in parentheses in Eq. (5.59) is approximately 0.5. Consequently, we can write

$$x \sim \sqrt{D_{\mathrm{Ba}}t} \tag{5.60}$$

The solution for Eq. (5.57) is symmetrical about $x = 0$, so $\sqrt{D_{\mathrm{Ba}}t}$ is also the characteristic distance for diffusion of Ba into the core. This simple relationship, which shows that the length scale for diffusion is proportional to $\sqrt{Dt}$, is one of the most important results presented in this chapter. Because of the square root dependence, diffusion over short length scales occurs much more rapidly than diffusion over long length scales. Similarly, the rate of compositional change will be fastest in the early stages of diffusion and will slow with time (Fig. 5.15).

Diffusion penetrates farther than just $\sqrt{Dt}$ for a given time t, so it is common to take a multiple of $\sqrt{Dt}$ as the distance for "significant" diffusion; $2\sqrt{Dt}$ is often used (Fig. 5.15). Exactly analogous reasoning applies for heat conduction. From Eq. (5.13), we can take $2\sqrt{kt}$ as the length scale for significant cooling of the intrusion, or for heating of the country rock.

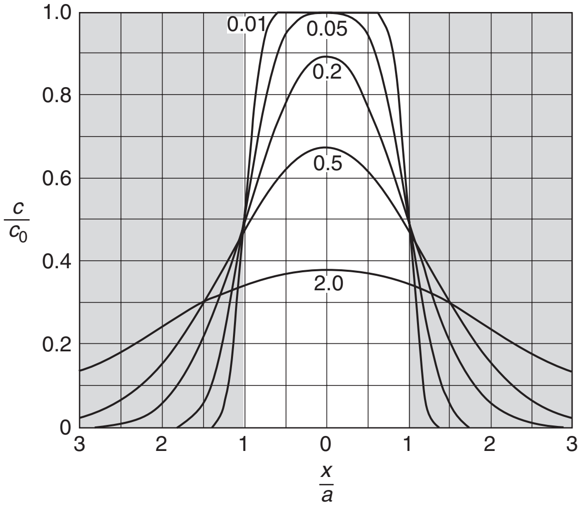


Fig. 5.16 Diffusion profiles across a K-rich exsolution lamella in a Na-rich alkali feldspar host that has started to homogenize on being heated. The lamella has a half-width of *a*, and distance is given as the ratio of *x/a*, with *x* being measured from the center of the lamella. Time is indicated on the gradients by the dimensionless term *Dt/a*^{2}. The gradients satisfy Eq. (5.61). See text for discussion.

For the solution presented in Eq. (5.57) to be valid, material with the initial concentrations must continue to be present at ever-increasing distances from the initial boundary. When this condition is no longer valid, a formulation equivalent to Eq. (5.20) must be used for diffusion. Consider, for example, a xenolith of granite that is suddenly incorporated into a hot basaltic dike. If the granite contains alkali feldspar crystals that contain potassium-rich exsolution lamellae with a composition of Or in the host alkali feldspar of composition Or , the heat from the basalt will cause the exsolution lamellae to start to diffuse into the surrounding crystal (Fig. 5.16). The distance *x* is measured from the center of a lamella, which has a half-width of *a*. We will assume that the lamellae are far enough from each other that the host is able to remain at a composition of Or_2 at some distance from a lamella. This problem is similar to that of the cooling of a dike (Eq. (5.20)). We can write the equivalent equation for diffusion of the K and Na through the feldspar as

$$\frac{\mathrm{Or} - \mathrm{Or}_2}{\mathrm{Or}_1 - \mathrm{Or}_2} = \frac{1}{2}\left[\operatorname{erf}\left(\frac{a-x}{2\sqrt{Dt}}\right) + \operatorname{erf}\left(\frac{a+x}{2\sqrt{Dt}}\right)\right] \tag{5.61}$$

The compositional gradients across the lamella are shown in Figure 5.16 for different times represented by the dimensionless term Dt/a^2. As before, the characteristic length scale for diffusion is $\sqrt{Dt}$.

Problem 5.15 and 5.16 make it clear that diffusion through crystals cannot be responsible for regional-scale transport of material. Nature, however, provides many examples where large-scale compositional change has occurred. Skarn deposits in contact metamorphic zones, for example, often testify to the introduction of large amounts of iron into carbonate rocks that can be of economic value, and many sulfide ore deposits are surrounded by large alteration haloes, which are often used in prospecting for these deposits. These large-scale compositional changes must involve advection, where migrating fluids have acted as transporting agents (see Chapter 21). Diffusion through crystals, of course, remains an important process at the meter to centimeter scale; that is, at the scale at which rocks develop their grain size and texture. Furthermore, as we will see in Chapter 21, diffusion through pore fluids may occur at the meter scale or larger, thus facilitating significant transfer of material between rock layers.

5.7 PROBLEMS

5.1 Calculate the thermal gradient in °C km^{-1} in a continental area and on a mid-oceanic ridge where the heat flux at the Earth's surface is 60 and 320 mW m^{-2}, respectively. Assume the thermal conductivity of the rocks in both areas to be 2.0 W m^{-1} $°\text{C}^{-1}$.

5.2 If a temperature gradient of 1 °C m^{-1} causes heat to flow into a rock having a thermal conductivity of 2.0 W m^{-1} $°\text{C}^{-1}$, a heat capacity of 1.0 kJ kg^{-1} $°\text{C}^{-1}$, and a density of 3 Mg m^{-3}, how long would be required for the temperature of 1 m^3 of the rock to be raised 1 °C? What is the thermal diffusivity of this rock, and how is it related to the answer to the first part of this question?

5.3 The Palisades along the west bank of the Hudson River opposite New York City are formed by a 300-m-thick lower Jurassic diabase sill. Along most of its length, a layer 15 m above the lower contact contains high concentrations of olivine that may have accumulated by sinking from the overlying magma (Section 14.3). By calculating the length of time the sill took to solidify up to this layer, we can set a limit on how long the crystals would have had to settle. If the sill was intruded at 1200 °C into sedimentary rocks at 50 °C, calculate the time necessary for the zone of solidification to advance 15 m up from the base. Solidification is assumed to have occurred when the temperature reached 950 °C. The thermal diffusivity of magma, diabase, and country rock is taken to be 10^{-6} m^2 s^{-1}. Ignore the latent heat of crystallization. Equation (5.13) can be used if magma in the center of the sill was still at its initial temperature. The magma is assumed not to have been convecting at the time (in Section 14.4, we will see that this is not a good assumption).

5.4 Using Figure 5.6, calculate the length of time after intrusion for the temperature at the center of the Palisade sill to start decreasing. All conditions are the same as in Problem 5.3. Was the assumption that allowed Eq. (5.13) to be used in Problem 5.3 justified?

5.5 Repeat Problem 5.3, but include a latent heat of crystallization of 418 kJ kg^{-1}, which can be incorporated by raising T_0. The heat capacity is 0.8 kJ kg^{-1} $°\text{C}^{-1}$. Compare your answer with the answer obtained in Problem 5.3.

5.6 In late 1959 and early 1960 the eruption of Kilauea Iki, Hawaii, formed a 100-m-deep lava lake. The surface of the lake was rapidly covered with a crust that began to thicken with cooling, but lava with the initial temperature of 1150 °C still existed at depth in the lake for many months following eruption.

- **(a)** Using the 1065 °C isotherm as the base of the crust (50% crystalline), calculate and plot a graph of the crustal thickness versus the square root of cooling time in months, for the early period during which magma with the initial temperature still existed at depth in the lake. Assume that the surface of the lava lake remained at a constant air temperature of 25 °C, the thermal diffusivity of molten and solid lava is 10^{-6} m^2 s^{-1}, and the latent heat of crystallization can be ignored. (Hint: Equation (5.13) can be used if it is recognized that the surface temperature of the lava, like the contact temperature of an intrusion, remains constant. The trick is to adjust the initial temperatures so that the surface of the lava lake is at one-half the initial magma temperature. It will then satisfy Eq. (5.13). It may help to pretend that imaginary country rock overlies the lava lake.)
- **(b)** Compare and discuss your theoretical cooling curve with the actual measured data given in Figure 2.3.

5.7 Continuous flow (especially turbulent) of magma past a contact can heat country rocks to temperatures where partial melting occurs. This is particularly common where diabase feeder dikes intrude quartzo-feldspathic rocks. If the continuous flow of magma maintains the contact temperature at 1000 °C, calculate the length of time required for partial melting to progress outward 2 m from the contact. The country rocks are at a temperature of 100 °C prior to intrusion and begin to melt at 700 °C. Assume the thermal diffusivity of all materials to be 10^{-6} m^2 s^{-1}, and the latent heat of fusion of the country rocks can be ignored as long as the amount of partial melting is small.

5.8 If a basaltic lava lake is formed with an initial temperature of 1200 °C, calculate the temperature gradients in the upper part of this lake for hourly intervals during the first 5 h. The air temperature (hence the upper surface of the lava lake) is 25 °C, and the thermal diffusivity of the basaltic magma and solid crust is 10^{-6} m^2 s^{-1}. Latent heat of crystallization will be ignored. This problem can be solved using Eq. (5.13) if it is remembered that in this solution the contact temperature remains constant, as does the surface temperature of the lava lake. You can therefore think of the air above the lava as an imaginary country rock. The trick is to adjust temperatures correctly.

5.9 Repeat Problem 5.8, but use the first numerical method outlined in Section 5.4. Use 20-cm-deep cells, and calculate the temperature gradients for hourly intervals. Compare the results with those obtained in Problem 5.8. If a computer is used to solve this problem, explore what happens if the size of the cells or the time intervals are changed. Can you make the numerical solution more closely approach the results in Problem 5.8?

5.10 Repeat Problem 5.9, but use the Crank–Nicolson finite difference technique to solve the problem.

5.11 With the same spreadsheet as used in Problem 5.10, solve Problem 5.7 numerically. Be certain that country rock at 100 °C always exists at some distance from the contact and that the contact temperature is maintained at 1000 °C. Note that the temperature scale does not need to be adjusted in the numerical solution.

5.12 The object of this problem is to calculate the cooling history of the 15-m-thick Icelandic basaltic lava flow shown in Figure 4.6 using the numerical method of

Shaw *et al.* (1977) and to determine if the thermal maximum in the flow during cooling corresponds to the colonnade–entablature boundary. The spreadsheet generated in Problem 5.9 can be used, but in this problem, the latent heat of crystallization will be taken into account. Consequently, the formulas used for calculating the temperature must involve **If** statements to determine whether Eq. (5.26) or (5.29) should be used to calculate the temperature change during each time increment. Use a depth increment of 2 m and a time interval of 10 days. Extend the depth to 60 m and the time to 600 days. The surface temperature of the lava and the temperature of the rocks beneath the lava are initially at 10 °C. Set the temperature on the surface of the lava and at a depth of 60 m to 10 °C at all times. Temperatures at other depths can change as heat is released from the surface of the lava or into the underlying rock. The lava is erupted at its liquidus temperature of 1200 °C, and its solidus is 1000 °C. The latent heat of crystallization is 400 kJ kg^{-1}, which is released linearly between the liquidus and solidus. The heat capacity of the lava is 1.0 kJ kg^{-1} K^{-1}. The thermal diffusivity of the lava and all rocks is 10^{-6} m^2 s^{-1}. Graph thermal profiles to a depth of 20 m at various times during the first 600 days. Draw a line tracking the position of the temperature maximum on each of these profiles. If the tips of the fractures propagating down from the surface and up from the base of the flow follow the 800 °C isotherm, where would you expect the colonnade–entablature boundary to be located? Measure the position of this boundary in Figure 4.6 and comment on the agreement between the positions of the observed and calculated colonnade–entablature boundary.

5.13 Many small particles of molten material ejected from volcanoes cool and solidify to glass before landing on the surface. If a 10^{-4} kg spherical particle of lava with emissivity of 0.5, heat capacity of 0.8 kJ kg^{-1} °C^{-1}, and density of 3 Mg m^{-3} is ejected from a volcano at 1127 °C, how long would it have to travel through the air to cool by radiation to 827 °C, at which temperature it becomes a rigid glass? Neglecting air resistance, how high would the particle have to be ejected? (Volume of sphere = $4/3\pi r^3$ and surface area of sphere = $4\pi r^2$.)

5.14 From the plot of $\log_{10}D$ versus T^{-1} K (Fig. 5.13), calculate the activation energy for diffusion of Ca in wollastonite.

5.15 Two large alkali feldspar crystals, one of Or_{100} composition and the other of Or_{60} composition, are in contact along a planar surface. If they are suddenly heated to 600 °C by being incorporated as a xenolith in a body of magma, how many years would it take the Or_{100} crystal to change its composition to Or_{90} 1 mm from the initial contact between the crystals? Feldspar of both Or_{100} and Or_{60} composition continue to exist at some distance from the initial boundary. The K–Na interdiffusion coefficient is 3×10^{-21} m^2 s^{-1}.

5.16 If an alkali feldspar of $\mathrm{Ab}_{10}\mathrm{Or}_{90}$ composition contains widely spaced 10-μm-thick exsolution lamellae of $\mathrm{Ab}_{90}\mathrm{Or}_{10}$, how many years would it take at 650 °C, where the K–Na interdiffusion coefficient is 10^{-19} m^2 s^{-1}, for the composition in the center of a lamella to reach a composition of $\mathrm{Ab}_{15}\mathrm{Or}_{85}$? Assume that the lamellae are far enough apart that host crystal of composition $\mathrm{Ab}_{10}\mathrm{Or}_{90}$ is always present.

5.17 At low pressure, olivine does not normally contain water. However, Demouchy *et al.* (2006) found that olivine in mantle-derived nodules brought up from a depth of ~70 km by an alkali basalt in Patagonia, Chile, contained up to 312 ppm H_2O. This indicates that at high pressure hydrogen diffuses into the olivine structure and that water is present at this depth in the mantle. As the nodule approached the surface and the pressure decreased, hydrogen diffused out of the olivine crystals to produce dehydration rims. The resulting profiles resemble those shown in Figure 5.16.

(a) If the concentration of H_2O in the center of a 2-mm-wide crystal (distance measured parallel to [001]; assume a great distance in other dimensions) had been decreased during ascent from 312 to 105 ppm at a temperature of 1245 °C, where the diffusion coefficient for hydrogen through olivine parallel to [001] is 5.2×10^{-11} m^2 s^{-1}, and the concentration of H_2O in equilibrium on the rim of the crystal is zero, how long must diffusion have been operating?

(b) Knowing the length of time that hydrogen had to diffuse out of the olivine, and if diffusion began as soon as the olivine started to rise from its source at a depth of 70 km, calculate the average rate of ascent of the olivine. Is this rate comparable to the velocities of basaltic magmas calculated in Chapter 3?

6 Classification of igneous rocks

6.1 INTRODUCTION

Much has been written about the classification of igneous rocks, and many older petrology texts devote lengthy chapters to this topic. Historically, in any of the natural sciences, classification is extremely important because it forms the basis on which the science grows. At first, definitions are necessary simply to permit communication. The organizing and systematizing of the defined quantities then lead to the development of genetic models that attempt to explain the apparent order revealed by the classification. At this stage, modification of some of the original definitions may improve the classification by simplifying the genetic models. Through such reiterations it is hoped the order in classifications can be made to match more closely the order that presumably exists in nature. Some classifications have approached this goal more closely than others, whereas some have completely failed and are no longer used. To the uninitiated, the subject of classification may appear shrouded in controversy. It has, indeed, involved considerable debate as classifications evolved. But today the matter raises little controversy and does not warrant lengthy consideration.

What is required of a modern classification of igneous rocks? First, it must provide definitions of rock types that are easy to apply. Second, it should not introduce divisions or groupings that are unnatural; that is, ones that violate basic physical chemical laws, such as those of thermodynamics. Finally, the classification should be consistent with present models of the genesis of rocks. With these aims, the International Union of Geological Sciences (IUGS) Subcommission on the Systematics of Igneous Rocks has recommended a classification and nomenclature for igneous rocks (Streckeisen, 1976, 1979; Le Bas *et al*., 1986). This is the classification that will be dealt with mainly in this chapter and used elsewhere in this book. A full discussion of early classifications can be found in Johannsen (1931).

Before discussing the classification of igneous rocks it is necessary to consider their basic constitution. Earth, like the other terrestrial planets (Mercury, Mars, Venus, and Moon), consists essentially of those elements that were present in the original solar nebula, with the exception of the volatile elements, which were either not accreted in large amounts during planet formation or have since been lost from the planet. Table 6.1 lists the average composition of the solar system based on spectral analyses of the Sun. The relative concentrations of the nonvolatile elements in the Sun can be seen to be similar to those of the elements in the estimated bulk composition of the Earth (Table 6.1). Oxygen is by far the most abundant element in the Earth and, in terms of volume, constitutes more than 90% of the planet. Because of its preponderance, most rock and mineral analyses are more conveniently expressed in terms of oxides rather than of elements. Silicon is the next most abundant element. It is not surprising, then, that with the exception of the metallic core, most of Earth is composed of silicates. These silicates, which involve primarily the next four most abundant elements, Mg, Fe, Al, and Ca, include clino- and orthopyroxenes, olivine, and garnet. Na and K also form silicates (feldspar and feldspathoid), but these are concentrated primarily in the crust of Earth, as is quartz. Some of the primordial hydrogen is retained in Earth as water, which can combine with the abundant elements to form hydrous silicates, such as amphiboles and micas.

Differentiation of Earth into core, mantle, and crust caused major redistribution of the accreted elements. The core formed soon after accretion, and the formation of the continental crust was essentially complete by 2.1 Ga before present (see Section 13.3). Differentiation involving sinking of Fe and Ni to form Earth's core concentrated the remaining elements in the mantle above. Low-density partial melts formed in the mantle during the Archean rose toward the surface to form Earth's crust. This differentiation produced a crust enriched in silicates of Al (Sial) and a mantle enriched in silicates of Mg and Fe (Sima). The estimated average compositions of the mantle and continental crust are given in Table 6.1.

6.1.1 Meteorites

Much of our understanding of the composition of Earth is based on the study of meteorites. *Meteoroids* are meteors that enter Earth's atmosphere. Most break up and disintegrate due to intense frictional heating. However, a few of the larger ones make it to Earth's surface; we call these *meteorites*. Most meteorites are derived from asteroids. The asteroid belt between Mars and Jupiter is the primary source, although asteroids are also found scattered throughout the solar system. Meteorites and the composition of Earth are enormous topics; we have space only for a brief introduction. The reader is referred to Ringwood (1975), Brown and Mussett (1981), McDonough and Sun (1995), Davis (2003), Dalrymple (2004), and references cited therein for more complete coverage.

Table 6.1 *Abundance of elements normalized to Si = 1000 atoms*

	Solar	Whole Earth	Mantle	Continental crust
H	2.2×10^7	–	–	–
He	1.4×10^6	–	–	–
O	1.5×10^4	3407	3685	2949
C	9.3×10^3	–	–	–
Ne	3.4×10^3	–	–	–
N	1.9×10^3	–	–	–
Si	1000	1000	1000	1000
Mg	890	1062	1292	144
Fe	710	1226	149	88
S	360	99	–	–
Ar	120	–	–	–
Al	74	93	92	310
Ca	50	79	77	158
Ni	43	65	4	–
Na	43	12	12	85
Cr	11	17	7	–
P	7	13	1	3
Cl	7	–	–	–
Mn	6	2	2	3
K	3	1	1	38
Ti	2.5	3	2	12
Co	2	3	–	–

– indicates abundance less than 1.
Data from Henderson (1982)

Meteorites are classified largely on the basis of iron content and degree of differentiation. Three main types are commonly recognized. *Iron meteorites*, or irons, are composed predominantly of Fe–Ni alloy. Slow cooling produces the *Widmanstätten* exsolution pattern, characterized by intergrown crystals of Ni-poor and more Ni-rich alloy (Fig. 6.1(A)). Some irons are almost certainly fragments of the cores of asteroids that were large enough to undergo differentiation. Differentiation processes also produced Earth's core and mantle. Seismic studies reveal that the density of Earth's core is similar to that of Fe–Ni alloy at high pressures and temperatures. However, the outer core is somewhat less dense than predicted for Fe–Ni, so many have suggested that a lighter element, such as O, S, Si, C, or H, is also present. Indeed, some iron meteorites do contain lighter elements, such as the sulfur contained in the mineral troilite (FeS) pictured in Figure 6.1(A).

Stony iron meteorites are rich in pyroxene or olivine as well as Fe–Ni alloy (Fig. 6.1(B)). Many such meteorites probably represent the core–mantle boundaries of asteroid-sized bodies. As such, they may provide valuable information about the core–mantle boundary of Earth. The stony irons and irons were originally at very high temperatures and reflect differentiation (e.g. core segregation, mantle formation) in the interiors of their parent bodies.

The third major subdivision, the *stony meteorites* (or stones), have received more scientific scrutiny than the other types of meteorites. There are two broad varieties. *Chondrites* contain mineral clusters called *chondrules* (Figs. 6.1(C), 6.2), whereas *achondrites* lack chondrules. Chondrules are composed mostly of high-temperature minerals like pyroxene and olivine. They are often spherical, ranging in diameter from ~100 μm to several mm. Many chondrules are very fine grained or even glassy. This, together with their spherical shapes, suggests that they formed by the rapid cooling of liquid droplets. Such chondrules may have condensed directly from the solar nebula, or represent rapidly quenched droplets of melt generated by impacts during planet formation. Irregularly shaped, more coarsely crystalline chondrules are also found, and these presumably cooled more slowly than the fine-grained or glassy varieties.

Chrondrites are divided into three main classes: *carbonaceous chondrites*, *ordinary chondrites*, and *enstatite chondrites*. Undoubtedly the most widely studied of these are the carbonaceous chondrites. They are rich in water and carbon, and contain complex organic compounds including amino acids. They are very "primitive," in the sense that they have not been heated and metamorphosed to more than ~200 °C. If they had been heated to higher temperatures, much of their volatile content would have been driven off. Ordinary chondrites, by far the most commonly found meteorites, are in many ways similar to carbonaceous chondrites, but they have been heated to higher temperatures (<~400 °C) and are often brecciated. Enstatite chondrites are the least primitive; they are mostly impact breccias that were heated to temperatures of ~600 °C to, in some cases, nearly 900 °C.

Because the carbonaceous chondrites have not been heated significantly and are undifferentiated, they presumably are much closer to the composition of the primordial solar nebula than irons or stony irons. In fact, carbonaceous chondrites are chemically very similar to the spectroscopically determined composition of the Sun's visible surface. A major difference, however, is that the chondrites are much poorer in the highly volatile elements H, He, O, C, and N. It turns out that the composition of Earth, exclusive of the highly volatile elements, is also comparable to the solar composition and, hence, to carbonaceous chondrites. Therefore, it was proposed that the bulk composition of Earth was the same as that of carbonaceous chondrites.

It has been found, however, that while carbonaceous chondrites are a useful general model for bulk Earth composition, they are not exact analogs for Earth's chemical makeup (cf. Ringwood, 1975; McDonough and Sun, 1995). For example, the bulk Earth is depleted in some of the moderately volatile elements, like Na, K, Cs, and Rb, relative to carbonaceous chondrites. Furthermore, the oxygen isotopic compositions of Earth and the Moon differ from carbonaceous chondrites (Clayton, 1977). These and other differences have led many to suggest that the carbonaceous chondrites formed in a separate, chemically and isotopically distinct part of the solar nebula.

Achondrites are differentiated meteorites that lack chondrules and, unlike the other meteorite varieties, resemble some igneous rock types found on Earth or the Moon. Some, for example, are very similar to olivine-rich basalts, whereas others have ultramafic compositions and resemble

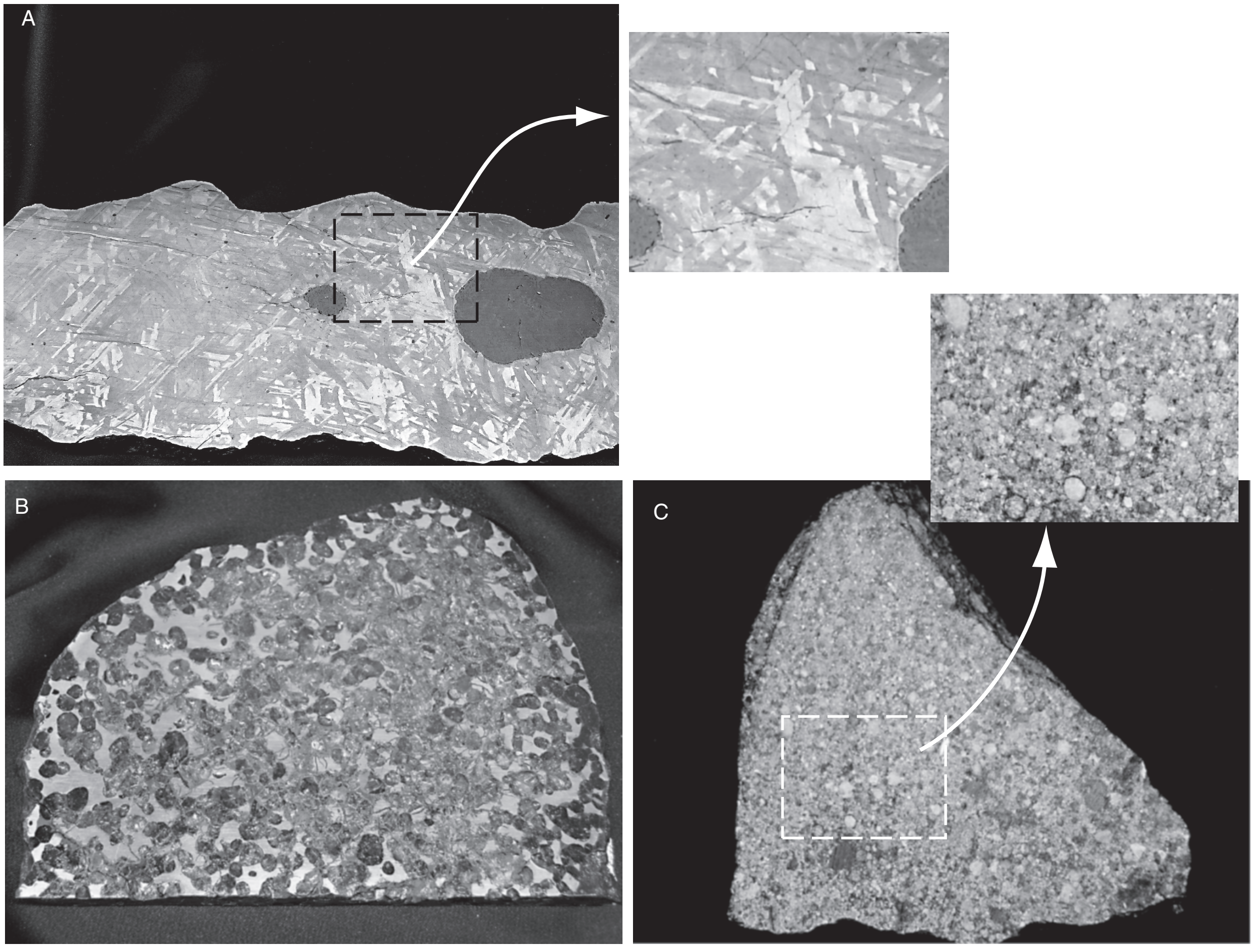


Fig. 6.1 Examples of meteorites. (Photos ©2007, J. J. Ague, courtesy Yale Peabody Museum.) (**A**) Cut slab of iron meteorite that has been etched to highlight the Widmanstätten pattern. The crystals are bent and somewhat irregularly shaped, attesting to impact-related deformation of the parent asteroid (e.g. inset). Note rounded inclusions of troilite (FeS). Meteorite found in the Sacramento Mountains, New Mexico, in about 1890. Field of view is about 20.5 cm. Specimen P274. (**B**) Cut slab of stony iron meteorite containing olivine crystals (dark) set in Fe–Ni alloy; Brenham, Kansas; found in 1882. Slab is about 17 cm wide. Specimen P599. (**C**) LL3 stone meteorite (ordinary chondrite) that fell in Parnallee, India, in 1857. Note the abundant chondrules (inset). LL stands for low iron and low total metal content. The number after the LL represents the degree of alteration of the chondrules, with 3 being the smallest and 7 the largest. Sample is about 8.5 cm wide. Specimen P191.

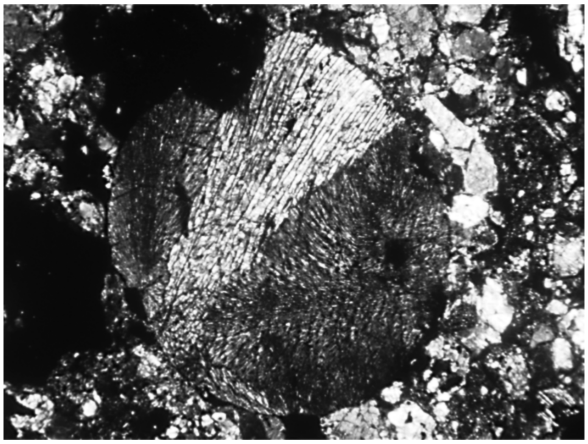


Fig. 6.2 Photomicrograph of an orthopyroxene (bronzite) chondrule surrounded mostly by olivine and Fe–Ni alloy. Note radiating pattern of orthopyroxene crystals. H4 chondrite, Monroe, North Carolina; fell in 1849. The H indicates high Fe–Ni content; the 4 a small to moderate degree of chondrule alteration. Field of view = 2.1 mm. (Photo ©2007, J. J. Ague, courtesy Yale Peabody Museum. Thin section SP11.)

fine-grained peridotites. Brecciation as a result of impacts is common. Many achondrites were derived from asteroids, but a few rare examples came from the Moon or Mars. Meteorite impacts dislodged rock material from the surfaces of these bodies that later was captured in Earth's gravity field. We know some of these meteorites came from the Moon because we have lunar samples returned by the Apollo missions for comparison. The Martian meteorites contain gas that has isotopically heavy nitrogen, just like the Martian atmosphere measured by the Viking landers.

6.1.2 Compositions of magmas and rocks

Elements are conveniently divided into three groups on the basis of their abundance (expressed as weight percent of the oxide). A *major element* typically constitutes more than about 1.0% of a rock, whereas a *minor element* constitutes from about 1.0% to 0.1%, and a *trace element* is less than about 0.1%. Although the divisions are arbitrary, elements in these three divisions play different roles in a rock. The major elements play an essential role, determining such important properties as the phase relations, melting points, densities, and viscosities. These properties are critical in determining whether a magma will form, or whether it will have properties that will allow it to ascend in the Earth. Minor elements commonly form accessory minerals in a rock, such as apatite (P) and zircon (Zr), but their abundance is too low to affect the phase relations significantly. Trace elements have such low concentrations that they simply substitute for major elements in the common minerals but do not affect the phase relations. Although the minor and trace elements may impart characteristic chemical signatures to rocks that may be of considerable practical importance in distinguishing or correlating rock types, the major elements are the important ones in classification.

Most crustal and mantle rocks consist of seven major oxides, SiO_2, Al_2O_3, FeO (Fe_2O_3), MgO, CaO, Na_2O, and K_2O, the relative abundances of which vary considerably. In addition to the crust–mantle differentiation, magmatic processes further broaden the range of compositions of igneous rocks. Certain compositions are more common than others, but a complete gradation exists between the compositional extremes of igneous rocks. Even where physical chemical laws prevent certain compositions of magma from existing, rocks with these compositions can still form through mixing of magma of another composition with appropriate crystals. Indeed, accumulations of crystals from magma can produce igneous rocks that have compositions very different from any liquid that may have existed in the Earth.

The seven major oxide components of magmas crystallize to form the *rock-forming minerals*, which include clinopyroxene, orthopyroxene, olivine, garnet (only at high pressure), amphibole, mica, quartz, plagioclase, alkali feldspar, feldspathoid, and spinel (commonly magnetite). Most of the minerals belong to solid solution series, and thus they can adjust their compositions to fit a wide range of bulk rock compositions. As a result, most rocks are composed of no more than four of the rock-forming minerals, and some contain fewer.

Classifications of igneous rocks have been based on the seven major oxides and the eleven essential rock-forming minerals. Both the chemical and mineralogical classifications have their advantages and disadvantages. Because most rocks contain no more than four rock-forming minerals, the seven major chemical constituents provide for greater differentiation between rock types. But a chemical classification is of little use in the field, because it requires analytical data. Minerals, on the other hand, are readily identifiable in the field, if sufficiently coarse-grained. They may also provide valuable information about the environment in which a rock forms; this usually is not provided by a chemical classification. Coesite and quartz, for example, are chemically identical, but coesite forms only under extremely high pressures. A mineralogical classification, however, is of little use if the minerals are too fine-grained to be identifiable, or the rock is glassy. The IUGS classification therefore uses both the mineralogical and chemical classifications, but emphasizes the mineralogical one wherever possible.

6.2 MODE AND NORM

The mineralogical composition of a rock is known as its *mode*. Mineral abundances are normally expressed in terms of volume percent, for this is what the eye actually perceives. Accurate modes are obtained by *point counting* thin sections of rock under the petrographic microscope. This method involves identifying the mineral beneath the crosshairs of the microscope each time the sample is advanced a metered amount by a mechanical stage. The metered increments essentially create a grid of points on the sample, and the fraction of the total number of points occupied by a particular mineral is proportional to its volume percent. In the case where individual minerals have a significantly different color, image analysis programs, such as NIH *ImageJ* (http://rsb.info.nih.gov/ij), can be used to determine the mode. This is particularly useful for determining the relative abundance of dark (mafic) and light (felsic) minerals. This ratio is commonly referred to as the *color index* (see Problem 6.1).

The *norm*, by contrast, is a calculated mineralogical composition based on the conversion of a whole-rock chemical analysis into various common minerals. This calculation can be done following a prescribed set of rules, such as those of the CIPW norm (see below), or by making use of the MELTS computer program (Ghiorso and Sack, 1995; Asimow and Ghiorso, 1998; Ghiorso *et al.*, 2002), which uses thermodynamic properties of minerals and melts to calculate stable assemblages of minerals that can crystallize from any given melt composition.

The most widely used norm was devised by the four petrologists Cross (United States Geological Survey), Iddings (University of Chicago), Pirsson (Yale), and Washington (Private Laboratory) (Johannsen, 1931). It is therefore commonly referred to as the *CIPW norm*. Because it takes the weight percentages of the oxides and recasts them as the weight percent of anhydrous end-member formulas of the common rock-forming and accessory minerals, it is also known as a *weight norm*. The normative percentages are not directly comparable to those of the mode, which are in volume percent. The discrepancy is particularly noticeable for dense minerals, such as magnetite and ilmenite. The weight norm, however, can be converted to volume proportions by dividing the weight percentage of each normative mineral by its density and recalculating to percentages. Norms can also be given in molecular proportions; these have the advantage of deviating less from the volume mode. Because all such calculations are now carried out on computers, there is no reason not to convert the normative mineral percentages to volume percent if comparison with a mode is to be made. The IUGS classification is based on volume percentages.

The task of distributing the 13 oxides normally given in a rock analysis among a multitude of common minerals at first seems impossible, or at least arbitrary. For example, where should CaO be allotted? Anorthite, clinopyroxene, and wollastonite all contain CaO, as do the accessory minerals apatite and calcite. Distribution of silica is still more of a problem because most of the common minerals are silicates, but they do

Table 6.2 *CIPW normative minerals*

Mineral	Abbreviation	Formula	Mol. to wt. factor
Quartz	Q	SiO_2	$SiO_2 \times 60.09$
Corundum	C	Al_2O_3	$Al_2O_3 \times 101.96$
Zircon	Z	$ZrO_2 \cdot SiO_2$	$ZrO_2 \times 183.31$
Orthoclase	or	$K_2O \cdot Al_2O_3 \cdot 6SiO_2$	$K_2O \times 556.67$
Albite	ab	$Na_2O \cdot Al_2O_3 \cdot 6SiO_2$	$Na_2O \times 524.46$
Anorthite	an	$CaO \cdot Al_2O_3 \cdot 2SiO_2$	$CaO \times 278.21$
Leucite	lc	$K_2O \cdot Al_2O_3 \cdot 4SiO_2$	$K_2O \times 436.50$
Nepheline	ne	$Na_2O \cdot Al_2O_3 \cdot 2SiO_2$	$Na_2O \times 284.11$
Kaliophilite	kp	$K_2O \cdot Al_2O_3 \cdot 2SiO_2$	$K_2O \times 316.33$
Acmite	ac	$Na_2O \cdot Fe_2O_3 \cdot 4SiO_2$	$Na_2O \times 462.02$
Na metasilicate	ns	$Na_2O \cdot SiO_2$	$Na_2O \times 122.07$
	wo	$CaO \cdot SiO_2$	$CaO \times 116.17$
Diopside	en	$MgO \cdot SiO_2$	$MgO \times 100.39$
	fs	$FeO \cdot SiO_2$	$FeO \times 131.93$
Wollastonite	wo	$CaO \cdot SiO_2$	$CaO \times 116.17$
Hypersthene	en	$MgO \cdot SiO_2$	$MgO \times 100.39$
	fs	$FeO \cdot SiO_2$	$FeO \times 131.93$
Olivine	fo	$2MgO \cdot SiO_2$	$MgO \times 70.35$
	fa	$2FeO \cdot SiO_2$	$FeO \times 101.89$
Ca orthosilicate	cs	$2CaO \cdot SiO_2$	$CaO \times 86.12$
Magnetite	mt	$FeO \cdot Fe_2O_3$	$FeO \times 231.54$
Chromite	cm	$FeO \cdot Cr_2O_3$	$Cr_2O_3 \times 223.84$
Hematite	hm	Fe_2O_3	$Fe_2O_3 \times 159.69$
Ilmenite	il	$FeO \cdot TiO_2$	$TiO_2 \times 151.75$
Sphene (titanite)	tn	$CaO \cdot TiO_2 \cdot SiO_2$	$TiO_2 \times 196.07$
Perovskite	pf	$CaO \cdot TiO_2$	$TiO_2 \times 135.98$
Rutile	ru	TiO_2	$TiO_2 \times 79.90$
Apatite	ap	$3(3CaO \cdot P_2O_5) \cdot CaF_2$	$P_2O_5 \times 336.21$
Fluorite	fr	CaF_2	$F \times 39.04$
Pyrite	pr	FeS_2	$S \times 59.98$
Calcite	cc	$CaO \cdot CO_2$	$CO_2 \times 100.09$

not all require the same amounts of silica. Of course, no unique solution to this problem would exist if there were no restrictions on the minerals that form. Petrographic experience, however, indicates that nature places severe restrictions on the possible mineral assemblages that can occur. On the basis of these restrictions, then, Cross, Iddings, Pirsson, and Washington devised their scheme for calculating the norm. It is a credit to their skills as petrographers that the norm calculation has stood the test of time and gives results that are similar to those now obtained using thermodynamic principles (Ghiorso and Sack, 1995). Although the calculation need not be memorized, valuable petrologic insight can be gained from a general appreciation of the steps involved, in particular those dealing with the distribution of silica (Problem 6.2 and 6.4).

6.3 CIPW NORM CALCULATION

To begin the CIPW norm calculation, the weight percentages of the oxides in the rock analysis are converted to mole proportions by dividing each oxide weight percent by its molecular weight. Because trace amounts of Mn and Ni substitute for Fe in the common ferromagnesian minerals, the mole proportions of MnO and NiO are added to that of FeO; similarly, the mole proportions of BaO and SrO are added to that of CaO. Using the formulas in Table 6.2, the mole proportions are distributed among the normative minerals according to the rules given below.

This calculation is tedious and the reader will be pleased to know that they can find web sites from which to download the CIPW norm calculation. The rules for carrying out the calculation, however, are given here because they are based on a considerable body of petrographic experience, and the authors believe that they are instructive. Pay particular attention to rules 18 and 19, because they embody the main classification of igneous rocks and involve important physical chemical principles that relate to the origin of rocks. This will become apparent after the discussion of thermodynamics and phase equilibria in Chapters 7 to 10.

The calculations can be carried out conveniently by hand in a simple balance sheet, as shown in Table 6.3. Designing a computer program that takes into account all of the rules and contingency is not difficult but does involve many nested *if statements* (Kelsey, 1965). Many CIPW norm calculation programs can be downloaded from the Web, but most of these give erroneous results when dealing with extremely silica-undersaturated rocks. The version by Jake Lowenstern at the USGS Volcano Hazards Program site gives reliable results (http://volcanoes.usgs.gov/staff/jlowenstern/other/software_jbl.html).

Table 6.3 is a simple balance sheet in which the first column gives the chemical analysis of the rock in weight percent; this is followed by the molecular weights of the oxides; the third column gives the mole proportions. The remaining columns, one for each normative mineral formed, show the amount of an oxide that remains after the necessary amount has been allocated to the particular mineral. Eventually, each of the oxides is reduced to zero at the right side of the table after all of the oxides have been allotted to the appropriate minerals. The weight percentages of the normative minerals are then calculated by multiplying the appropriate oxide by the factor given in Table 6.2. As a check on the calculation, the sum of the normative minerals should be the same as the sum of the weight percentages of the oxides used in the original analysis (H_2O is ignored).

In the following rules, the name of an oxide refers to its mole proportion. Thus, a rule stating that the SiO_2 is reduced by $0.5 \times MgO$ indicates that the mole proportion of SiO_2 is reduced by an amount equal to one-half the mole proportion of MgO.

1. P_2O_5 is allotted to apatite, and CaO is reduced by $3.33 \times P_2O_5$.
2. S is allotted to pyrite, and FeO is reduced by $0.5 \times S$.
3. Cr_2O_3 is allotted to chromite, and FeO is reduced by Cr_2O_3.
4. TiO_2 is allotted to ilmenite, and FeO is reduced by TiO_2. If TiO_2 exceeds FeO, the excess is allotted to provisional sphene (tn′), and CaO and SiO_2 are both reduced by an amount equal to the excess of TiO_2; this step is carried out only if CaO remains after formation of anorthite (step 10). If TiO_2 still remains, it is calculated as rutile.

Table 6.3 *Example of CIPW norm calculation (see text for explanation)*

				If mole proportion is affected by the formation of the normative mineral, the new value is entered here											
										di[a]			hy		
Oxide	Wt%	Mol. wt. oxide	Mole prop.	ap	il	or	ab	an	mt	wo	en	fs	en	fs	Q
SiO_2	53.8	60.09	0.8953	–	–	0.7999	0.5095	0.3655	–		0.2463			0.1133	0.0000
TiO_2	2.0	79.90	0.0250	–	0.0000										
Al_2O_3	13.9	101.96	0.1363	–	–	0.1204	0.0720	0.0000							
Fe_2O_3	2.6	159.69	0.0163	–	–	–	–	–	0.0000						
FeO	9.3	71.85	0.1294	–	0.1072	–	–	–	0.0909	–	–	0.0628	–	0.0000	
MnO	0.2	70.94	0.0028												
MgO	4.1	40.31	0.1017	–	–	–	–	–	–	–	0.0702	–	0.0000		
CaO	7.9	56.08	0.1409	0.1316	–	–	–	0.0596	–	0.0000					
Na_2O	3.0	61.98	0.0484	–	–	–	0.0000								
K_2O	1.5	94.20	0.0159	–	–	0.0000									
P_2O_5	0.4	141.95	0.0028	0.0000											
Total	98.7														
Mole proportion of normative mineral				0.0028	0.0250	0.0159	0.0484	0.0720	0.0163	0.0596	0.0315	0.0281	0.0702	0.0628	0.1133
Mole to weight conversion factor				336.21	151.75	556.67	524.46	278.21	231.54	116.17	100.39	131.93	100.39	131.93	60.09
Wt% of normative minerals				0.94	3.79	8.85	25.38	20.03	3.77	6.92	3.16	3.71	7.05	8.28	6.81
Sum of normative minerals				98.7											

[a] Calculation of diopside:
In diopside, wo = (en + fs) = 0.0596
but en/fs = 0.1017/0.0909
therefore en = 0.0315
and fs = 0.0281

5. F is allotted to fluorite, and CaO is reduced by 0.5 × F.
6. CO_2 is allotted to calcite, and CaO is reduced by CO_2.
7. ZrO_2 is allotted to zircon, and SiO_2 is reduced by ZrO_2.
8. K_2O is allotted to provisional orthoclase (or′), and Al_2O_3 is reduced by K_2O, and SiO_2 is reduced by 6 × K_2O.
9. Al_2O_3 remaining from step 8 is combined with an equal amount of Na_2O to form provisional albite (ab′), and silica is decreased by six times this amount. If there is insufficient Al_2O_3, proceed to step 11.
10. Al_2O_3 remaining from step 9 is combined with an equal amount of CaO to form provisional anorthite (an′), and silica is decreased by twice this amount. If Al_2O_3 exceeds CaO, it is calculated as corundum.
11. If Na_2O exceeds Al_2O_3 in step 9, an amount of Fe_2O_3 equal to the excess is allotted to acmite, and silica is decreased by four times this amount.
12. If Na_2O still remains after step 11, the remaining Na_2O forms sodium metasilicate, and silica is reduced by the amount of the remaining Na_2O (this is extremely rare).
13. All remaining Fe_2O_3 is allotted to magnetite, and the FeO is decreased by Fe_2O_3. If Fe_2O_3 exceeds FeO, the excess is calculated as hematite.
14. All remaining MgO and FeO forms pyroxenes and olivines. At this point, MgO and FeO are added together, but their proportions are maintained in calculating the amounts of the Mg and Fe end-member components of the pyroxenes and olivines.
15. The CaO remaining from step 10 forms provisional diopside (di′), which decreases MgO + FeO by an amount equal to CaO, and silica by twice this amount.
16. If CaO exceeds MgO + FeO, the excess forms provisional wollastonite (wo′), and silica is decreased by the excess CaO.
17. If the MgO + FeO in step 15 exceeds the CaO, the excess forms provisional hypersthene, and silica is decreased by the excess MgO + FeO.
18. If SiO_2 is still positive, remaining SiO_2 is calculated as quartz.
19. If SiO_2 is negative, the rock contains insufficient silica for the provisionally formed silicate minerals. Some of these minerals must therefore be converted to ones containing less silica, until the silica deficiency is eliminated. The order in which this is done is as follows: first, hypersthene is converted to olivine, then sphene to perovskite, albite to nepheline, orthoclase to leucite, wollastonite and diopside to calcium orthosilicate and olivine, and finally, leucite to kaliophilite. Let the deficiency in silica be D; the amounts of the provisional normative minerals are designated with a prime (see Table 6.2 for abbreviations).
20. If $D < hy'/2$, set $ol = D$, and $hy = hy' - 2D$.

    If $D > hy'/2$, all provisional hypersthene is converted to olivine ($ol = hy'$), and the new silica deficiency, D_1, is $D - hy'/2$.
21. If $D < tn'$, set $pf = D_1$ and $tn = tn' - D_1$.

    If $D > tn'$, all provisional sphene is converted to perovskite ($pf = tn'$), and the new silica deficiency, D_2, is D ′.

    If no provisional sphene was made in step 4, simply set $D_2 = D_1$.
22. If $D_2 < 4ab'$, some of the provisional albite is converted to nepheline, such that $ne = D_2/4$ and $ab = ab' - D_2/4$.

    If $D_2 > 4ab'$, all provisional albite is converted to nepheline ($ne = ab'$), and the new silica deficiency, D_3, is $D_2 - 4ab'$.
23. If $D_3 < 2or'$, some provisional orthoclase is converted to leucite, such that $lc = D_3/2$ and $or = or' - D_3/2$.

    If $D_3 > 2or'$, all provisional orthoclase is converted to provisional leucite ($lc' = or'$) and the new silica deficiency, D_4, is $D_3 - 2or'$.
24. If $D_4 < wo'/2$ in provisional wollastonite (not in diopside), some provisional wollastonite is converted to calcium orthosilicate, such that $cs = D_4$, and $wo = wo' - 2\ D_4$.

    If $D_4 > wo'/2$, all provisional wollastonite is converted to calcium orthosilicate ($cs = wo'/2$), and the new silica deficiency, D_5, is $D_4 - wo'/2$.
25. If $D_5 < di'$, some of the diopside is converted to calcium orthosilicate and olivine which are added to that previously formed; set $cs = D_5/2$, $ol = D_5/2$, and $di = di' - D_5$, remembering to add the amounts of cs and ol to those already formed in steps 24 and 20, respectively.

    If $D_5 > di'$, all provisional diopside is converted to calcium orthosilicate and olivine, such that $cs = di'/2$ and $ol = di'/2$ (add to amounts formed in steps 24 and 20), and the new silica deficiency, D_6, is $D_5 - di'$.
26. Finally, if silica is still deficient, leucite is converted to kaliophilite; set $kp = D_6/2$, and $lc = lc' - D_6/2$.

Once the silica deficiency has been eliminated (one of the steps between 20 and 26), the norm calculation is completed by multiplying the mole proportion of the first oxide in the formula of each normative mineral formed by the weight-conversion factor given in Table 6.2. Only rare rock types have low enough silica contents to cause the norm calculation to proceed beyond step 22. For most rocks, the calculation is therefore quite short (Problem 6.2).

6.4 MELTS PROGRAM AND NORMATIVE MINERALS

The MELTS program of Ghiorso and Sack (1995) provides another means of calculating the mineral composition of a rock based on the rock's chemical composition. Unlike the CIPW norm, which redistributes the oxides in a whole-rock analysis amongst simple pure anhydrous end-member minerals that are known, from field observations, to crystallize together, the MELTS program uses thermodynamic data for silicate liquids and minerals to calculate actual phase compositions that would stably coexist at any given temperature and pressure. It thus allows you to trace the crystallization of a magma from the liquidus, where the magma is totally liquid, to the solidus, where it is completely solid. It is the solidus mineral assemblage that we are interested in if we are to classify a rock based on its chemical composition. In addition to calculating more realistic mineral compositions, MELTS takes into account the effect pressure has on mineral

Table 6.4 *General descriptive petrographic terms*

Mineralogical	Chemical	Obsolete	Color	Color index
Felsic	= SiO_2-rich	= Acid	= Leucocratic	= 0
Mafic		= Basic		│
Ultramafic	= (Mg,Fe)O-rich	= Ultrabasic	= Melanocratic	= 100

assemblages, and it allows for hydrous minerals. In contrast, CIPW normative minerals are those formed at low pressure and are anhydrous; they therefore most closely resemble volcanic mineral assemblages. We will discuss the justification for the MELTS program in Chapter 10.

The MELTS program can be downloaded from the Web at http://melts.ofm-research.org. It has a user-friendly interface and the instructions are simple to follow. To calculate the mineral composition of a rock, the chemical analysis is entered into the program, and a starting high temperature and final low temperature are chosen, along with the cooling increments for which you wish the calculation to be performed. Since we are only interested at this point in obtaining the final solid mineral assemblage it is important only to choose a final temperature that is low enough to ensure complete crystallization. It should be pointed out that the MELTS program sometimes gets hung up trying to solve equations that eliminate the final few percent of liquid. If this happens, the proportion of minerals just prior to complete crystallization can be taken as a reasonable approximation to the rock's mineral content. The pressure under which the crystallization will occur must also be set. If the rock analysis does not give ferrous and ferric iron separately but only total iron, then it is necessary to set what is known as the *oxygen fugacity*; this can be thought of as the effective partial pressure of oxygen – it tells the magma how oxidizing the environment is. This is an important parameter because it affects the oxidation state of iron, which changes dramatically what minerals will form. Most rocks crystallize with an oxygen fugacity close to that provided by the reaction of fayalite + O_2 to form quartz and magnetite (see Section 11.8 and Eq. (11.42)). The assemblage of quartz + fayalite + magnetite is known as the QFM oxygen buffer and is one of the oxygen fugacites provided in MELTS. The program can then be run and the final mineral assemblage calculated. Problem 6.5 leads you through such a calculation and allows you to compare the mineral assemblage obtained with that calculated using the CIPW norm.

6.5 GENERAL CLASSIFICATION TERMS

It will be recalled from Chapter 4 that bodies of igneous rock are divided into plutonic, hypabyssal, and extrusive, depending on their depth of emplacement. The rocks in these bodies are similarly classified as plutonic, hypabyssal, and extrusive or volcanic. No definite division separates plutonic from hypabyssal rocks, but in general the distinction is based on grain size. In plutonic rocks, minerals are readily identifiable in the hand specimen because grain sizes are either medium or coarse (>1 mm). Hypabyssal and volcanic rocks are fine-grained or even glassy, and mineral identification in hand specimens is difficult. For this reason hypabyssal and volcanic rocks are commonly classified together.

The grain size of igneous rocks is described as *coarse* if the average grain diameter is greater than 5 mm, *medium* if it is between 5 and 1 mm, and *fine* if it is less than 1 mm. In general, these categories correspond respectively to rocks in which minerals can be identified by the unaided eye, with the aid of a hand lens, and with a microscope. Rocks that are so fine-grained that individual crystals are difficult to discern even with a hand lens are referred to as *aphanitic*. Many igneous rocks contain crystals of different size. Such rocks are described as *porphyritic*; the large crystals are termed *phenocrysts* and the finer ones *groundmass*. Igneous rocks exhibit a large number of special textures (see Philpotts, 1989), some of which are discussed separately at various places in the text, especially in Chapter 10.

The general composition of a rock can be indicated by a number of different but equivalent terms (Table 6.4). Rocks rich in quartz, feldspars, or feldspathoids are said to be *felsic*, whereas those rich in ferromagnesian minerals are *mafic* or even *ultramafic* if they are totally devoid of felsic minerals. Felsic rocks are rich in SiO_2, whereas ultramafic rocks are rich in MgO and FeO. Early in the development of petrology, silica in magmas was thought to combine with water to form siliceous acid, whereas MgO and FeO were thought to form bases. This led to the terms *acid*, *basic*, and *ultrabasic*. Although these are obsolete, they are still commonly used. Application of any of these terms in the field is usually based on the simple observation of the color of the hand specimen, and thus it seems more to the point to use terms that indicate this. Light-colored rocks are referred to as *leucocratic*, whereas dark ones are *melanocratic*. A quantitative statement of the color of a rock can be made by using the *color index*, which is simply the volume percentage of the dark ferromagnesian minerals (Problem 6.1). The color index can also be based on the percentage of ferromagnesian minerals in the CIPW norm if an accurate mode is not available.

6.6 IUGS CLASSIFICATION OF PLUTONIC IGNEOUS ROCKS

The classification of plutonic rocks is less controversial than that of volcanic ones, because the constituent minerals are readily identifiable and thus can be used for modal classification. The following is a summary of the classification of plutonic rocks recommended by the IUGS Subcommission on the Systematics of Igneous Rocks (Streckeisen, 1976).

The IUGS classification is based on the modal amounts of the common minerals, which are divided into five groups:

Q quartz
A alkali feldspar, including albite with up to 5 mol% anorthite ($<An_5$)
P plagioclase with composition from An_5 to An_{100} and scapolite (a common alteration product of plagioclase)
F feldspathoids: nepheline, sodalite, analcite, leucite, pseudoleucite, kalsilite (kaliophilite), nosean, hauyne, and cancrinite
M mafic minerals: olivine, pyroxenes, amphiboles, micas, monticellite, melilite, opaque minerals, and accessory minerals, such as zircon, apatite, sphene, epidote, allanite, garnet, and carbonate.

Rocks containing less than 90% mafic minerals ($M < 90$) are classified on the basis of their proportions of Q, A, P, and F; rocks with $M > 90$ are classified on the basis of the proportions of the major mafic minerals. The division at $M = 90$ is arbitrary, but few rocks have compositions near this division. Ultramafic rocks typically have $M \gg 90$, whereas other rocks fall well below this limit.

Because no rocks contain both Q and F, nonultramafic rocks can be classified in terms of three components, either QAP or FAP. The three components can be represented in triangular plots or in the orthogonal plot of Figure 6.3. The horizontal axis in this diagram indicates the proportion of plagioclase in the total feldspar ($P/\{P + A\}$), and the vertical axis indicates the modal percentage of quartz (Q) measured upward and the amount of feldspathoid (F) measured downward from the zero line. The diagram is subdivided into areas marking the compositional extent of the main rock types (see also Table 6.5). Although the divisions are arbitrary, they are positioned so that rocks that most petrologists associate with a particular name occupy the center of a field. Thus, a rock containing equal amounts of alkali feldspar and plagioclase and 10% quartz would be named a *quartz monzonite*; a rock containing 15% nepheline and having feldspar that is only 5% plagioclase would be named a *nepheline syenite*.

Some rocks cannot be uniquely defined in terms of only QAP or FAP, and further diagnostic properties must be used. For example, as plagioclase becomes more abundant to the right in Figure 6.3 it becomes more anorthite rich. But in some rocks it becomes more anorthite rich than in others. This difference serves to distinguish *diorite* from *gabbro*; the average plagioclase composition in diorite is less than An_{50}, whereas in gabbro it is greater than An_{50}. (Zoning of plagioclase crystals may make this distinction difficult to apply.)

Most rocks become more mafic with increase in plagioclase content, as indicated in the graphs at the top and bottom of Figure 6.3. Rocks rich in quartz have the lowest contents of mafic minerals, whereas those containing feldspathoids tend to have the highest. Also, gabbroic rocks are generally more mafic than dioritic ones. For any particular rock type there is a normal range of concentration of mafic minerals, as shown in Figure 6.3. Rocks containing more or less than this can have the prefix *mela*- or *leuco*-, respectively, added to the name. Thus, a gabbro containing a large amount of plagioclase would be described as a leucogabbro. If the amount of plagioclase is exceptionally high (>90), the rock is given the special name *anorthosite*. Further definition of a rock type can be given by specifying the names of the major mafic minerals present, with the most abundant being placed closest to the rock name. Thus, a *biotite-hornblende diorite* contains more hornblende than biotite.

Although gabbroic rocks plot in Figure 6.3, their compositional variation cannot be expressed in terms of the components of this diagram. The essential minerals in gabbroic rocks are plagioclase, orthopyroxene, clinopyroxene, olivine, and hornblende, which can be used to classify them into a number of major types (Fig. 6.4). Normal gabbroic rocks contain between 35% and 65% mafic minerals; if they contain more or less than this they are prefixed by the terms *mela*- or *leuco*-, respectively. In addition to the major components, gabbroic rocks may also contain magnetite, ilmenite, spinel, biotite, and garnet, which can be added as qualifiers to the rock name, again with the most abundant mineral being placed closest to the rock name. The plagioclase in gabbroic rocks has an average composition $>An_{50}$, but more albitic compositions may occur on the rims of zoned crystals. Some anorthosites are an exception to this rule. Although anorthosites in layered gabbroic intrusions contain plagioclase in the bytownite range (Section 15.5), large Precambrian anorthosite massifs may consist entirely of andesine plagioclase (some are composed of labradorite). Nonetheless, massif-type anorthosites have been classified with gabbros rather than with diorites, but many of their associated rocks bear more affinities to the diorite-granodiorite suite of rocks than to the gabbroic suite (Section 15.8).

Ultramafic rocks are composed essentially of olivine, orthopyroxene, clinopyroxene, and hornblende. The names used for the various types of ultramafic rock are shown in Figure 6.5. Small amounts of spinel, garnet, biotite, magnetite, or chromite can be indicated, for example, as *spinel-bearing peridotite* if the amount is < 5% and *spinel peridotite* if >5%. If the amount exceeds 50%, as might happen, for example, with chromite, the rock is referred to as a *peridotitic chromitite*, and if the amount exceeds 95%, it is named a *chromitite*. One unusual type of phlogopite-rich peridotite that contains mantle-derived xenoliths is known as *kimberlite*, because of its occurrence at Kimberley, South Africa, where it forms the host rock for diamonds.

Many igneous rocks, especially the more mafic feldspathoid-bearing ones, contain carbonate that has formed as a late-crystallizing phase or as a late alteration product, but it rarely exceeds a few percent. A group of igneous rocks known collectively as *carbonatites*, however, are composed essentially of carbonates (>50% carbonate). These have such low silica contents that they cannot be classified by the schemes for silicate rocks. Most are composed of mixtures of calcite and dolomite, but iron carbonate may be present as well. Recent eruptions of Oldoinyo Lengai volcano in northern Tanzania have produced a *natrocarbonatite*, composed essentially of sodium carbonate. This is not known from any other igneous body. Its rarity in the geologic record is undoubtedly due to the high solubility of sodium carbonate in water, which drastically reduces its chances of preservation. Carbonatites in which the carbonate fraction is more than 90% calcite are known as *sovite* if coarse-grained and *alvikite* if finer-grained. Carbonatites in which the carbonate fraction is more than 90% dolomite are known as *beforsite*. Intermediate compositions containing from 10% to 50% dolomite are referred to as *dolomite-calcite carbonatite*, and from 50% to 90% dolomite as *calcite-dolomite carbonatites*. The presence of other minerals, such as olivine, monticellite, diopside, phlogopite, perovskite, pyrochlore, and apatite can be placed in front of the name if their abundance is significant or attention is to be drawn to a particular mineral.

6.7 IUGS CLASSIFICATION OF VOLCANIC AND HYPABYSSAL ROCKS

Because the modal composition of many volcanic rocks is difficult, if not impossible, to determine, classifications of volcanic rocks have made use of chemical as well as modal data. This has introduced more variables and complicated the classification. As a result, there is no single generally accepted classification of volcanic rocks as there is of plutonic ones. Nonetheless, the IUGS Subcommission on the Systematics of Igneous Rocks has recommended a classification for adoption (Streckeisen, 1979; Table 6.5).

The IUGS recommends that volcanic rocks be classified on the basis of their modes wherever possible, and that the divisions in the classification should correspond to those in the plutonic rock classification. Where modes cannot be obtained, chemical data can be used to classify the rocks, but here again the divisions in the classification should correspond to those for plutonic rocks. The normative calculation provides a simple means of converting chemical data into mineralogical proportions that can be used for classification purposes. Figure 6.3 includes the recommended names for the volcanic rocks in each one of the fields defined by the plutonic rocks. Each of the main plutonic rocks has a volcanic equivalent, except for those in the ultramafic category and anorthosites. These plutonic rocks are normally formed by the accumulation of crystals from a magma, and they have never existed as a liquid. For this reason, they do not appear as lavas. The rare volcanic rock known as *komatiite*, which occurs almost exclusively in Archean terranes, is, however, the volcanic equivalent of peridotite.

Figure 6.3 allows for the simple naming of volcanic rocks based on their modal or normative positions in the diagram. Rocks falling within field 2 should be referred to as *alkali-feldspar rhyolites* rather than *alkali rhyolites*, because the latter term has been used to indicate the presence of an alkali pyroxene or alkali amphibole. The same applies to fields 6*, 6, and 6′. The term *rhyodacite* can be used for rocks transitional between fields 3 and 4. Volcanic rocks having compositions similar to *tonalite* are rare. They can be described as *dacite*, but the common *dacites* have the composition of granodiorite. The distinction between *basalt* and *andesite* based on plagioclase composition is difficult, because of the problem of determining average plagioclase compositions when crystals are zoned. If an analysis is available, the normative plagioclase composition can be used. Also, the silica content serves to distinguish the two, basalts typically containing less than 52 wt% SiO_2 and andesites more. In general, basalts have color indexes greater than 35, whereas most andesites have lower values. The common andesites of strato volcanoes plot in field 9*, whereas most quartz-bearing basalts fall in fields 10 and 10*. So-called *alkali basalts*, which are associated with many oceanic islands and rift valleys on continents, plot in field 10′. Many volcanic rocks contain glass, which should be mentioned in a rock description. Depending on the percentage of glass present, the rock can be described variously as glass-bearing (0% to 20% glass), glass-rich (20% to 50%), glassy (50% to 80%), or by a special name such as obsidian (for a silica-rich glass) if the glass content is from 80% to 100%.

Most hypabyssal rocks have similar grain sizes to volcanic rocks and are equally difficult to classify. Volcanic rock names are used to describe most of them. Basaltic dikes, if thick enough, may be medium-grained and the rock commonly develops a texture known as *ophitic*, in which laths of plagioclase are embedded in larger crystals of pyroxene or olivine. The synonymous terms *dolerite* and *diabase* are used for such rocks. They plot in fields 10 and 10* of Figure 6.3.

One distinct group of hypabyssal rocks that have their own nomenclature are the *lamprophyres*, which are melanocratic, porphyritic rocks containing phenocrysts of a hydrous mafic mineral, biotite or amphibole, and possibly clinopyroxene and olivine in a fine-grained groundmass. Feldspars, if present, occur only in the groundmass, but many lamprophyres contain no feldspar at all. These have such low silica contents that feldspathoids or melilite are present instead. The classification of lamprophyres is given in Table 6.6.

Because volcanic and hypabyssal rocks form on or near Earth's surface where they can interact with circulating groundwater, hydrothermal alteration commonly results in their feldspars being converted to sericite or saussurite (zoisite or epidote) and the mafic minerals to chlorite, serpentine, or talc. Despite this alteration, the rock should be classified according to what the rock was prior to alteration, if this is determinable from textures or relict grains; the prefix *meta*- can be added to indicate that the rock has undergone change. Thus, a metabasalt might consist essentially of albite, epidote, and chlorite. One group of altered rocks for which special terms are used are those formed on the ocean floor. *Spilites* are originally basaltic rock, *keratophyres* originally intermediate volcanic rocks, and *quartz keratophyres* originally silicic volcanic rocks. They have all undergone sodium metasomatism, with the plagioclase being converted entirely into albite, and the mafic minerals are mostly converted to chlorite. Spilites are commonly pillowed, and the

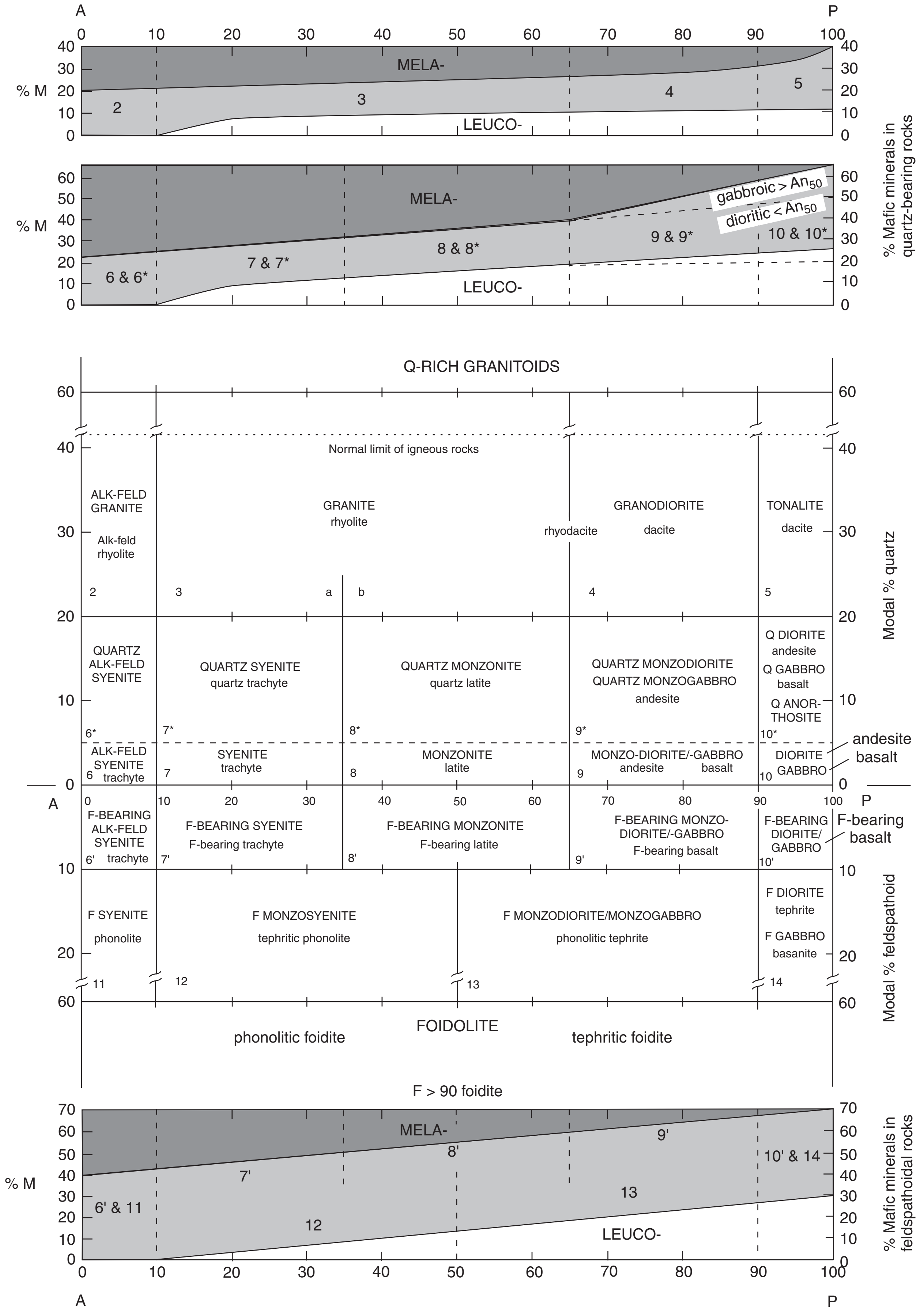

A
P
0
10
20
30
40
50
60
70
80
90
100
% M
MELA-
LEUCO-
2
3
4
5
% Mafic minerals in quartz-bearing rocks
MELA-
gabbroic > An_{50}
dioritic < An_{50}
6 & 6*
7 & 7*
8 & 8*
9 & 9*
10 & 10*
LEUCO-
Q-RICH GRANITOIDS
Normal limit of igneous rocks
ALK-FELD GRANITE
Alk-feld rhyolite
GRANITE
rhyolite
rhyodacite
GRANODIORITE
dacite
TONALITE
dacite
2
3
a
b
4
5
Modal % quartz
QUARTZ ALK-FELD SYENITE
QUARTZ SYENITE
quartz trachyte
QUARTZ MONZONITE
quartz latite
QUARTZ MONZODIORITE
QUARTZ MONZOGABBRO
andesite
Q DIORITE
andesite
Q GABBRO
basalt
Q ANOR-THOSITE
6*
7*
8*
9*
10*
andesite
basalt
ALK-FELD SYENITE
trachyte
SYENITE
trachyte
MONZONITE
latite
MONZO-DIORITE/-GABBRO
andesite
basalt
DIORITE
GABBRO
6
7
8
9
10
F-BEARING ALK-FELD SYENITE
trachyte
F-BEARING SYENITE
F-bearing trachyte
F-BEARING MONZONITE
F-bearing latite
F-BEARING MONZO-DIORITE/-GABBRO
F-bearing basalt
F-BEARING DIORITE/ GABBRO
F-bearing basalt
6'
7'
8'
9'
10'
F SYENITE
phonolite
F MONZOSYENITE
tephritic phonolite
F MONZODIORITE/MONZOGABBRO
phonolitic tephrite
F DIORITE
tephrite
F GABBRO
basanite
11
12
13
14
Modal % feldspathoid
FOIDOLITE
phonolitic foidite
tephritic foidite
F > 90 foidite
% M
MELA-
9'
8'
7'
10' & 14
6' & 11
13
12
LEUCO-
% Mafic minerals in feldspathoidal rocks

Table 6.5 *Common rock names, with IUGS recommended terms in bold print*[a]

Adamellite	3:8*	Plutonic
Alaskite	3:2, 1	Plutonic, no mafics
Alk-feld. granite	3:2	Plutonic, name feldspar
Alk-feld. rhyolite	3:2	Volcanic, name feldspar
Alk-feld. syenite	3:6	Plutonic, name feldspar
Alk-feld. trachyte	3:6	Volcanic, name feldspar
Alkali granite	3:2, 3	Plutonic, alkali amph or pyrox
Alkali rhyolite	3:2, 3	Volcanic, alkali amph or pyrox
Alkali syenite	3:6*, 7*	Plutonic, alkali amph or pyrox
Alkali trachyte	3:6*, 7*	Volcanic, alkali amph or pyrox
Alnöite	Table 6.6	Hypabyssal
Andesite	3:10, 10*, 9, 9*	Volcanic
Ankaramite	3:10	Volcanic, porph. ol-aug-rich
Anorthosite	4:1	Plutonic cumulate
Aplite	3:3	Sugary texture
Basalt	3:10*, 10, 10′	Volcanic
Basanite	3:14	Volcanic
Benmoreite	3:8, 8′	Volcanic
Camptonite	Table 6.6	Hypabyssal
Carbonatite	see text	Plutonic, rare volc. carbonate
Charnockite	3:3b	Plutonic, hypersthene
Comendite	3:2	Volcanic, peralkaline
Cortlandite	4:6	Plutonic, poikilitic texture
Dacite	3:4	Volcanic
Diabase	3:10, 10*	Hypabyssal
Diorite	3:10, 10*	Plagioclase $<An_{50}$
Dolerite	3:10, 10*	Hypabyssal
Dunite	5:1	Plutonic, cumulate
Essexite	3:10′	Plutonic
Fenite	3:7	Aegerine, metasomatic
Foid syenite	3:11	Plutonic, name feldspathoid
Fourchite	Table 6.6	Monchiquite but no olivine
Foyaite	3:11	Plutonic
Gabbro	3:10	Plutonic
Granite	3:3	Plutonic
Granodiorite	3:4	Plutonic
Granophyre	3:3	Hypabyssal, micrographic texture
Harzburgite	5:4	Plutonic, cumulate, or restite
Hornblendite	5:9	Plutonic
Icelandite	3:9*	Volcanic, Fe-rich
Ijolite	3:15	Nepheline + aegerine, plutonic
Jacupirangite	3:15	Titanaugite + nepheline, plutonic
Jotunite	3:9	Plutonic, hypersthene
Keratophyre	3:8*	Metavolcanic, albite
Kersantite	Table 6.6	Hypabyssal
Kimberlite		Ultramafic, phlogopite, hypabyssal
Komatiite	5:3	Volcanic
Lamproite	Table 6.6	Similar to lamprophyre, volcanic, K_2O – MgO-rich
Larvikite	3:6, 6*	Plutonic
Latite	3:8	Volcanic
Lherzolite	5:3	Plutonic, cumulate
Mangerite	3:8	Plutonic, hypersthene, mesoperthite
Melilitite	Table 6.6	Melilite, volcanic
Minette	Table 6.6	Hypabyssal
Monchiquite	Table 6.6	Hypabyssal
Monzonite	3:8	Plutonic
Monzodiorite	3:9	Plutonic
Mugearite	3:9, 9′	Volcanic
Neph. syenite	3:11	Plutonic
Nephelinite	3:15	Volcanic
Nordmarkite	3:7, 7*	Plutonic
Norite	4:3	Plutonic
Pantellerite	3:2, 6*	Volcanic, peralkaline
Peridotite	5:2, 3, 4	Plutonic, cumulate
Phonolite	3:11	Volcanic
Picrite	3:10	Volcanic, porph. olivine-rich
Polzenite	Table 6.6	Hypabyssal
Pulaskite	3:6′	Plutonic
Pyroxenite	5:5, 6, 7	Plutonic, cumulate
Qtz Diorite	3:10*	Plutonic
Qtz Keratophyre	3:3	Metavolcanic, albite
Qtz Latite	3:8*	Volcanic
Qtz Monzonite	3:8*	Plutonic
Qtz Syenite	3:7*	Plutonic
Qtz Trachyte	3:7*	Volcanic
Rapakivi	3:3	Plutonic, plagioclase rimming K-feld
Rhyodacite	3:3–4	Volcanic
Rhyolite	3:3	Volcanic
Sannaite	Table 6.6	Hypabyssal
Shonkinite	3:7′	Plutonic, hypabyssal, mafic
Spessartite	Table 6.6	Hypabyssal
Spilite	3:10	Metabasalt, albite
Syenite	3:7	Plutonic
Syenodiorite	3:9	Plutonic
Tephrite	3:14	Volcanic, plagioclase $< An_{50}$
Teschenite	3:14	Hypabyssal
Theralite	3:14	Plutonic
Tholeiite	3:10*	Volcanic
Tinguaite	3:11	Hypabyssal, analcite
Tonalite	3:5	Plutonic
Trachyte	3:7	Volcanic
Tristanite	3:8′	Volcanic
Troctolite	4:5	Plutonic, plagioclase cumulate
Trondhjemite	3:4	Plutonic, sodic
Ugandite	3:15	Volcanic, mafic, leucite
Vogesite	Table 6.6	Hypabyssal
Websterite	5:6	Plutonic, cumulate
Wehrlite	5:2	Plutonic, cumulate
Wyomingite	3:15	Volcanic, phlogopite, leucite

[a] The first number refers to either Figure 6.3, 6.4, or 6.5; numbers following the colon give the compositional field in which the rock plots in the figures.

tectonic setting of these rocks indicates that they formed on the ocean floor. Their high sodium content may therefore have resulted from interaction with seawater.

More recently, the IUGS Subcommission on the Systematics of Igneous Rocks (Le Bas *et al.*, 1986) has recommended adoption of a simple chemical classification of volcanic rocks based on the two parameters, total alkalis

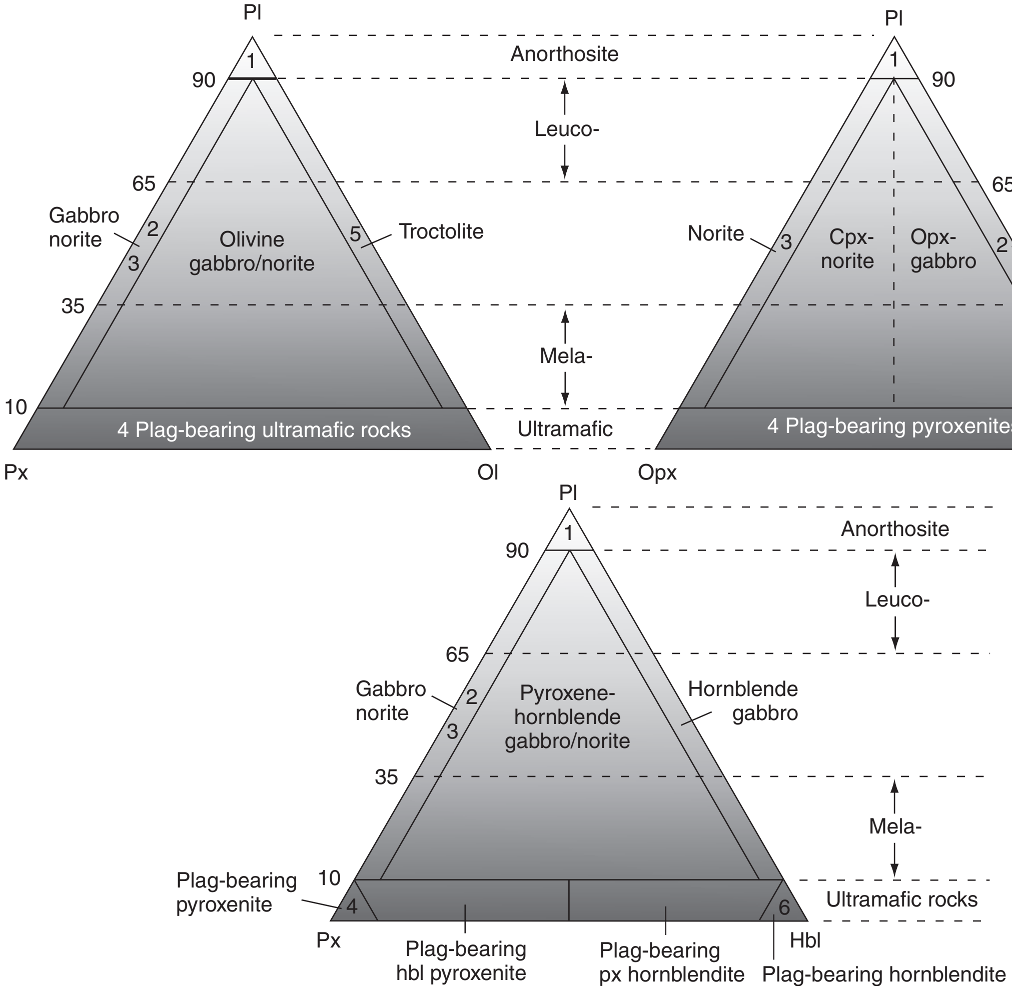


Fig. 6.4 IUGS classification of gabbroic rocks. Shading indicates the approximate color index of the rocks.

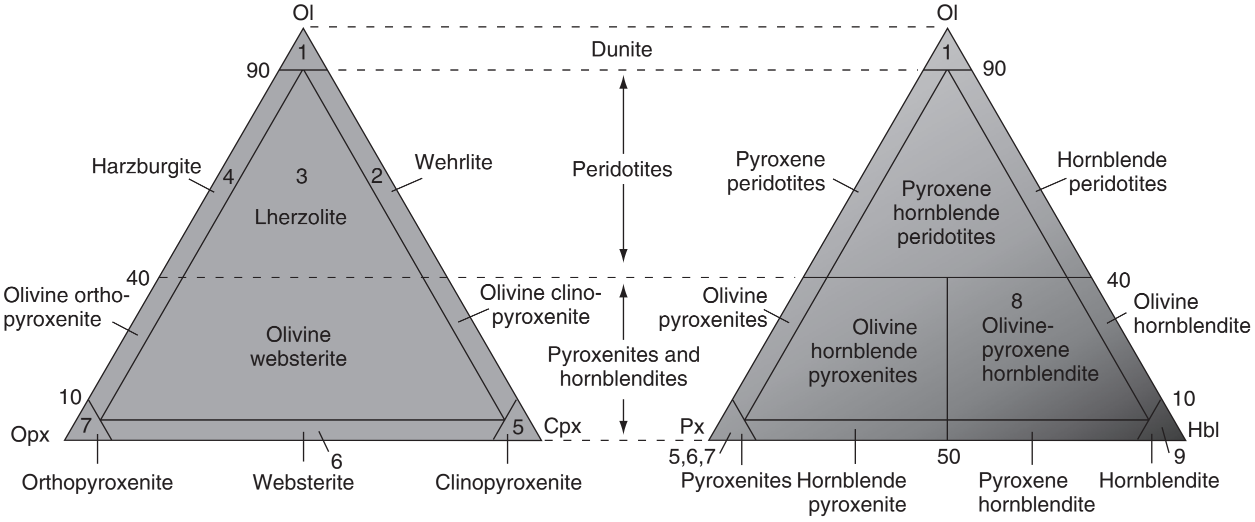


Fig. 6.5 IUGS classification of ultramafic rocks.

and silica content (TAS). The TAS diagram (Fig. 6.6) is divided into 15 fields by a series of straight lines. To plot a rock in this diagram, the analysis is recalculated to 100% on a H_2O- and CO_2-free basis. Rocks containing more than 2% H_2O+ (water driven off by heating above 105 °C) and more than 0.5% CO_2 are considered altered, and their classification by this scheme may be erroneous. Rocks falling in the *trachybasalt* field can be further classified as *hawaiite* if

Table 6.6 *Classification of lamprophyres after Streckeisen (1979)*

Felsic constituent		Predominant mafic mineral			
Feldspar	Foid	biotite diopside augite ±olivine	hornblende diopside augite ±olivine	amphibole (barkevikite kaersutite) titanaugite olivine biotite	melilite biotite ±titanaugite ±olivine ±calcite
or > pl	–	Minette	Vogesite		
pl > or	–	Kersantite	Spessartite		
or > pl	fsp > foid			Sannaite	
pl > or	fsp > foid			Camptonite	
–	foid			Monchiquite	Polzenite
–	–				Alnöite

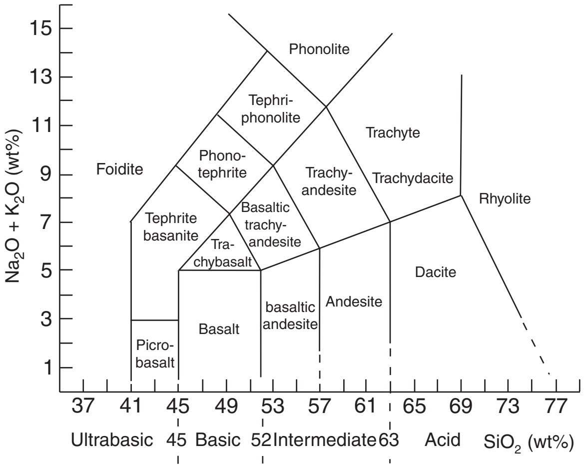


Fig. 6.6 Compositional fields of volcanic rocks in terms of total alkalis and silica. (After Le Bas *et al.*, 1986; published by permission of Oxford University Press.)

$(Na_2O - 2) > K_2O$ and as *potassic trachybasalt* if $(Na_2O - 2) < K_2O$. Similarly, the field of *basaltic trachyandesites* can be divided into *mugearite* (Na) and *shoshonite* (K), and the field of *trachyandesite* into *benmoreite* (Na) and *latite* (K). Classification by this scheme is almost totally consistent with that based on the QAPF diagram (Fig. 6.3).

6.8 THE IRVINE–BARAGAR CLASSIFICATION OF VOLCANIC ROCKS

In recent years the classification proposed by Irvine and Baragar (1971) has gained wide acceptance. It sets up divisions between different rock types based solely on common usage; that is, in practice most geologists associate a particular rock name with a certain compositional range. The scheme also incorporates the well-established fact that volcanic rocks fall into a number of distinct genetic series, which can be distinguished by simple chemical parameters. These series have the added significance that they can be correlated with distinct tectonic environments.

Volcanic rocks are classified by Irvine and Baragar into three main groups (Fig. 6.7), the *subalkaline*, the *alkaline*, and the *peralkaline* (alkali-rich). Most rocks belong to the first two groups, which are each subdivided into two subgroups. Assigning a rock to any one of these groups is based on simple chemical parameters or normative compositions. Before this is done, however, the chemical effects of alteration must be taken into account, if possible. Many volcanic rocks become oxidized, hydrated, or carbonated by hydrothermal activity during burial or during later metamorphism. These chemical changes can seriously affect the normative composition of a rock, which may, in turn, affect its classification. For example, conversion of ferrous iron to ferric during alteration results in smaller amounts of iron silicates being calculated in the norm; this then produces a norm that appears more saturated in silica than was the original rock. This type of alteration, however, can be corrected for, because in many unaltered volcanic rocks there is a strong positive correlation between the TiO_2 and Fe_2O_3 contents. The primary wt% Fe_2O_3 in many volcanic rocks is given approximately by (wt% $TiO_2 + 1.5$). H_2O and CO_2 are subtracted from the analysis and the total recalculated to 100%. Norm calculations are carried out according to the CIPW rules, but Irvine and Baragar chose to recalculate the normative minerals to molecular rather than weight percentages. Thus, instead of multiplying the mole proportions by the weight factors given in Table 6.2, the mole proportions are simply recalculated to 100%. Finally, in expressing feldspar compositions, nepheline is recast as albite. Thus, the normative anorthite content is given by $100 \times An/(An + Ab + 5/3Ne)$. Analyses of typical samples of each of the main rock types in Irvine and Baragar's classification are given in Table 6.7.

Division into the three main groups is based on the alkali content of the rocks. Rocks in which the molecular amounts of $(Na_2O + K_2O) > Al_2O_3$ fall into the peralkaline group. These rocks typically contain aegerine or a sodic amphibole. The alkali content that separates the subalkaline from the alkaline groups varies with the silica content of the rock (Fig. 6.8). The equation for the boundary between these groups is given by

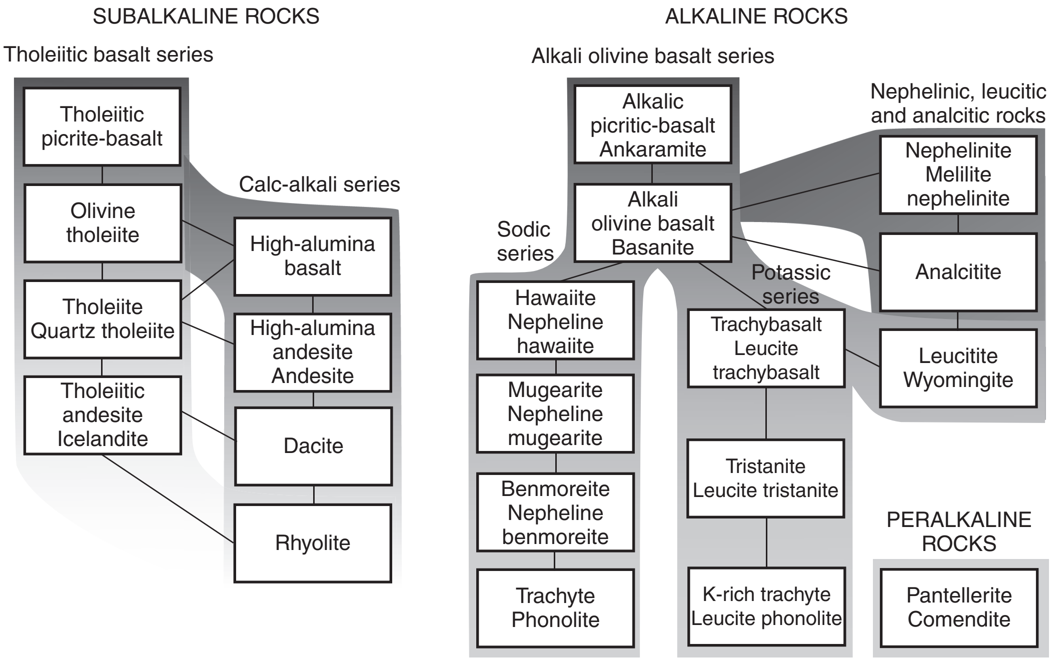


Fig. 6.7 General classification scheme for the common volcanic rocks. Lines joining boxes link commonly associated rocks. The names within the boxes refer to variants of the main rock. (After Irvine and Baragar, 1971; published by permission of *Canadian Journal of Earth Sciences*.)

$$\begin{aligned}\mathrm{SiO_2} = & -3.3539\times10^{-4}\times \mathrm{A}^6 + 1.2030\times10^{-2}\times \mathrm{A}^5\\ & -1.5188\times10^{-1}\times \mathrm{A}^4 + 8.6096\times10^{-1}\times \mathrm{A}^3\\ & -2.1111\times \mathrm{A}^2 + 3.9492\times \mathrm{A} + 39.0\end{aligned}$$

where A = (Na_2O + K_2O). These two groups can also be distinguished in a plot of the normative contents of olivine–nepheline–quartz (Fig. 6.9). To plot a rock in this diagram the normative minerals are recast as follows: $\mathrm{Ne}' = \mathrm{Ne} + \frac{3}{5}\mathrm{Ab}$, $\mathrm{Q}' = \mathrm{Q} + \frac{2}{5}\mathrm{Ab} + \frac{1}{4}\mathrm{Opx}$, and $\mathrm{Ol}' = \mathrm{Ol} + \frac{3}{4}\mathrm{Opx}$. The subalkaline rocks plot on the quartz side of the boundary line, whereas the alkaline ones plot on the nepheline side of it.

The subalkaline rocks are divided into the *calcalkali* and *tholeiitic* series on the basis of their iron contents in the AFM plot (Fig. 6.10), where A = Na_2O + K_2O, F = FeO + 0.8998 × Fe_2O_3, and M = MgO (all in wt%). This plot distinguishes intermediate members of these series very well, but at the mafic and felsic ends there is considerable overlap. Calcalkali basalts and andesites, however, contain 16% to 20% Al_2O_3, which is considerably more than occurs in tholeiitic basalts and andesites, which contain from 12% to 16%. At the extreme felsic end there is no satisfactory way of distinguishing calcalkali and tholeiitic members; thus all granitic rocks are assigned to the calcalkali series. The alkaline rocks are divided into the *alkali olivine basalt* series and the *nephelinitic–leucitic–analcitic* series. Rocks of the latter series typically contain less than 45% SiO_2, have normative color indices greater than 50, and may contain normative leucite.

The naming of rocks within the various subgroups is based on normative plagioclase composition and on normative color index. In the various subalkaline series, the rocks range from *basalt* through *andesite* and *dacite* to *rhyolite* with decreasing normative anorthite content and decreasing normative color index (Fig. 6.11). Two series of rock names are used for the alkaline rocks, depending on whether they are sodic or potassic. This division is made on the basis of the normative feldspar composition (Fig. 6.12). With decreasing normative anorthite content, the sodic series passes from *alkali basalt* through *hawaiite*, *mugearite*, and *benmoreite* to *trachyte* (Fig. 6.13 (A)), whereas the potassic series passes from *alkali basalt* through *trachybasalt* and *tristanite* to *trachyte* (Fig. 6.13(B)). At the mafic end of all of these series, basalts containing more than 25% normative olivine are named *picrites*; these rocks contain abundant phenocrystic olivine. *Ankaramites*, which belong to the alkaline group, contain an abundance of augite phenocrysts, which causes the norm to have more than 20% clinopyroxene. Basalts containing more than 5% normative nepheline are named *basanite* if they contain modal nepheline, or *basanitoid* if nepheline is not visible. Finally, nepheline-bearing trachyte is known as *phonolite*.

Although this classification is descriptive, it has considerable genetic significance. The calcalkali series is characteristic of orogenic belts and gives rise to the volcanic rocks of island arcs. These rocks are clearly related to subduction zones. The tholeiitic rocks are prominently developed in zones of crustal extension where they commonly develop thick sequences of flood basalts. They constitute the major rock type along oceanic ridges (mid-ocean ridge basalt, MORB) and on many large oceanic islands, such as Hawaii. Many alkali olivine basalts and associated rocks occur in areas of continental rifting and in regions overlying deeply subducted plates; they also occur at intraplate hotspots in both oceanic and continental regions.

Table 6.7 *Typical analyses of the rocks listed in Figure 6.7*

	Subalkaline rocks										Peralkaline rocks	
	Tholeiitic basalt series					*Calcalkali series*						
	Tholeiitic picrite	*Ol tholeiite*	*Tholeiite*	*Tholeiitic andesite*	*Icelandite*	*High-Al basalt*	*High-Al andesite*	*Andesite*	*Dacite*	*Rhyolite*	*Pantellerite*	*Comendite*
SiO_2	46.4	49.2	53.8	58.3	61.8	49.1	58.6	60.0	69.7	73.2	69.8	75.2
TiO_2	2.0	2.3	2.0	1.7	1.3	1.5	0.8	1.0	0.4	0.2	0.4	0.1
Al_2O_3	8.5	13.3	13.9	13.8	15.4	17.7	17.4	16.0	15.2	14.0	7.4	12.0
Fe_2O_3	2.5	1.3	2.6	3.4	2.3	2.8	3.2	1.9	1.1	0.6	2.4	0.9
FeO	9.8	9.7	9.3	6.5	5.8	7.2	3.5	6.2	1.9	1.7	6.1	1.2
MnO	0.2	0.2	0.2	0.2	0.2	0.1	0.1	0.2	0.0	0.0	0.3	0.1
MgO	20.8	10.4	4.1	2.3	1.8	6.9	3.3	3.9	0.9	0.4	0.1	0.0
CaO	7.4	10.9	7.9	5.6	5.0	9.9	6.3	5.9	2.7	1.3	0.4	0.3
Na_2O	1.6	2.2	3.0	3.9	4.4	2.9	3.8	3.9	4.5	3.9	6.7	4.8
K_2O	0.3	0.5	1.5	1.9	1.6	0.7	2.0	0.9	3.0	4.1	4.3	4.7
P_2O_5	0.2	0.2	0.4	0.5	0.4	0.3	0.2	0.2	0.1	0.0	0.2	0.1

	Alkaline rocks														
	Alkali olivine basalt series											*Nephelinites etc.*			
	Alkalic picrite	*Ankaramite*	*K-poor alk basalt*	*K-rich alk basalt*	*Trachybasalt*	*Hawaiite*	*Mugearite*	*Tristanite*	*Benmoreite*	*Trachyte*	*Phonolite*	*Nephelinite*	*Analcitite*	*Leucitite*	*Wyomingite*
SiO_2	46.6	44.1	45.4	42.4	46.5	47.9	49.7	55.8	55.6	60.7	60.6	39.7	49.0	46.2	54.1
TiO_2	1.8	2.7	3.0	4.1	3.1	3.4	2.1	1.8	0.9	0.5	0.0	2.8	0.7	1.2	2.3
Al_2O_3	8.2	12.1	14.7	14.1	16.7	15.9	17.0	19.0	16.4	20.5	18.3	11.4	13.0	14.4	9.9
Fe_2O_3	1.2	3.2	4.1	5.8	4.1	4.9	3.4	2.6	3.1	2.3	2.7	5.3	4.9	4.1	3.1
FeO	9.8	9.6	9.2	8.5	7.3	7.6	9.0	3.1	4.9	0.4	1.2	8.2	4.5	4.4	1.5
MnO	0.1	0.2	0.2	0.2	0.2	0.2	0.3	0.1	0.2	0.2	0.2	0.2	0.1	0.0	0.1
MgO	19.6	13.0	7.8	6.7	4.6	4.8	2.8	2.0	1.1	0.2	0.1	12.1	8.3	7.0	7.0
CaO	9.4	11.5	10.5	11.9	9.4	8.0	5.5	4.5	2.9	1.4	0.8	12.8	11.5	13.2	4.7
Na_2O	1.6	1.9	3.0	2.8	3.8	4.2	5.8	5.2	6.1	6.2	8.9	3.8	3.9	1.6	1.4
K_2O	1.2	0.7	1.0	2.0	3.1	1.5	1.9	4.1	3.5	6.7	5.1	1.2	3.0	6.4	11.4
P_2O_5	0.3	0.3	0.4	0.6	0.9	0.7	0.5	0.4	0.7	0.0	0.0	0.9	1.1	0.4	1.8

From Irvine and Baragar (1971).

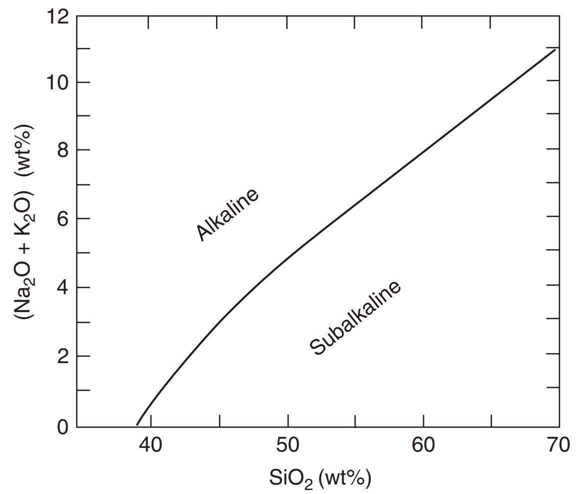


Fig. 6.8 Alkali–silica plot with line separating fields of alkaline and subalkaline rocks (after Irvine and Baragar, 1971).

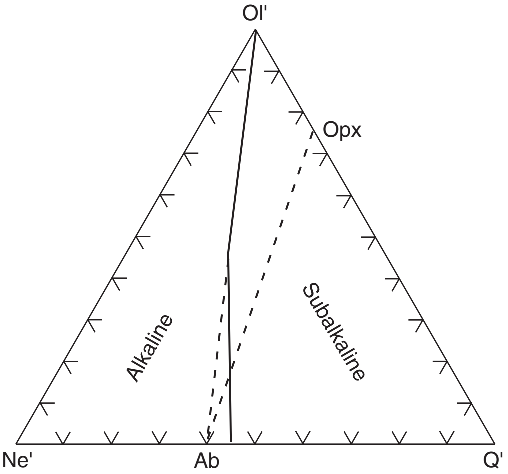


Fig. 6.9 Ol′–Ne′–Q′ projection with line separating fields of alkaline and subalkaline rocks. Plot in percent cation equivalents; see text for explanation (after Irvine and Baragar, 1971).

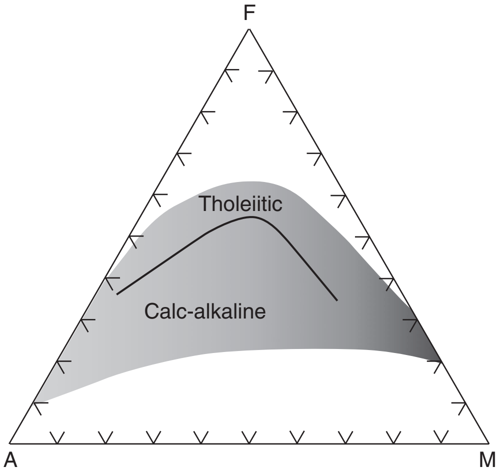


Fig. 6.10 AFM plot showing line separating fields of tholeiitic and calcalkaline rocks as proposed by Irvine and Baragar (1971). A, Na_2O + K O; F, FeO + 0.8998Fe O ; M, MgO; all in weight percent.

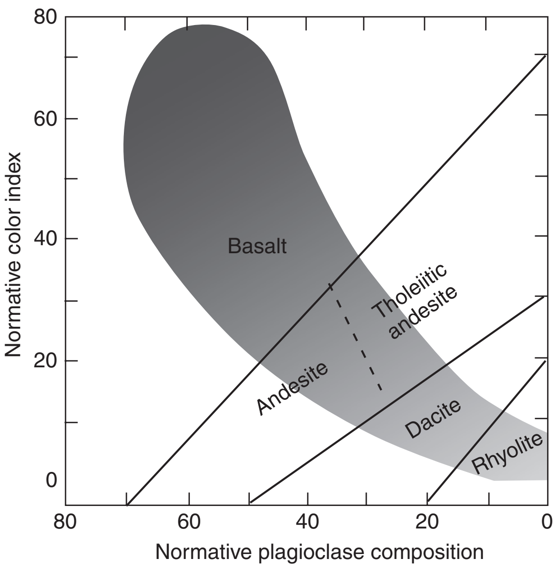


Fig. 6.11 Irvine and Baragar's (1971) subdivision of the subalkaline rocks in a plot of normative color index versus normative plagioclase composition. Plot in percent cation equivalents. Normative color index = ol + opx + cpx + mt + il + hm. Normative plagioclase composition = 100 An/(An + Ab + 5/3Ne); this converts nepheline into albite.

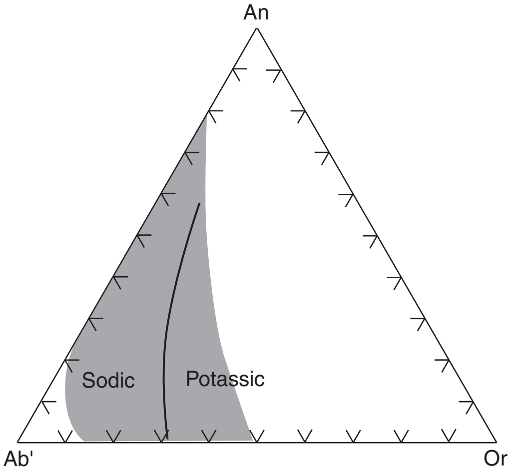


Fig. 6.12 Plot of normative An–Ab′–Or with dividing line separating sodic and potassic alkaline rocks. Ab′ = Ab + 5/3Ne. Plot in percent cation equivalents (after Irvine and Baragar, 1971).

6.9 IGNEOUS ROCK NAMES

Despite the relatively small number of rock names proposed in the IUGS and Irvine and Baragar classifications, the petrologic literature contains hundreds of rock names. Although most of these are no longer used, it is necessary at least to be able to find their definitions in order to read the literature. The list in Table 6.5 includes the most commonly encountered rock names. They have been defined, if possible, by relating them to the IUGS classification with a pair of numbers. The first number, 3, 4, and 5, refers to Figures 6.3, 6.4, and 6.5, and the second number indicates the field in which the rock plots in the particular figure. Thus a *jotunite* (3:9) is a

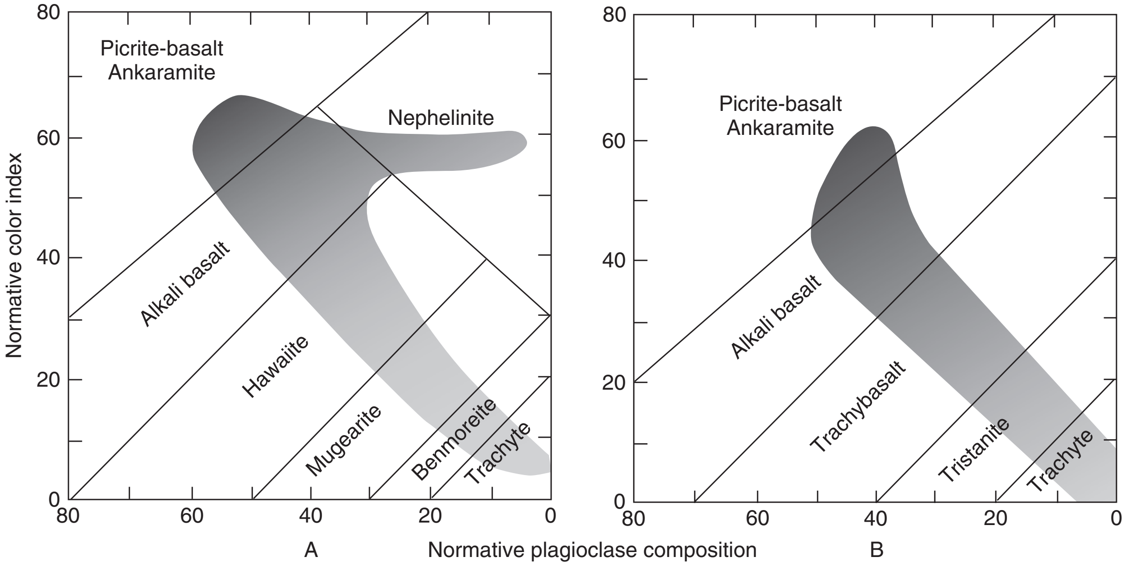


Fig. 6.13 Plots of normative color index versus normative plagioclase composition (defined as in Fig. 6.9) for (**A**) sodic alkaline rocks and (**B**) potassic alkaline rocks (after Irvine and Baragar, 1971).

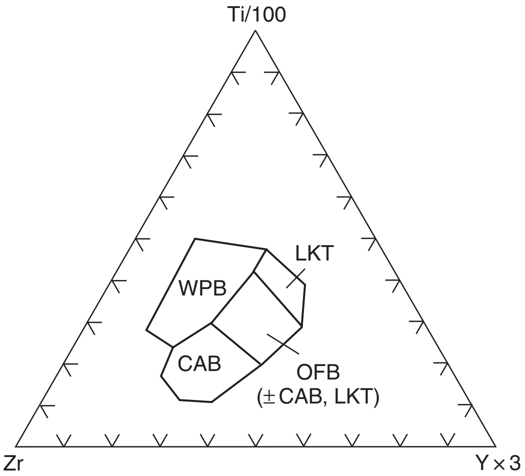


Fig. 6.14 Compositional fields for basaltic rocks from various plate tectonic settings in terms of Ti, Zr, and Y (after Pearce and Cann, 1973). The fields are: within-plate basalts (WPB), calcalkaline basalts (CAB), ocean-floor basalts (OFB), and low-potassium basalts (LKT). Some calcalkaline basalts and low-potassium tholeiites plot in the OFB field.

hypersthene monzodiorite and plots in Fig. 6.3, field 9. Some rock names are based on textural features, in which case these are briefly stated. The object of this tabulation is not to present a catalog of names to be memorized but simply to provide a convenient list to which reference can be made quickly. The important rock names, including for example those in the IUGS classification, have been placed in bold print, and their definitions should be learned.

6.10 CHEMICAL DISCRIMINANTS OF ROCK TYPES

In addition to classifying rocks on the basis of their minerals and general chemical composition, various chemical schemes have been used. Although the IUGS classification has displaced these from common usage, a number of chemical discriminants are used, not to name rocks, but to distinguish possible parental associations and, in the case of ancient rocks, to determine possible plate tectonic settings. For example, Pearce and Cann (1973) used the ratio of the trace elements Ti, Zr, and Y to distinguish basaltic rocks formed in different plate tectonic settings (Fig. 6.14). These particular elements are relatively immobile during metamorphism and weathering, so even when the major element composition has been altered, these trace elements have a better chance of preserving the original values. This is particularly useful in studying ancient volcanic rocks, which rarely survive with pristine compositions.

Granites, according to the IUGS classification, are rocks composed of alkali feldspar with at least 20% quartz. Rocks fitting this category, however, may contain many other minor minerals or have chemical signatures that can allow them to be further subdivided. Chappell and White (1992) found that granites in the Lachlan fold belt of eastern Australia could be divided into two groups based primarily on their alumina content. The *peraluminous* varieties, in which the mole proportions of $Al_2O_3/(K_2O + Na_2O + CaO) > 1$ (i.e. the Al_2O_3 content is in excess of that needed to form feldspar) were believed to have formed from the partial melting of aluminous sedimentary rocks such as shales and thus were referred to as *S-type granites*. These rocks contain, in addition to the essential quartz and feldspar of granite, alumina-rich minerals such as muscovite, garnet, cordierite, and Al_2SiO_5 polymorphs. In the CIPW norm, these aluminous minerals appear as normative corundum. In contrast, metaluminous granites in which the mole proportions of $Al_2O_3/(K_2O + Na_2O + CaO) < 1$ appear to have formed from the melting of mafic igneous rocks and may even involve a mantle-derived component. These granites were designated *I-type*. They have higher Na_2O and CaO than do S-type granites, and consequently amphibole is their common mafic mineral. S- and I-type granites are both found in continental orogenic zones. Some granites, however, occur in anorogenic zones, such as over hot spots or along continental rift valleys and these form a distinct chemical group known as *A-type granites* (Pitcher, 1997). These typically are rich in alkalis, especially K, and they are also rich in high-field-strength cations (high charge to ionic radius) such as Zr, Nb, and Ta and the rare earth elements.

Ocean-floor basalts (MORBs), while having a rather restricted range of composition in terms of their major elements, can have significantly different isotopic signatures of Nd, Sr, Pb, and He, which allows them to be subdivided (Zindler and Hart, 1986). These differences point to the involvement of four or five different components in the generation of magmas in the mantle. Although we are not in a position to discuss these isotopic data at this stage in the book (see Chapters 13 and 23), it is worth pointing out in closing this chapter that classification of rocks using the latest analytical techniques (mass spectrometry) is leading to a deeper understanding of the processes that have led to the formation of our planet.

6.11 PROBLEMS

6.1 The object of this problem is to measure the abundance of dark minerals (color index) in a photomicrograph of the diapir and host rock shown in Figure 4.81(B) using the NIH *ImageJ* program (see Problem 4.5 concerning this program). First, import the image from the text's web site (www.cambridge.org/philpotts) under Problem 6.1, and then convert to an 8-bit grayscale image. In the pull-down menu at the top of the screen, click on **Analyze** and select **Set measurements**; click only the **Area fraction**, and then click OK. Next, in the pull-down menu under **Image**, select **Adjust ▶ Threshold**. The grayscale spectrum shows two prominent peaks; by adjusting the threshold select only the left peak (0–107); this will make only the dark minerals appear red. With the rectangular select tool, outline homogeneous areas within the diapir and in the surrounding host rock and click on **Measure** for each area (do not include the margin of the diapir or any amygdules). Your results table will give you the fraction of dark minerals within each area that you have measured. From these measurements, give average color indexes of the rock forming the diapir and the host rock.

6.2 Calculate the CIPW norms of the following rocks listed in Table 6.7: dacite, tholeiitic picrite, comendite, hawaiite, phonolite, leucitite.

6.3 Using the CIPW norms from Problem 6.2, determine the names that should be given to these rocks according to the IUGS classification set forth in Figures 6.3 and 6.6.

6.4 With insight gained from doing the norm calculations, explain why the following mineral assemblages are never found in rocks:

- **(a)** quartz and nepheline
- **(b)** quartz and forsteritic olivine
- **(c)** albite and leucite
- **(d)** anorthite and acmite
- **(e)** diopside and corundum.

6.5 The object of this problem is to use the MELTS program to calculate what minerals would crystallize from magmas with the dacite and hawaiite compositions in Table 6.7. Open the MELTS program and select the default low-pressure version. First, enter the analysis of the dacite. From the **Intensive Variables** menu select **T, P**... and set a starting temperature of 1200 °C, a stop temperature of 850 °C and a temperature increment of 10 °C. Set the starting and stop pressures at 1 bar. Click **Done** when variables are set. From the **Commands** menu, select **Find Liquidus**. This will give the temperature at which a melt of this composition first begins to crystallize on cooling. Again, from the **Commands** menu select **Execute**. The melt will then crystallize and give the amounts and compositions of the minerals formed. Repeat the process for the Hawaiite analysis, but set a stop temperature of 950 °C. In both rocks the phosphorus is calculated as the mineral whitlockite [$Ca_3(PO_4)_2$]; apatite appears only if the analysis contains Cl or F. Compare your results with the CIPW norms calculated in Problem 6.2, and discuss any differences.

6.5 Recalculate the CIPW norms in Problem 6.2 to molecular norms; then, using the criteria of Irvine and Baragar (1971), determine whether these rocks were properly named in Table 6.7.

7 Introduction to thermodynamics

7.1 INTRODUCTION

Thermodynamics, the study of energy, is one of the most important subjects in all of science. Historically, it evolved from the desire to understand the efficiency of machines, in particular of steam engines. Much of its terminology, therefore, centers around heat and work, especially work associated with expanding gas. Thermodynamics, however, deals with the transfer of other forms of energy, such as that associated with chemical reactions. Although heat and mechanical work done by expanding gas are important in geology, for example in the cooling of a magma or the explosion of a volcano, it is in the study of chemical energies that thermodynamics is of greatest value to petrology. It is particularly useful in the study of processes that take place within the Earth, where they cannot be observed directly. The increased availability in recent years of thermodynamic data for the common minerals and magmas has resulted in a rapid growth in the application of thermodynamics to petrologic problems, and computer programs now use these data to calculate the compositions of minerals crystallizing from magmas and the mineral assemblages that can form in metamorphic rocks under any given temperature and pressure.

The general applicability of thermodynamics stems from the fundamental nature of the principles on which it is based, namely simple observations on the behavior of energy. For example, although energy can be converted from one form to another (kinetic to potential, chemical to thermal, etc.), it can never be destroyed. Furthermore, experience tells us that heat flows from hot to cold bodies, and never the reverse. The first observation, which concerns the conservation of energy, is embodied in the *first law of thermodynamics*, whereas the second one, which deals with the natural direction of processes, leads to the *second law of thermodynamics*. These laws can be expressed in simple mathematical forms, which can then be combined and manipulated to give useful functions from which the equilibrium conditions for a process or reaction can be calculated. In this way, it is possible, for example, to determine melting points of minerals, compositions of minerals crystallizing from magma, temperatures and pressures of metamorphic reactions, relative stabilities of minerals with respect to chemical weathering, and compositions of ore-forming solutions.

Little more than a descriptive treatment of petrology could be given if thermodynamics were to be omitted. However, an entire book would be required to fully develop all thermodynamic relations encountered in the petrologic literature. In this and the following two chapters, only some of the more important fundamental concepts are covered. Standard physical chemistry texts will provide the reader with a more extensive coverage of the topic (e.g. Castellan, 1983, and Denbigh, 1957).

7.2 ENERGY IN THE FORM OF HEAT AND WORK

When discussing the energy of processes, it is important to specify the extent of the material being considered. This is done by using the term *system* to designate that part of space under consideration. A system may have real boundaries, such as the walls of a magma chamber, or imaginary ones, as did the small control volume used in Chapter 2 to derive the rate of flow of magma. The system is chosen to suit the particular problem. Systems are *isolated* if they have no interaction with the surroundings, *closed* if they exchange only heat, and *open* if they exchange both heat and material. Truly isolated systems are difficult to find, but their concept plays an important role in derivations of certain theoretical relations. Many geological systems can be considered closed, as, for example, a small rapidly cooling dike. A large batholith, on the other hand, might exchange considerable amounts of water and other mobile constituents with its surroundings while cooling and would be considered an open system. Many metamorphic reactions involve the loss of a volatile phase from the rock and are therefore open.

Energy can be expressed in the form of either heat or work, ignoring for the moment energy tied up with chemical reaction. *Heat* is the quantity of energy that flows across the boundary of a system in response to a temperature gradient. *Work* is the quantity of energy that crosses the boundary of a system and is converted entirely into mechanical work in the surroundings, such as the lifting of a weight (Fig. 7.1); a geological example would be the explosive removal of the top of Mount St. Helens. By convention, energy put into a system in the form of heat, Q, is positive, whereas that in the form of work, W, is negative. Positive work, then, is done on the surroundings. This convention is inherited from the early days of thermodynamics when there was interest in how much work a machine could do on its surroundings. Some texts (e.g. Kern and Weisbrod, 1967) have changed the sign convention for work, so that, like heat, it is positive when done on the system. Because the designation of sign is purely

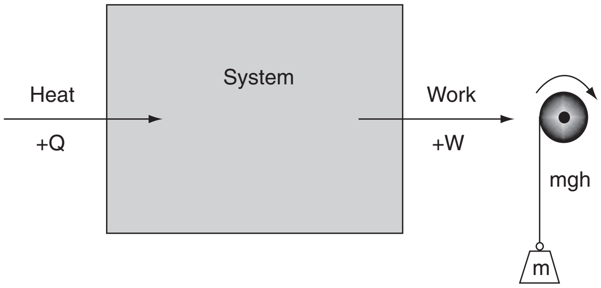


Fig. 7.1 Sign convention for energy transferred into or out of a thermodynamic system in the form of heat or work.

arbitrary, the sign convention does not affect thermodynamic conclusions. Care should be taken, however, in reading thermodynamic texts to ascertain the sign convention used.

The type of work most commonly encountered in petrologic processes is that known as *work of expansion*. For example, when a rock melts at some depth in the Earth, the approximate 10% expansion involved with the phase change results in work being done as the volume expands against the opposing pressure (P_{op}) of the surrounding rock. This work of expansion is given by

$$\begin{aligned} W_{exp} &= (\text{force}) \times (\text{distance}) = (P_{op} \times \text{area}) \times (\text{distance}) \\ &= P_{op}\, \Delta V \end{aligned} \tag{7.1}$$

where ΔV is the volume change. Other types of work include electrical and magnetic, but these are not normally involved in petrologic processes (now would be a good time to do Problem 7.1 and 7.2).

To appreciate work of expansion, consider a gas bubble with volume V_1, pressure P_1, and temperature T in a magma that is suddenly erupted onto the Earth's surface where the pressure is P_2. The bubble expands to V_2 against the opposing pressure P_2, but its temperature is kept constant by the thermal buffering of the surrounding hot magma. This change can be represented by the simple mechanical analog shown in Figure 7.2. The work of expansion done by this bubble on the surrounding magma is given by

$$W_{exp} = P_2(V_2 - V_1)$$

If the gas behaves ideally ($PV = nRT$), this isothermal expansion can be represented by the P versus V plot in Figure 7.2, in which the shaded area represents the amount of work done.

If the magma had stopped at some intermediate depth where the bubble could have expanded against an intermediate pressure P_i and then erupted onto the surface to complete its expansion, the work done in this two-stage decompression would be

$$W_{exp} = P_i(V_i - V_1) + P_2(V_2 - V_i)$$

This amount of work (Fig. 7.3 (A)) is greater than that done by the single-stage expansion. If the bubble had expanded in three stages, the work would have been still greater (Fig. 7.3 (B)). Clearly, the maximum amount of work that could be obtained from this expansion results from an infinite number of infinitesimal steps (Fig. 7.3(C)), in which case the work done would be

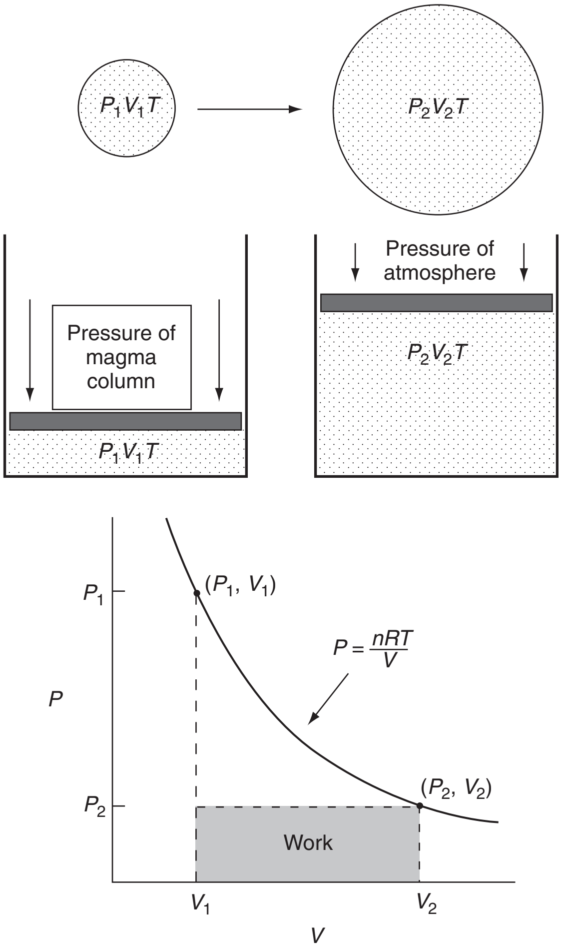


Fig. 7.2 Mechanical analog of a gas bubble expanding isothermally against a pressure of 1 atm (10^5 Pa), and a graphical representation of the amount of work done during expansion ($P\Delta V$) assuming that the gas behaves ideally ($PV = nRT$).

$$W_{exp}^{max} = \int_1^2 P_{op} \mathrm{d}V$$

where P_{op} is the opposing pressure at any stage of expansion. If the gas is taken to be ideal, P_{op} can be replaced by nRT/V, in which case

$$W_{exp}^{max} = \int_1^2 \frac{nRT}{V} \mathrm{d}V = nRT \int_1^2 \frac{\mathrm{d}V}{V} = nRT \ln \frac{V_2}{V_1} \tag{7.2}$$

Supposing that the gas bubble were now to be compressed back to its original state, work would have to be done on the system. This could be done in a single stage of compression by suddenly increasing the pressure to P_1, in which case the amount of work done (Fig. 7.3(D)) would be given by

$$W_{comp} = P_1(V_1 - V_2)$$

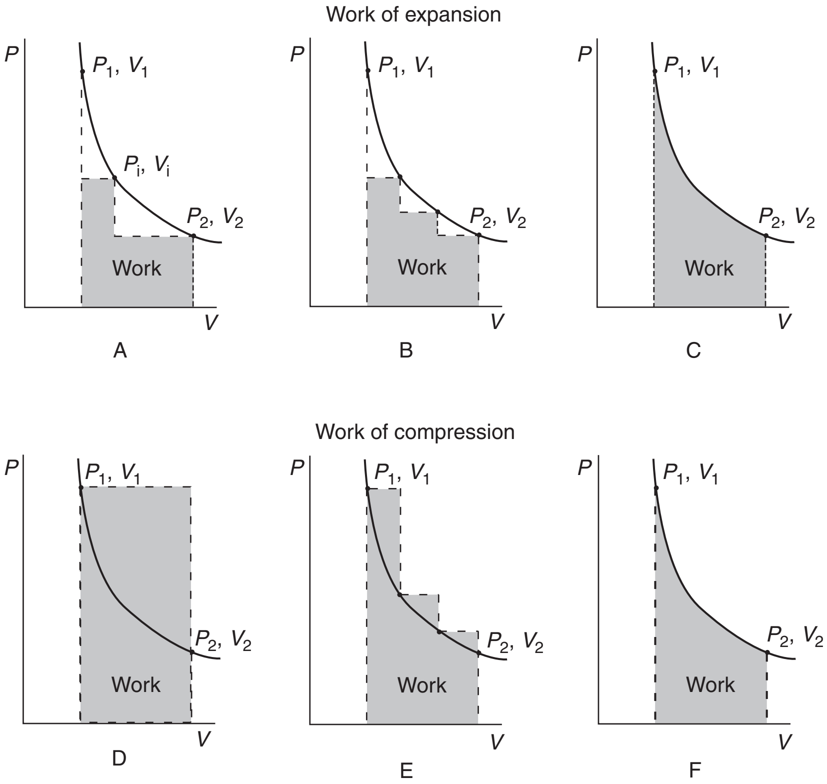


Fig. 7.3 Work of expansion (shaded area) produced when decompression of an ideal gas takes place in (**A**) two stages, (**B**) three stages, (**C**) an infinite number of infinitesimal steps, and the amount of work necessary to return this gas to its original state by isothermally compressing it in (**D**) one stage, (**E**) three stages, and (**F**) an infinite number of infinitesimal steps.

If the bubble were compressed in a series of stages (Fig. 7.3 (E)) the amount of work required would obviously be less, and the minimum amount of work (Fig. 7.3(F)) would be done when the pressure at each stage was increased infinitesimally, giving

$$W_{\text{comp}}^{\min} = \int_2^1 P_{\text{op}}\,\mathrm{d}V = nRT\int_2^1 \frac{\mathrm{d}V}{V} = nRT\,\ln\frac{V_1}{V_2} \\ = -nRT\ln\frac{V_2}{V_1} \tag{7.3}$$

We have followed a bubble through an isothermal cycle of expansion and compression. During the first half of the cycle, the bubble did work on the surroundings, but during the second part, work was done on the bubble to restore it to its initial state. From Figure 7.3, the amount of work produced during expansion in a finite number of steps is clearly less than the amount of work that has to be performed on the system to restore it to its former state. Therefore, processes such as this, carried out in a finite number of steps, always result in work having to be done on the system. On the other hand, if the cycle of decompression and compression could be carried out in an infinite number of infinitesimal steps, the work of expansion (Eq. (7.2)) would equal exactly the work required to compress the gas (Eq. (7.3)); that is, the areas under the PV curves ($\int P\mathrm{d}V$) would be equal. Because the compression part of the cycle carried out in this manner is the exact reverse of the decompression part of the cycle, processes of this type are described as *reversible*, whereas processes carried out in a finite number of steps are *irreversible*. All natural processes are irreversible; reversible processes are unnatural because they would require an infinite amount of time to take place. The concept of a reversible reaction, however, plays an important role in determining conditions of equilibrium, as will be seen later in this chapter.

7.3 FIRST LAW OF THERMODYNAMICS

In the isothermal cyclic process described above, irreversible expansion or compression of the bubble results in turbulence in the gas, which can be equated with heat. For the temperature to remain constant, this heat has to be liberated into the surroundings. Hence, all natural cycles (natural = irreversible), in addition to requiring work to be done on the system, produce heat in the surroundings. Experience tells us that these two quantities of energy are always equal. This fact was first clearly enunciated by the German physicist Mayer in 1842 and eight years later was quantified in Joule's classic experiment on the *mechanical equivalent of heat*. Observations such as these led to the formulation of the *first law of thermodynamics*, which states that *for any cyclical process (reversible or irreversible), the work produced in the surroundings is equal to the heat removed from the surroundings*.

Mathematically the first law can be expressed as

$$\oint \text{đ}W = \oint \text{đ}Q$$

or

$$\oint (\text{đ}Q - \text{đ}W) = 0 \tag{7.4}$$

where $đQ$ and $đW$ are the differentials of heat and work involved in the cyclic process, and the symbol $\oint$ indicates a cyclic integral; that is, the sum of all the $đQ$ and $đW$ around the cycle. Note that $đ$ is used to indicate an inexact differential, because the integrals of $đQ$ or $đW$ have no definite values unless the path is specified (Figure 7.3). Equation (7.4), however, reveals that although $\oint đQ$ and $\oint đW$ by themselves do not have definite values, $\oint (đQ - đW)$ does have a definite value, which for a cyclic process is zero. This means that although the quantities of work and heat taken separately can have various values depending on the path followed by the process, the value of their combination is independent of the path; its value is determined only by the initial and final states of the system. Because of this behavior, $đQ - đW$ is said to be a *state property*; that is, a property which is dependent only on the state of the system and not on the path followed. For convenience, this state property is given the name *internal energy*, *E*, and is defined as

$$dE \equiv đQ - đw \tag{7.5}$$

Note that the internal energy is not an independent quantity that can be equated with heat and work; it is simply the sum of the energy put into the system in the form of heat ($đQ$) and work ($-đW$). Note also that it is defined in differential form; only changes in the internal energy can be measured, not absolute amounts. Integration of Eq. (7.5) results in a constant of integration that cannot be evaluated. This lack of knowledge of the absolute value of internal energy is of no importance because all thermodynamic calculations eventually deal only with differences, and thus the constant of integration disappears. For example, the change in internal energy associated with a reaction going from state 1 to state 2 is given by

$$\int_1^2 dE = \int_1^2 đQ - \int_1^2 đW$$

or

$$E_2 - E_1 = \Delta E = Q - W \tag{7.6}$$

So far, we have used, for illustrative purposes, the purely physical process of the expansion and compression of a gas bubble. Equation (7.5), however, is equally applicable to any process in which there is a change in energy, whether it is physical or chemical. The process could, for example, involve a change of state from solid to liquid or from one mineral polymorph to another, or a chemical reaction between several minerals. Each of these processes will involve the transfer of energy in the form of both heat and work and must obey Eq. (7.5).

Many reactions of interest to petrologists involve work of expansion ($P\Delta V$) at constant pressure – for example, the melting of a rock at a specific depth in the Earth. In such a case, Eq. (7.6) can be expressed as

$$E_2 - E_1 = Q_P - P(V_2 - V_1)$$

where Q_P is the heat involved with the reaction at constant pressure. On rearranging, we obtain

$$Q_P = (E_2 + PV_2) - (E_1 + PV_1) \tag{7.7}$$

which shows that the heat involved with a reaction taking place at constant pressure is the difference between two groups of terms which describe the energy, pressure, and volume of the final and initial states of the system. We have already seen that the internal energy is a state property. But PV is also a state property; for example, in the case of an ideal gas it would be equal to nRT. Consequently, $E + PV$ must also be a state property. It is therefore given a special name, *enthalpy* (*H*), which is defined as

$$H \equiv E + PV \tag{7.8}$$

Equation (7.7) now becomes

$$Q_P = H_2 - H_1 = \Delta H \tag{7.9}$$

The enthalpy change in a reaction is, therefore, the heat withdrawn from the surroundings at constant pressure.

If heat is given out during a reaction it is said to be *exothermic* and ΔH is negative. An *endothermic* reaction is one that takes heat from the surroundings and hence ΔH is positive. For example, if forsterite were to react with quartz at 298 K (25 °C) to form enstatite, 7.4 kJ mol^{-1} of forsterite would be liberated, and hence $\Delta H_{298} = -7.3$ kJ mol^{-1}. In contrast, the reaction from low quartz (α) to high quartz (β) at 848 K (575 °C) is endothermic, and $\Delta H_{848} = +1.2$ kJ mol^{-1}.

7.4 STANDARD HEATS OF FORMATION

Because enthalpy involves the internal energy (Eq. (7.8)), absolute values cannot be known. This, however, is not a problem because thermodynamic calculations deal only with changes in enthalpy. For example, in the reaction above where forsterite reacts with quartz to form enstatite, the absolute values of the enthalpies of the minerals cannot be known, but the enthalpy change ($\Delta H = -7.4$ kJ mol^{-1}) accompanying the reaction can be determined and used to calculate the conditions under which this reaction will occur.

Although absolute values of enthalpy cannot be known, it is convenient to think of substances as having such values. By convention, then, we assign an arbitrary "absolute" value of zero to the enthalpy of each of the elements in their standard stable form at 298.15 K (25 °C) and a pressure of 10^5 Pa (1 bar). This is represented by $H^\circ_{298,\text{element}} \equiv 0$, where the superscript $^\text{o}$ indicates 10^5 Pa pressure. With this arbitrary base level, it is then possible to define the enthalpy of a mineral in terms of the enthalpy change accompanying the formation of that mineral from the elements at 298 K and 10^5 Pa pressure. An enthalpy defined in this way is referred to as the *standard heat* (enthalpy) *of formation* of the mineral, $H^\circ_{f,298}$. For example, the reaction to form quartz would be

$$\text{Si}_{\text{crystal}} + \text{O}_{2\ \text{gas}} \xrightarrow{\Delta H^\circ_{f,298\text{Q}}} \text{SiO}_{2\ \text{crystal}}$$

and the standard heat of formation would be given by

$$\Delta H^\circ_{f,298,\mathrm{Q}} = H^\circ_{298,\mathrm{Q}} - \left(H^\circ_{298,\mathrm{Si}} + H^\circ_{298,\mathrm{O_2}}\right)$$

But the terms in parentheses have values of zero, since they refer to the elements in their stable states at 298 K and 10^5 Pa pressure. The enthalpy of quartz under these conditions is therefore equal to the enthalpy change of the reaction, which is –910.83 kJ mol^{-1}. Note that in this reaction the stable form of silicon under these conditions is a crystalline metal, whereas the stable form of oxygen is the diatomic gas (O_2). Standard heats of formation of minerals are given in Table 7.1.

Although the heat of formation of quartz in the previous reaction was given for a temperature of 298 K, this reaction is not likely to proceed rapidly at this temperature – if it did, pocket calculators and computers would have a very short life expectancy. High temperatures are required before silicon metal will react rapidly with oxygen. Therefore, if we were interested in measuring the heat involved with this reaction, it would be necessary to carry out the experiment at high temperatures. How, then, would we determine, from the high-temperature experiments, the enthalpy change associated with the reaction at 298 K?

To answer this we make use of the fact that enthalpy is a state property, and therefore its value is independent of the path followed by the reaction. For instance, instead of trying to react oxygen with silicon at 298 K, we can heat these materials to a high temperature, 1800 K for example. Quartz forms rapidly at this temperature. Once the reaction is complete, the quartz can be cooled to 298 K. Thus, the temperature of the starting materials and end product will both be 298 K, even though the reaction took place at 1800 K. The enthalpy change between starting materials and end product at 298 K will be the same whether the reaction proceeded directly at 298 K or followed the high-temperature path. These two possible ways of carrying out this reaction can be illustrated as follows:

$$\begin{tikzcd}
\text{(1800 K)} & \mathrm{Si}_{\mathrm{crystal}} & + & \mathrm{O}_{2\ \mathrm{gas}} \arrow[r, "\Delta H^\circ_{f,1800,\mathrm{Q}}"] & \mathrm{SiO}_{2\ \mathrm{Q}} \arrow[d, "-(H^\circ_{1800} - H^\circ_{298})_{\mathrm{Q}}"] \\
\text{(298 K)} & \mathrm{Si}_{\mathrm{crystal}} \arrow[u, "(H^\circ_{1800} - H^\circ_{298})_{\mathrm{Si}}"] & + & \mathrm{O}_{2\ \mathrm{gas}} \arrow[u, "(H^\circ_{1800} - H^\circ_{298})_{\mathrm{O_2}}"] & \mathrm{SiO}_{2\ \mathrm{Q}} \arrow[l, "\Delta H^\circ_{f,298,\mathrm{Q}}"']
\end{tikzcd}$$

Because enthalpy is a state property, its integral around a cyclic process must be zero. Thus, if we were to heat Si and O_2 from 298 K to 1800 K, react them together to form quartz, cool the quartz to 298 K, and then break down the quartz to form Si and O_2 again, the sum of the enthalpy changes of all of the steps around this cycle would be zero; that is,

$$\left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{Si}} + \left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{O_2}} + \Delta H^\circ_{f,1800,\mathrm{Q}} + \left[-\left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{Q}}\right] + \left[-\left(\Delta H^\circ_{f,298,\mathrm{Q}}\right)\right] = 0$$

By rearranging this, we obtain the standard heat of formation of quartz at 298 K,

$$\Delta H^\circ_{f,298,\mathrm{Q}} = \Delta H^\circ_{f,1800,\mathrm{Q}} - \left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{Q}} + \left[\left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{Si}} + \left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{O_2}}\right] \quad (7.10)$$

Most of the enthalpy data presented in Table 7.1 have been collected at temperatures other than 298 K and have had to be corrected to this standard temperature. Moreover, many of the reactions have followed more complicated paths than ones simply requiring heating and cooling. Some have involved dissolving the elements and minerals in acid or high-temperature metallic melts. The heats of solution in these solvents have then been used to calculate the $\Delta H^\circ_{f,298}$. Again, because enthalpy is a state function, the actual path taken in the reaction does not affect the enthalpy change of the overall reaction, which depends only on the initial and final states.

The enthalpy data presented in Table 7.1 are normally used to calculate the enthalpy of formation of a mineral at a particular temperature of interest. For example, we might wish to know the enthalpy of formation of quartz at 1800 K and 10^5 Pa pressure. This can be determined by rearranging Eq. (7.10) and inserting the value for the heat of formation of quartz at 298 K obtained from thermodynamic tables. This gives

$$\Delta H^\circ_{f,1800,\mathrm{Q}} = \Delta H^\circ_{f,298,\mathrm{Q}} + \left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{Q}} - \left[\left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{Si}} + \left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{O_2}}\right] \quad (7.11)$$

High-temperature heats of formation can also be read directly from thermodynamic tables, but care must be exercised in doing this. Enthalpy data are commonly presented in two different ways. One lists values calculated according to Eq. (7.11) (Robie *et al*., 1978). The other also uses Eq. (7.11), but the enthalpy terms for the elements (those in square brackets) are dropped (Helgeson *et al*., 1978); enthalpies calculated this way are referred to as *apparent enthalpies of formation* from the elements. A simple example will illustrate the justification for dropping these terms. To determine the enthalphy change associated with the transformation from low to high quartz at 848 K we need know only the difference in the heats of formation from the elements of these two forms of quartz at this temperature. Values obtained from Eq. (7.11) for the two polymorphs will have identical terms for the elements (those in brackets) because the same elements are involved in both minerals. These terms, therefore, cancel when we take the difference in the enthalpies of formation of the two polymorphs. The same argument applies to other more complicated reactions, because the same elements are present on both sides of the reaction.

To calculate a high-temperature heat of formation from Eq. (7.11), the change in enthalpy associated with changes in temperature (e.g. $\left(H^\circ_{1800} - H^\circ_{298}\right)_{\mathrm{Q}}$ in Eq. (7.11)) must be evaluated. This can be determined from the heat capacity of the mineral at constant pressure (C_{P}), because

$$\int_{H^\circ_{298}}^{H^\circ_T} \mathrm{d}H = \int_{298}^{T} C_{\mathrm{P}}\,\mathrm{d}T \quad (7.12)$$

Table 7.1 *Molar thermodynamic data for common minerals at standard temperature and pressure, i.e. 298.15 K and 10^5 Pa (1 bar)*

Mineral	Formula	Formula wt (kg)	Volume (J bar^{-1})	ΔH_f^o(kJ)	S° (J K^{-1})	ΔG_f^o(kJ)	a (kJ K^{-1})	$b \times 10^5$ (kJ K^{-2})	c (kJ K)	d (kJ K$^{-1/2}$)
Acmite	$NaFeSi_2O_6$	0.23101	6.459	−2586.65	170.6	−2419.31	0.3071	1.6758	−1685.5	−2.1258
Akermanite	$Ca_2MgSi_2O_7$	0.27264	9.254	−3866.36	212.5	−3668.89	0.3854	0.3209	−247.5	−2.8899
Albite	$NaAlSi_3O_8$	0.262224	10.006	−3934.56	210.1	−3711.91	0.452	−1.3364	−1275.9	−3.9536
Almandine	$Fe_3Al_2Si_3O_{12}$	0.49775	11.511	−5263.52	340	−4939.8	0.6773	0	−3772.7	−5.044
Analcite	$NaAlSi_2O_6 \cdot H_2O$	0.220155	9.74	−3309.9	232	−3090.97	0.6435	−1.6067	9302.3	−9.1796
Andalusite	Al_2SiO_5	0.162046	5.153	−2588.8	92.7	−2440.97	0.2773	−0.6588	−1914.1	−2.2656
Andradite	$Ca_3Fe_2Si_3O_{12}$	0.505184	13.204	−5768.13	318	−5424.33	0.6386	0	−4955.1	−3.9892
Annite	$KFe_3(AlSi_3O_{10})(OH)_2$	0.51189	15.432	−5151.67	418	−4796.02	0.8157	−3.4861	19.8	−7.4667
Anorthite	$CaAl_2Si_2O_8$	0.27821	10.079	−4233.48	200	−4007.51	0.3716	1.2615	−4110.2	−2.0384
Anthophyllite	$Mg_7Si_8O_{22}(OH)_2$	0.780872	26.54	−12069.2	536	−11342.22	1.2773	2.5825	−9704.6	−9.0747
Antigorite	$Mg_{48}Si_{34}O_{85}(OH)_{62}$	4.536299	175.48	−71424.31	3591	−70622.39	9.621	−9.1183	−35941.6	−83.0342
Aragonite	$CaCO_3$	0.100089	3.415	−1207.58	89.5	−1128.03	0.1923	−0.3052	1149.7	−2.1183
Brucite	$Mg(OH)_2$	0.058327	2.463	−924.92	64.5	−834.31	0.1584	−0.4076	−1052.3	−1.1713
Calcite	$CaCO_3$	0.100089	3.689	−1207.47	92.5	−1128.81	0.1409	0.5029	−950.7	−0.8584
Carbon dioxide	CO_2	0.04401	2478.920	−393.51	213.7	−394.3	0.0878	−0.2644	706.4	−0.9989
Carbon monoxide	CO	0.02801	2478.920	−110.53	197.67	−137.13	0.0457	−0.0097	662.7	−0.4147
Chloritoid (Fe)	$FeAl_2SiO_5(OH)_2$	0.25191	6.98	−3215.38	155	−2973.74	0.4846	−1.3808	−198.9	−4.7622
Chloritoid (Mg)	$MgAl_2SiO_5(OH)_2$	0.22037	6.875	−3551.42	140	−3313.56	0.4644	−1.2654	−1147.2	−4.341
Chrysotile	$Mg_3Si_2O_5(OH)_4$	0.277134	10.746	−4359.03	221.3	−4030.75	0.6247	−2.077	−1721.8	−5.6194
Clinochlore	$Mg_5Al_2Si_3O_{10}(OH)_4$	0.48777	21.09	−8912.41	430.5	−8263.35	1.1618	1.0133	−7657.3	−9.6909
Clinozoisite	$Ca_2Al_3Si_3O_{12}(OH)$	0.622882	13.63	−6898.15	301	−6502.98	0.567	1.8063	−7034	−2.603
Coesite	SiO_2	0.060085	2.064	−905.47	40.8	−850.89	0.0965	−0.0577	−444.8	−0.7982
Cordierite	$Mg_2Al_3(AlSi_5O_{18})$	0.584969	23.322	−9163.37	407.5	−8653.24	0.8213	4.3339	−8211.2	−5.00
Cordierite hydrous	$Mg_2Al_3(AlSi_5O_{18}).H_2O$	0.602984	23.322	−9446.98	487.3	−8891.08	0.8697	5.1995	−7723.7	−5.2512
Corundum	Al_2O_3	0.101961	2.558	−1675.25	50.9	−1581.72	0.1395	0.589	−2460.6	−0.5892
Cristobalite	SiO_2	0.060085	2.61	−905.99	46.5	−853.12	0.0979	−0.335	−636.2	−0.774
Daphnite	$Fe_5Al_2Si_3O_{10}(OH)_4$	0.64548	21.34	−7134.85	565	−6535.56	1.2374	1.3594	−3743	−11.25
Diamond	C	0.012011	0.342	2.07	2.3	3.13	0.0243	0.6272	−377.4	−0.2734
Diopside	$CaMg(SiO_3)_2$	0.21656	6.619	−3202.76	142.7	−3027.8	0.3145	0.0041	−2745.9	−2.0201
Dolomite	$CaMg(CO_3)_2$	0.184411	6.434	−2324.43	156	−2161.51	0.3589	−0.4905	0	−3.4562
Enstatite	$Mg_2(SiO_3)_2$	0.200792	6.262	−3090.47	132.5	−2915.53	0.3562	−0.299	−596.9	−3.1853
Epidote	$Ca_2FeAl_2Si_3O_{12}(OH)$	0.651747	13.91	−6463.21	328	−6076.41	0.5446	2.4781	−11230	−1.1921
Fayalite	Fe_2SiO_4	0.203778	4.631	−1478.15	151	−1378.98	0.2011	1.733	−1960.6	−0.9
Ferrosilite	$Fe_2(SiO_3)_2$	0.263862	6.592	−2388.63	190.6	−2234.53	0.3987	−0.6579	1290.1	−4.058
Forsterite	Mg_2SiO_4	0.140708	4.366	−2172.2	95.1	−2052.75	0.2333	0.1494	−603.8	−1.8697
Gehlenite	$Ca_2Al_2SiO_7$	0.274205	9.024	−3986.88	202	−3784.82	0.4057	−0.7099	−1188.3	−3.1744
Glaucophane	$Na_2Mg_3Al_2Si_8O_{22}(OH)_2$	0.78355	26.05	−11969.47	535	−11233.99	1.7175	−12.107	7075	−19.272
Graphite	C	0.012011	0.53	0	5.85	0	0.051	−0.4428	488.6	−0.8055
Grossular	$Ca_3Al_2Si_3O_{12}$	0.450454	12.535	−6644.15	255	−6280.94	0.626	0	−5779.2	−4.0029
Hedenbergite	$CaFe(SiO_3)_2$	0.248106	6.795	−2844.16	174.2	−2680.39	0.3402	0.0812	−1047.8	−2.6467
Hematite	Fe_2O_3	0.159692	3.027	−825.71	87.4	−743.73	0.1639	0	−2257.2	−0.6576
Hercynite	$FeAl_2O_4$	0.173809	4.075	−1959.15	107.5	−1843.85	0.2833	−0.5376	609.8	−2.7136
Hydrogen	H_2	0.002016	2478.920	0	130.7	0	0.0233	0.4627	0	0.0763
Ilmenite	$FeTiO_3$	0.151745	3.169	−1231.3	108.9	−1154.63	0.1389	0.5081	−1288.8	−0.4637
Jadeite	$NaAl(SiO_3)_2$	0.20214	6.04	−3027.85	133.5	−2849.1	0.3011	1.0143	−2239.3	−2.0551
Kalsilite	$KalSiO_4$	0.158167	6.04	−2121.92	134	−2005.98	0.242	−0.4482	−895.8	−1.9358
Kaolinite	$Al_2Si_2O_5(OH)_4$	0.258161	9.934	−4122.18	203.7	−3801.72	0.4367	−3.4295	−4055.9	−2.6991
Kyanite	Al_2SiO_5	0.162046	4.414	−2593.11	83.5	−2442.59	0.2794	−0.7124	−2055.6	−2.2894
Laumontite	$CaAl_2Si_4O_{12} \cdot 4H_2O$	0.470441	20.37	−7268.47	457	−6707.45	1.0134	−2.1413	−2235.8	−8.8067
Lawsonite	$CaAl_2Si_2O_7(OH)_2 \cdot H_2O$	0.3142	10.132	−4869.14	230	−4513.04	0.6878	0.1566	375.9	−7.1792
Leucite	$KAlSi_2O_6$	0.218248	8.828	−3029.16	200	−2866.19	0.3698	−1.6332	684.7	−3.6831
Magnesite	$MgCO_3$	0.084321	2.803	−1111.36	65.1	−1027.74	0.1864	−0.3772	0	−1.8862
Magnetite	Fe_3O_4	0.231539	4.452	−1115.51	146.1	−1012.31	0.2625	−0.7204	−1926.2	−1.6557
Margarite	$CaAl_2(Al_2Si_2O_{10})(OH)_2$	0.398187	12.964	−6241.23	267	−5856.99	0.7444	−1.68	−2074.4	−6.7832
Merwinite	$Ca_3Mg(SiO_4)_2$	0.328719	9.847	−4546.42	253.1	−4317.73	0.4175	0.8117	−2923	−2.3203
Methane	CH_4	0.016043	2478.920	−74.81	186.26	−50.66	0.1501	0.2062	3427.7	−2.6504
Microcline	$KAlSi_3O_8$	0.278337	10.892	−3975.11	216	−3750.19	0.4488	−1.0075	−1007.3	−3.9731
Monticellite	$CaMgSiO_4$	0.156476	5.148	−2253.05	108.1	−2134.63	0.2507	−1.0433	−797.2	−1.9961
Muscovite	$KAl_2(AlSi_3O_{10})(OH)_2$	0.398313	14.083	−5984.18	292	−5603.71	0.7564	−1.984	−2170	−6.9792
Nepheline	$NaAlSiO_4$	0.145227	5.419	−2095.08	124.4	−1980.35	0.2727	−1.2398	0	−2.7631
Oxygen	O_2	0.031999	2478.920	0	205.2	0	0.0483	−0.0691	499.2	−0.4207
Paragonite	$NaAl_2(AlSi_3O_{10})(OH)_2$	0.382201	13.211	−5946.34	276	−5565.09	0.803	−3.158	217	−8.151
Pargasite	$NaCa_2Mg_4Al(Al_2Si_6O_{22})(OH)_2$	0.835858	27.19	−12720.65	601	−11986.81	1.2802	2.2997	−12359.5	−8.0658
Periclase	MgO	0.040311	1.125	−601.6	26.9	−569.34	0.0605	0.0362	−535.8	−0.2992
Phlogopite	$KMg_3\ (AlSi_3O_{10})(OH)_2$	0.417286	14.964	−6219.16	328	−5837.42	0.7703	−3.6939	−2328.9	−6.5316
Prehnite	$Ca_2Al(AlSi_3O_{10})(OH)_2$	0.412389	14.026	−6203.18	292.8	5825.13	0.7249	−1.3865	−2059	−6.3239
Pyrope	$Mg_3Al_2Si_3O_{12}$	0.40313	11.318	−6284.72	266.3	−5933.62	0.6335	0	−5196.1	−4.3152
Pyrophyllite	$Al_2Si_4O_{10}(OH)_2$	0.360316	12.81	−5640.64	239.4	5266.87	0.7845	−4.2948	1251	−8.4959
Pyroxene Ca–Al	$CaAl_2SiO_6$	0.218125	6.356	−3307.03	138	−3129.29	0.3476	−0.6974	−1781.6	−2.7575
Quartz	SiO_2	0.060085	2.269	−910.83	41.5	−856.46	0.1107	−0.5189	0	−1.1283
Rutile	TiO_2	0.079899	1.882	−944.18	50.6	−888.92	0.0904	0.29	0	−0.6238
Sanidine	$KAlSi_3O_8$	0.277337	10.9	−3964.96	230	−3744.21	0.4488	−1.0075	−1007.3	−3.9731
Silica liquid	SiO_2	0.060085	2.64	−920.85	16.5	0.0825	0	0	0	
Sillimanite	Al_2SiO_5	0.162046	4.986	−2585.68	95.5	−2438.93	0.2802	−0.69	−1375.7	−2.3994
Spessartine	$Mn_3Al_2Si_3O_{12}$	0.40313	11.792	−5646.4	367	−5326.31	0.5846	−0.1593	−7516.7	−2.7501
Spinel	$MgAl_2O_4$	0.142273	3.978	−2300.72	81.5	−2175.64	0.2427	−0.6037	−2315.1	−1.6781
Staurolite (Fe)	$Fe_4Al_{18}Si_{7.5}O_{48}H_4$	1.69170	44.88	−23753.93	1010	−22282.23	2.88	−5.6595	−10642	−25.373
Stishovite	SiO_2	0.060085	1.401	−875.63	24.5	−816.2	0.0681	0.601	−1978.2	−0.0821
Talc	$Mg_3Si_4O_{10}(OH)_2$	0.379289	13.625	−5897.1	260	−5516.73	0.6222	0	−6385.5	−3.9163
Tremolite	$Ca_2Mg_5Si_8O_{22}(OH)_2$	0.81241	27.27	−12310.38	550	−11581.42	1.2602	0.383	−11455	−8.2376
Tridymite	SiO_2	0.060085	2.7	−906.69	46.1	−853.69	0.0979	−0.335	−636.2	−0.774
Ulvöspinel	Fe_2TiO_4	0.22359	4.682	−1497.49	175	−1401.79	−0.1026	14.252	−9144.5	5.2707
Wairakite	$CaAl_2Si_4O_{12} \cdot H_2O$	0.434411	19.04	−6666.42	375	−6220.05	0.8383	−2.146	−2272	−7.2923
Water (steam)	H_2O	0.0180153	2478.920	−241.81	188.8	−228.59	0.0401	0.8656	487.5	−0.2512
Wollastonite	$CaSiO_3$	0.116164	3.993	−1634.06	82.5	−1548.47	0.1593	0	−967.3	−1.0754
Zoisite	$Ca_2Al_3Si_3O_{12}(OH)$	0.622882	13.575	−6898.61	297	−6502.25	0.5957	6.2297	−5921.3	−3.3947

The coefficients a, b, c, and d are from the heat capacity polynomial $C_p = a + bT + cT^{-2} + dT^{-1/2}$ (Note that values of b have been multiplied by 10^5. The unit of volume, J bar^{-1} = 10 cm^3 = 10^{-5} m^3). From Holland and Powell (1998) and from the Thermocalc web page at www.earthsci.unimelb.edu.au/tpg/thermocalc.

Heat capacities, however, vary with temperature; thus C_P must be expressed as a function of T before Eq. (7.12) can be integrated. Variations in the heat capacity of most minerals can be fitted to an expression of the form

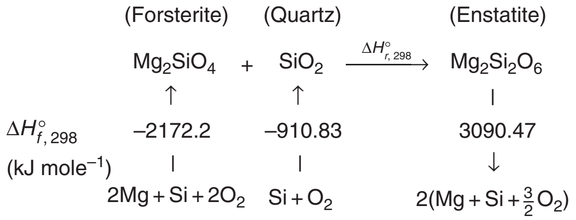
$$C_P = a + bT + \frac{c}{T^2} + \frac{d}{T^{1/2}} \tag{7.13}$$

Values of these coefficients are given in Table 7.1 for the common minerals (note that the values of b in Table 7.1 have been multiplied by 10^5). Substitution of Eq. (7.13) into Eq. (7.12) gives, on integration,

$$H^\circ_T - H^\circ_{298} = a(T - 298) + \frac{b}{2}\left(T^2 - 298^2\right) - c\left(\frac{1}{T} - \frac{1}{298}\right) + 2d\left(T^{1/2} - 298^{1/2}\right) \tag{7.14}$$

Calculation of the high-temperature enthalpy of formation of a mineral is therefore a simple matter using Eqs. (7.11) and (7.14) and the data in Table 7.1 (Problem 7.3). Because the calculations are tedious, they are best carried out by computer; they can be handled easily on the simple spreadsheet.

The actual reactions involved in forming minerals from the elements are of little interest in themselves, as most do not occur in nature. Elemental silicon, for example, is never found reacting with oxygen to form quartz. The enthalpies of these reactions, however, can be used to calculate the enthalpies of reaction (ΔH°_r) between other minerals, and herein lies the value of the standard heats of formation. To illustrate this, consider the petrologically important reaction of olivine with quartz to form orthopyroxene:



$$\mathrm{Mg_2SiO_4} + \mathrm{SiO_2} \xrightarrow{\Delta H^\circ_{r,298}} \mathrm{Mg_2Si_2O_6}$$

For each of these minerals it is possible to write a reaction for their formation from the elements. The enthalpies of these reactions are obtained directly from Table 7.1. These reactions provide another path between the reactants and products. Enstatite, for example, could be broken down into its constituent elements; these elements could then be recombined to form forsterite and quartz; reaction of forsterite with quartz returns us to enstatite. Because enthalpy is a state property, its integral around this cycle must be zero. We can determine the enthalpy change of the reaction ($\Delta H^\circ_{r,298}$), then, by summing all of these terms as we proceed around the cycle in one direction – clockwise, for example. In doing this, care must be taken to keep the signs of the enthalpy changes correct. Table 7.1 indicates that the enthalpy of formation of a mineral from the elements is negative; that is, heat is liberated into the surroundings when elements are combined to form the mineral. If the reaction takes place in the opposite direction, that is, the mineral breaks down into the elements, the enthalpy change must be positive. For this cycle we can write

$$\Delta H^\circ_{r,298} + 3090.47 + (-910.83) + (-2172.2) = 0$$

from which it follows that $\Delta H^\circ_{r,298} = -7.4$ kJ mol^{-1} of olivine. The enthalpy change of this reaction at higher temperatures can be calculated using the enthalpies of formation of the minerals at higher temperatures (Problem 7.4).

7.5 SECOND LAW OF THERMODYNAMICS

Determination of the enthalpy change accompanying a reaction is the first step to understanding the conditions under which a reaction will take place. There remains the important question of the direction of the reaction. Will mineral A change into B, or will B change into A? We know from experience that many everyday processes have a definite direction to them. When cream is stirred in coffee, mixing occurs; if the direction of stirring is reversed, the coffee and cream do not unmix. It is therefore a matter of experience that the process of stirring results in mixing; the opposite is never observed. This implies that some fundamental principle governs the direction of the process. If this principle can be determined, it could be used to indicate the directions of reactions with which we do not have everyday experience, such as those occurring in the Earth. This principle is embodied in the second law of thermodynamics, and it involves an obscure property of material known as entropy (S).

In discussing the first law of thermodynamics it was emphasized that the heat involved with a change from one state to another has no definite value unless the path is specified. This can be illustrated by considering different ways in which a gas can be expanded from one state to another. Imagine that this is done reversibly, although we know, in reality, such a process would take infinite time. Consider first the isothermal expansion of the gas from an initial state A to a final state B (Fig. 7.4). During this expansion a quantity of heat, Q_4, is absorbed from the surroundings in order to keep the temperature constant. If the temperature is not maintained constant during the expansion, many other paths can be followed between A and B. For example, the gas could expand in an insulated container where no heat would be absorbed from the surroundings, and as a result its temperature would fall, for example to T_1 (C in Fig. 7.4). Such a change, in which no heat is transferred in or out of the system, is *adiabatic*. From point C, the gas could expand isothermally to point D with the absorption of a quantity of heat Q_1. Adiabatic compression would then take the gas to point B. The expansion of the gas from A to B could also involve several isothermal steps, such as the path $AEFGHB$, in which two quantities of heat, Q_3 and Q_2, are absorbed along the isothermal lines T_3 and T_2, respectively. Even when the path does not follow adiabatic or isothermal lines, as along the irregular curve in Figure 7.4, the path can be treated as a large number of infinitesimal adiabatic and isothermal steps, with

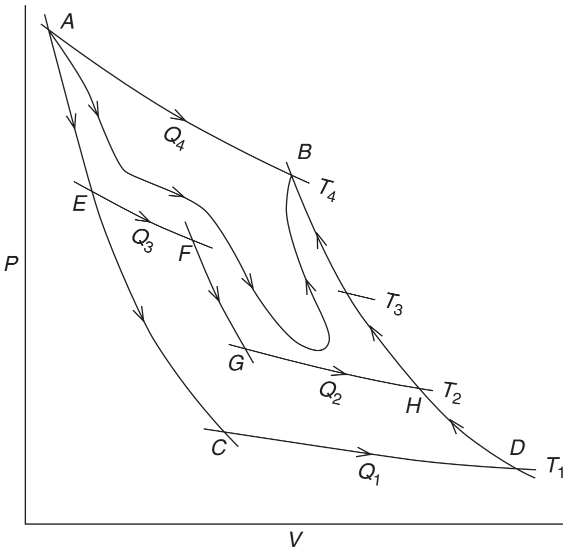


Fig. 7.4 Various paths that can be followed by a gas changing its pressure and volume from state *A* to *B*. The curves labeled T_1 to T_4 are isothermal lines arranged in order of increasing temperature, whereas the steeper curves are adiabatic lines.

each of the latter involving the absorption of a quantity of heat, $\text{đ}Q$, so that the total heat absorbed between *A* and *B* is $\int_A^B \text{đ}Q_{\text{rev}}$. The subscript $_{\text{rev}}$ indicates that the process takes place reversibly.

Each one of the paths between *A* and *B* in Fig. 7.4 involves a different quantity of heat; that is,

$$Q_4 \neq Q_1 \neq (Q_3 + Q_2) \neq \int_A^B \text{đ}Q_{\text{rev}}$$

Because the amount of heat involved in the change from *A* to *B* depends on the path followed, heat cannot be a state property. If, however, the various quantities of heat are divided by the absolute temperatures at which the heat absorption takes place, a function is created that is independent of the path followed. This function must, therefore, be a state function. Thus,

$$\frac{Q_4}{T_4} = \frac{Q_1}{T_1} = \frac{Q_3}{T_3} + \frac{Q_2}{T_2} = \int_A^B \frac{\text{đ}Q_{\text{rev}}}{T}$$

This state property is given the name *entropy* and symbol *S*. It is defined by the equation

$$\mathrm{d}S \equiv \frac{\text{đ}Q_{\text{rev}}}{T} \tag{7.15}$$

As with any state property, the total change in entropy accompanying a reversible cycle is zero; that is,

$$\oint \mathrm{d}S = \oint \frac{\text{đ}Q_{\text{rev}}}{T} = 0$$

A reversible cycle, however, is not a natural one. In an irreversible or natural cycle the amount of heat generated in the surroundings is greater than in a reversible one (review the sign convention in Fig. 7.1). Consequently, we can write

$$\oint \frac{\text{đ}Q_{\text{irrev}}}{T} < \oint \frac{\text{đ}Q_{\text{rev}}}{T} = \oint \mathrm{d}S = 0 \tag{7.16}$$

or in general form

$$\mathrm{d}S \geq \frac{\text{đ}Q}{T} \tag{7.17}$$

where the equality sign applies to the reversible case and the inequality sign to the irreversible one. For a real reaction to take place (irreversible), $\text{đ}Q/T$ must be less than d*S*. Also, for a process taking place in an isolated system, $\text{đ}Q_{\text{irrev}}$ must be zero, and therefore $\mathrm{d}S > 0$. *That is, for a real reaction to occur in an isolated system entropy must increase. The reaction will continue until equilibrium is attained, at which point* d*S* *becomes zero and the entropy is a maximum*. This statement is but one of many different ways of expressing the *second law of thermodynamics*.

It is important to emphasize that in Eq. (7.15), entropy is defined using the heat involved in a *reversible* reaction. The fact that a reversible reaction is not possible does not invalidate the definition. Entropy is a state function, and therefore its change in value depends only on the initial and final states and not on the path of the reaction or whether it was carried out reversibly or irreversibly. All natural reactions are irreversible, and this simply means that $\text{đ}Q_{\text{irrev}}/T < \mathrm{d}S$.

We will see later how entropy can be measured. But first, we will investigate the physical significance of entropy.

7.6 ENTROPY

Thermodynamic terms such as pressure, work, and heat are familiar from everyday experiences, but entropy, despite its importance, is not. But our expectations that stirring cream in coffee will cause mixing, or that oxygen in the air is unlikely suddenly all to move to one end of a room, or that heat will flow from high to low temperatures, are based on processes in which entropy strives for a maximum. This suggests that entropy is a measure of the degree of randomness in a system. This relation was first formalized by the Austrian physicist Boltzmann, who showed that entropy can be defined in terms of the number of possible arrangements of the particles constituting a system. Entropy, so defined, is given by

$$S = k \ln \Omega \tag{7.18}$$

where *k* is the Boltzmann constant (gas constant per molecule, $R/N_0 = 1.3806 \times 10^{-23}$ J K^{-1}) and Ω is the number of possible arrangements. From this relation, entropy is clearly related to the amount of disorder or randomness in a system.

To illustrate this relation, we will consider the entropy change accompanying the transformation from low- to high-temperature albite. Albite is a framework silicate with four asymmetric tetrahedral sites, three occupied by silicon and one by aluminum. In the low-temperature form, aluminum enters one specific site, but at high temperature, it may be in any of the four sites. Consequently, the high-temperature form has a greater capacity for randomness and should therefore have the higher entropy.

In low-temperature albite, the atoms can be arranged in only one way. Admittedly, the silicon atoms could be switched around in the silicon sites, but silicon atoms are indistinguishable, and thus this would not result in distinguishable arrangements. The entropy due to occupancy of the tetrahedral sites in low albite is therefore

$$S_{\mathrm{LowAb}} = k \ln 1 = 0$$

At high temperature, the aluminum can enter any of the tetrahedral sites. But 1 mole of albite ($NaAlSi_30_8$) contains N_0 (Avogadro's number = 6.022×10^{23}) atoms of aluminum and $3N_0$ atoms of silicon that must be distributed over $4N_0$ tetrahedral sites. The number of possible ways of arranging these is

$$\Omega = \frac{(4N_0)!}{(N_0)!(3N_0)!}$$

Hence,

$$S_{\mathrm{High\,Ab}} = k[\ln(4N_0)! - \ln(N_0)! - \ln(3N_0)!]$$

Because N is very large, we can use Stirling's approximation, that is,

$$\ln N! = N \ln N - N$$

The entropy of the high-temperature form is then

$$S_{\mathrm{High\,Ab}} = kN_o[4 \ln 4 - 3 \ln 3]$$

but $kN_0 = R$, the gas constant (8.31443 J K^{-1} mol^{-1}), so that

$$S_{\mathrm{HighAb}} = 18.70 \text{ J mol}^{-1} \text{ K}^{-1}$$

Therefore, the entropy change due to the disordering of the aluminum and silicon in the tetrahedral sites is $S_{\mathrm{High\,Ab}} - S_{\mathrm{Low\,Ab}}$, which is 18.70 J mol^{-1} K^{-1}.

It should be emphasized that this calculated entropy change is a maximum because no account is taken of any crystal chemical restrictions on the possible groupings of ions. In addition, this calculation pertains only to the change in the configuration of the aluminum and silicon in the tetrahedral sites. The albite structure may have other sources of randomness that contribute to the absolute entropy of this mineral (see Problem 7.5).

7.7 THIRD LAW OF THERMODYNAMICS AND THE MEASUREMENT OF ENTROPY

The third law of thermodynamics states that *the entropy of a pure, perfectly crystalline substance is zero at the absolute zero of temperature*. The entropy of such a substance at temperature T is then

$$S_T = \int_0^T \frac{đQ_{\mathrm{rev}}}{T} = \int_0^T \frac{C_{\mathrm{p}}}{T} \mathrm{d}T \tag{7.19}$$

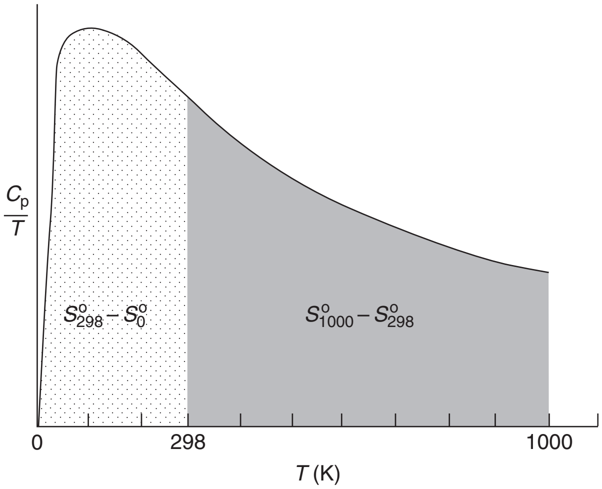


Fig. 7.5 The entropy of a substance is given by the area under the curve in a plot of C_p/T versus T. Entropies of substances at 298 K and 10^5 Pa (1 bar) pressure (stippled area) are listed in Table 7.1. Higher temperature entropies can be obtained by adding the area under the higher-temperature part of the curve (darker shaded area).

where C_{p}, the heat capacity at constant pressure, is a readily measured physical property. Entropy is normally determined by graphically evaluating the integral in Eq. (7.19). This is done by plotting C_{p}/T versus T and measuring the area under the curve (Fig. 7.5).

The terms *pure* and *perfectly crystalline* in the third law are very important. Substances such as glass or intermediate composition plagioclase would still have entropy at absolute zero because of their randomness in structure (*configurational entropy*). For any substance, the entropy can be thought of as consisting of two parts, one thermal and the other configurational; that is,

$$S = S_{\mathrm{thermal}} + S_{\mathrm{configurational}}$$

The S_{thermal} becomes zero at the absolute zero of temperature for all substances, but the $S_{\mathrm{configurational}}$ becomes zero at this temperature only for pure, perfectly crystalline substances. This need not concern us, however, for the entropy data at 298 K in Table 7.1 take this into account.

The entropy of a substance at high temperature and 10^5 Pa pressure can be calculated from the data in Table 7.1 through the following relation:

$$S^{\mathrm{o}}_T = \int_{298}^T \frac{C_{\mathrm{p}}}{T} \mathrm{d}T + S^{\mathrm{o}}_{298} \tag{7.20}$$

The heat capacities at high temperatures, however, are given by Eq. (7.13), with the coefficients being listed in Table 7.1. Integration of Eq. (7.20), therefore, gives

$$S^{\mathrm{o}}_T = \left[a \ln T + bT - \frac{c}{2T^2} - \frac{2d}{T^{1/2}} \right] \Bigg|_{298}^{T} + S^{\mathrm{o}}_{298} \tag{7.21}$$

7.8 GIBBS EQUATION: THERMODYNAMIC POTENTIALS

The first law of thermodynamics gives the relations between the various forms of energy, whereas the second law gives the sense of direction for reactions by introducing the concept of entropy, which is given absolute values by the third law. These can now be combined into a general relation governing reactions and equilibrium.

Rearranging Eq. (7.5) gives

$$-\,\mathrm{d}E - \text{đ}W + \text{đ}Q = 0$$

From the second law (Eq. (7.17)) $T\,\mathrm{d}S \geq \text{đ}Q$; hence,

$$-\,\mathrm{d}E - \text{đ}W + T\mathrm{d}S \geq 0$$

The work can be expressed as work of expansion ($P\,\mathrm{d}V$) plus any other form of work ($\mathrm{d}U$). Thus,

$$-\,\mathrm{d}E - P\mathrm{d}V - \mathrm{d}U + T\mathrm{d}S \geq 0 \tag{7.22}$$

This is a general relation indicating that at equilibrium, which is equivalent to the reversible situation, the left-hand side of the equation must be zero, and for a spontaneous reaction, it must be greater than zero.

Let us consider a reaction that takes place under constant pressure and constant temperature, a common condition encountered in petrologic problems. In this case, $P\mathrm{d}V$ and $T\mathrm{d}S$ can be written as $\mathrm{d}(PV)$ and $\mathrm{d}(TS)$. Equation (7.22) therefore becomes

$$-\,\mathrm{d}E - \mathrm{d}(PV) + \mathrm{d}(TS) \geq \mathrm{d}U$$

or

$$-\,\mathrm{d}(E + PV - TS) \geq \mathrm{d}U \tag{7.23}$$

The combination of terms ($E + PV - TS$) is a state variable and is given the name *Gibbs free energy* (G); that is,

$$G \equiv E + PV - TS = H - TS \tag{7.24}$$

Equation (7.23) becomes

$$-\,\mathrm{d}G \geq \mathrm{d}U$$

and in the case where only work of expansion is done, which is the most common geological situation,

$$-\,\mathrm{d}G \geq 0 \tag{7.25}$$

Thus, for a spontaneous reaction (irreversible) to occur at constant P and T, $-\mathrm{d}G$ must be positive; that is, the free energy must decrease. The reaction will proceed until equilibrium is attained (reversible), at which point $-\mathrm{d}G = 0$, and the free energy is a minimum.

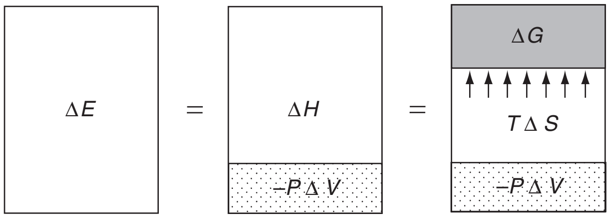


Fig. 7.6 Energy changes involved with a reaction at constant pressure and temperature. The reaction continues spontaneously until ΔG is reduced to zero.

Although various forms of energy are transferred during a reaction proceeding at constant P and T, it is the Gibbs free energy that controls the feasibility and direction of the reaction. This is illustrated graphically in Figure 7.6. The change in internal energy (ΔE) accompanying a reaction taking place at a particular pressure and temperature consists of the work done on the system ($-P\Delta V$) and the enthalpy change (ΔH). The enthalpy change can be further subdivided into thermal energy due to the entropy change ($T\Delta S$) and the change in free energy (ΔG). As the reaction proceeds, entropy is always increasing and striving for a maximum consonant with the state of the system; hence, $T\Delta S$ is also increasing and does so at the expense of ΔG. The free-energy change is therefore that part of the heat removed from the surroundings that can be used to increase the randomness of the system. As the reaction proceeds and the system becomes more random, this fraction of the enthalpy change decreases and becomes zero at equilibrium. At this point it is clear from Figure 7.6 that $T\Delta S = \Delta H$. This also follows directly from Eq. (7.24), which, for an infinitesimal change, would be

$$\mathrm{d}G = \mathrm{d}E + P\mathrm{d}V + V\mathrm{d}P - T\mathrm{d}S - S\mathrm{d}T \tag{7.26}$$

But, if P and T are constant,

$$\mathrm{d}G = \mathrm{d}E + P\mathrm{d}V - T\mathrm{d}S$$

which, from Eq. (7.8), gives

$$\mathrm{d}G = \mathrm{d}H - T\mathrm{d}S$$

which, in turn, for a finite change, becomes

$$\Delta G = \Delta H - T\Delta S \tag{7.27}$$

At equilibrium $\Delta G = 0$; hence,

$$T\Delta S = \Delta H$$

or

$$T_{\text{equil}} = \Delta H / \Delta S \tag{7.28}$$

Equation (7.28) gives a simple means of determining the equilibrium temperature for a reaction. Consider, for example, the reaction of cristobalite melting at 10^5 Pa pressure to form silica liquid. Data for this reaction at high temperature indicate that ΔH is +8071 J mol^{-1} and ΔS is +4.05 J mol^{-1} K^{-1}. Therefore, the equilibrium temperature or melting point would be $T_{\text{equil}} = \Delta H/\Delta S = 1993$ K or 1720 °C.

The equilibrium temperature for any reaction, such as that for the melting of cristobalite, is the temperature at which $\Delta G = 0$. This is shown in Figure 7.7 as the point of intersection of the free-energy curves for cristobalite and silica liquid (glass at low temperature). At higher temperatures, ΔG is

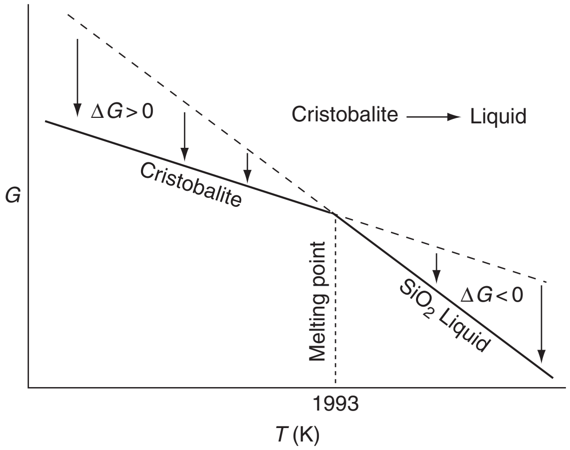


Fig. 7.7 Free energy versus temperature plots at constant pressure for cristobalite and silica liquid. The intersection of the two curves is the melting point of cristobalite. Below the melting point, cristobalite has a lower free energy than does the silica liquid and is therefore the stable phase; above the melting point, the liquid has a lower free energy and is therefore stable.

negative; hence, the reaction proceeds with the melting of cristobalite. Below this temperature, ΔG is positive, so the reaction can proceed only in the opposite direction, causing cristobalite to crystallize. The most stable form is always the one with the lowest free energy. Of course, thermodynamics indicates only what the equilibrium state should be, but kinetic factors may prevent this from being achieved. Volcanic glasses, for example, could lower their free energy by crystallizing, but the kinetics of this process are slow at low temperatures.

For many petrologic problems, we need to calculate the ΔG of reaction under conditions different from those for which the data in Table 7.1 are applicable. To do this it is necessary to know how free energy changes with temperature and pressure. From Eq. (7.22) it is clear that for a reversible reaction involving only work of expansion, dE is equivalent to TdS − PdV, which can be substituted for dE in Eq. (7.26), giving

$$\mathrm{d}G = T\mathrm{d}S - P\mathrm{d}V + P\mathrm{d}V + V\mathrm{d}P - T\mathrm{d}S - S\mathrm{d}T$$

which reduces to

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P \tag{7.29}$$

In a closed system – that is, one in which no matter is transferred in or out – G is a function only of temperature and pressure [$G = f(T, P)$]. We can express the total change in G (dG) resulting from a change in T and P as the sum of the change due to T and the change due to P. This is known as a *total differential* of the function, and it is represented as follows:

$$\mathrm{d}G = \left(\frac{\partial G}{\partial T}\right)_P \mathrm{d}T + \left(\frac{\partial G}{\partial P}\right)_T \mathrm{d}P \tag{7.30}$$

The terms in parentheses are known as *partial derivatives*, as they denote the variation in G with respect to only one of the two variables, while the other variable, shown as a subscript outside the parentheses, is held constant. Comparison of Eqs. (7.30) and (7.29) reveals that

$$\left(\frac{\partial G}{\partial T}\right)_P = -S \tag{7.31}$$

and

$$\left(\frac{\partial G}{\partial P}\right)_T = \bar{V} \tag{7.32}$$

The bar over the $\bar{V}$ signifies molar volume. This is introduced because values of G are typically given as molar quantities. Because all substances have positive entropy, free energy always decreases with increasing temperature at constant pressure, and because liquids have higher entropies than corresponding solids, their free energy decreases more rapidly than that of solids (Fig. 7.7). Molar volumes are also always positive, therefore increasing pressure at constant temperature causes the free energy to rise.

Similar relations can be derived for the free-energy change of a reaction. The change in the ΔG of a reaction with temperature is

$$\left(\frac{\partial \Delta G}{\partial T}\right)_P = -\Delta S \tag{7.33}$$

and the change with pressure is

$$\left(\frac{\partial \Delta G}{\partial P}\right)_T = \Delta V \tag{7.34}$$

In these cases, the change in the ΔG of the reaction is determined by the entropy change and volume change of the reaction. Both of these terms can be either positive or negative, so generalizations about the variation of the ΔG of a reaction cannot be made. However, in the case of reactions that evolve a gas such as many metamorphic reactions, ΔS and ΔV will both be positive, and therefore the ΔG of reaction will decrease with increasing temperature and increase with increasing pressure.

Other useful relations can be derived from Eq. (7.30) simply by utilizing two properties of a total differential, that is the *cross-derivative rule* and the *cyclic rule*. We have seen by comparing Eqs. (7.29) and (7.30) that

$$\left(\frac{\partial G}{\partial T}\right)_P = -S \quad \text{and} \quad \left(\frac{\partial G}{\partial P}\right)_T = \bar{V}$$

If we take derivatives of these expressions with respect to the variable held constant, we obtain

$$\left[\frac{\partial}{\partial P}\left(\frac{\partial G}{\partial T}\right)_P\right]_T = -\left(\frac{\partial S}{\partial P}\right)_T \qquad \text{and}$$

$$\left[\frac{\partial}{\partial T}\left(\frac{\partial G}{\partial P}\right)_T\right]_P = \left(\frac{\partial V}{\partial T}\right)_P$$

But, from the *cross-derivative rule* for a total derivative that is exact,

$$\left[\frac{\partial}{\partial P}\left(\frac{\partial G}{\partial T}\right)_P\right]_T = \left[\frac{\partial}{\partial T}\left(\frac{\partial G}{\partial P}\right)_T\right]_P \tag{7.35}$$

Therefore,

$$-\left(\frac{\partial S}{\partial P}\right)_T = \left(\frac{\partial V}{\partial T}\right)_P$$

but from Eq. (2.2), $(\partial V/\partial T)_P = \alpha V$, where α is the isobaric coefficient of thermal expansion. Therefore,

$$\left(\frac{\partial S}{\partial P}\right)_T = -\alpha V \tag{7.36}$$

This allows us to evaluate the variation in entropy with pressure from two easily measured physical properties, α and V.

Similarly, by differentiating Eq. (7.32) we obtain

$$\left(\frac{\partial^2 G}{\partial P^2}\right)_T = \left(\frac{\partial V}{\partial P}\right)_T$$

but from Eq. (2.8) $(\partial V/\partial P)_T = -\beta V$, where β is the isothermal coefficient of compressibility. Therefore,

$$\left(\frac{\partial^2 G}{\partial P^2}\right)_T = -\beta V \tag{7.37}$$

And by differentiating Eq. (7.31) we obtain

$$\left(\frac{\partial^2 G}{\partial T^2}\right)_P = -\left(\frac{\partial S}{\partial T}\right)_P$$

but $(\partial S/\partial T)_P = C_p/T$. Thus,

$$\left(\frac{\partial^2 G}{\partial T^2}\right)_P = -\frac{C_p}{T} \tag{7.38}$$

The *cyclic rule* for total differentials states that if any three variables, *x*, *y*, and *z*, for example, are connected by a functional relation, then the three partial derivatives satisfy the following relation:

$$\left(\frac{\partial x}{\partial y}\right)_z\left(\frac{\partial y}{\partial z}\right)_x\left(\frac{\partial z}{\partial x}\right)_y = -1 \tag{7.39}$$

This rule is easily remembered by writing the three variables in any order in a row, and then repeating them below so that none of the vertical columns match. These vertical pairs give the partial derivatives, with the subscripted variable, which is held constant, being the third variable. For example,

$$\begin{matrix} z\ y\ x \\ x\ z\ y \end{matrix} \quad \text{becomes} \quad \left(\frac{\partial z}{\partial x}\right)_y\left(\frac{\partial y}{\partial z}\right)_x\left(\frac{\partial x}{\partial y}\right)_z = -1$$

If the cyclic rule is applied to the three variables *P*, *T*, and *V*, we obtain

$$\left(\frac{\partial P}{\partial T}\right)_V\left(\frac{\partial T}{\partial V}\right)_P\left(\frac{\partial V}{\partial P}\right)_T = -1 \tag{7.40}$$

But from Eq. (2.2), $(\partial V/\partial T)_P = \alpha V$, and from Eq. (3.1), $(\partial V/\partial P)_T = -\beta V$, which, when substituted into Eq. (7.40), give

$$\left(\frac{\partial P}{\partial T}\right)_V\left(\frac{1}{\alpha V}\right)(-\beta V) = -1$$

Hence,

$$\left(\frac{\partial P}{\partial T}\right)_V = \frac{\alpha}{\beta} \tag{7.41}$$

Thermodynamic relations, then, can be manipulated with the cyclic and cross-derivative rules into useful forms for specific applications.

7.9 FREE ENERGY OF FORMATION AT ANY TEMPERATURE AND PRESSURE

Finally, we will derive an expression for the free energy of formation of a phase at any temperature and pressure. We define the molar free energy of formation of a phase from the elements at 298.15 K and 10^5 Pa (1 bar) as

$$\begin{aligned} \text{Free energy of formation} &= \Delta\bar{G}^\circ_{f,298} \\ &= \Delta H^\circ_{f,298} - 298\ \Delta S^\circ_{298} \end{aligned} \tag{7.42}$$

The ΔS in this expression refers to the difference in entropies of the phase and its constituent elements in their standard states $\left(S_{\text{phase}} - \sum S_{\text{elements}}\right)$. Values of $\Delta\bar{G}^\circ_{f,298}$ in column 7 of Table 7.1 are calculated according to Eq. (7.42). In Section 7.4 it was shown that if the ΔH°_r is used to calculate the ΔH of a reaction between phases, the terms for the enthalpy of the elements on the reactant and product sides of a reaction cancel. The same is true for the entropies of the elements, and in some compilations of thermodynamic data, $\Delta G^\circ_{f,298}$ is calculated ignoring the entropies of the elements. A value calculated in this manner is known as the *apparent free energy of formation*; it is given by

$$\Delta\bar{G}^\circ_{f,298} = \Delta H^\circ_{f,298} - 298.15 \times S^\circ_{298} \tag{7.43}$$

where S°_{298} is the entropy of the phase only. For example, the reaction to form quartz from the elements in their standard states is

$$S_i + O_2 \rightarrow SiO_2$$

According to Eq. (7.42), the $\Delta\bar{G}^\circ_{f,298}$ for quartz would be

$$\begin{aligned} \Delta\bar{G}^\circ_{f,298} &= \Delta H^\circ_{f,298} - 298.15\left(S^\circ_{298,\text{Qtz}} - \left(S^\circ_{298,\text{Si}} + S^\circ_{298,\text{O}_2}\right)\right) \\ &= -856.46\ \text{kJ} \end{aligned}$$

The apparent free energy of formation (Eq. (7.43)) would be

$$\Delta\bar{G}^\circ_{f,298} = \Delta H^\circ_{f,298} - 298.15 \times S^\circ_{298,\text{Qtz}} = -923.06\ \text{kJ}$$

Readers should take care to ascertain which type of free energy of formation is being used when they consult other texts.

The reason for using the apparent free energy of formation is that it reduces the amount of calculation necessary when determining the free energy of formation of a phase at elevated temperatures and pressures. It eliminates having to calculate the changes in the enthalpies and entropies of the elements when the temperature and pressure are raised above standard conditions. We will illustrate this by deriving the expression for the apparent free energy of formation of a phase at any temperature T (K) and pressure P (Pa).

We first determine $\Delta\bar{G}^{\circ}_{f,298}$ at standard conditions from Eq. (7.43). If the temperature is to be raised, appropriate terms must be added for changes in enthalpy, $\int \mathrm{d}H$ (Eq. (7.12)), and entropy, $\int \mathrm{d}S$ (Eq. (7.20)). If the pressure is to be changed, the free energy must also be adjusted according to Eq. (7.32). Combining all these terms into one expression gives

$$\Delta\bar{G}_{f,T,P} = \left(\Delta H^{\circ}_{f,298} + \int_{298}^{T} C_{\mathrm{p}}\mathrm{d}T\right) - T\left(S^{\circ}_{298} + \int_{298}^{T}\left(\frac{C_{\mathrm{p}}}{T}\right)\mathrm{dT}\right) + \int_{10^5}^{P} V\mathrm{d}P \qquad (7.44)$$

Substituting the polynomial expression for C_{p} (Eq. (7.13)) and integrating, we obtain

$$\begin{aligned}\Delta\bar{G}_{f,T,P} = {} & \Delta H^{\circ}_{f,298} - TS^{\circ}_{298} + a(T-298) + \frac{b}{2}\left(T^2 - 298^2\right) \\ & - c\left(\frac{1}{T} - \frac{1}{298}\right) + 2d\left(T^{1/2} - 298^{1/2}\right) \\ & - T\left[a\,\ln\left(\frac{T}{298}\right) + b(T-298)\right. \\ & \left. - \frac{c}{2}\left(\frac{1}{T^2} - \frac{1}{298^2}\right) - 2d\left(\frac{1}{T^{1/2}} - \frac{1}{298^{1/2}}\right)\right] \\ & + \int_{10^5}^{P} V\mathrm{d}P \end{aligned} \qquad (7.45)$$

If a computer has not yet been used in solving problems in this text, Eq. (7.45) will rapidly convince the reader of its advantages. The last term in this equation has not been integrated. If pressure remains constant, the term becomes zero. Because the compressibilities of minerals and magmas are extremely small, V can be considered a constant, so when pressure does vary, this term becomes $V(P - 10^5)$. But for a gas, V is certainly not a constant, and the variation of V with P must be known before we can integrate this term. This problem is dealt with in Section 8.3.

7.10 PROBLEMS

7.1 If the molar volume of a peridotite is 5×10^{-5} m^3, and its volume increase on totally melting is 10%, how much work is done when 1 mol of peridotite melts at a depth where the pressure is 2 GPa? Note that the system being considered is the 1 mol of rock. Be certain to get the sign convention correct (see Fig. 7.1).

7.2 If the molar volume of granitic magma is 7×10^{-5} m^3, and on crystallizing it decreases by 10%, compare the work done by magma crystallizing near the top of a batholith, where the pressure is 0.05 GPa, with magma crystallizing near the base of the batholith, where the pressure is 0.5 GPa. Be careful of the sign convention.

7.3 From data in Table 7.1, and using Eq. (7.14), determine the enthalpy of formation of kyanite and andalusite at 466 K and 10^5 Pa (1 bar). If kyanite were to change into andalusite under these conditions, what would be the enthalpy of reaction? Is the reaction exothermic or endothermic? (Be careful of the sign convention; write the reaction kyanite → andalusite, then ΔH_{r} is the final enthalpy minus the initial.)

7.4 Using the ΔH°_{f} at 298.15 K and heat capacity data in Table 7.1, calculate the enthalpy of reaction at 10^5 Pa and 1500 K for the reaction forsterite + quartz → enstatite.

7.5 In dolomite, calcium has two different possible sites to occupy at high temperatures, but at low temperatures it preferentially enters one of these sites, and magnesium occupies the other. Calculate the configurational entropy associated with the complete disordering of dolomite.

7.6 If entropy is a function of temperature and pressure, that is, $S = S(T,P)$, (**a**) write the total differential of $S(T,P)$, and (**b**) show that the total differential of $S(T,P)$ is given by

$$\mathrm{d}S = \left(\frac{C_{\mathrm{p}}}{T}\right)\mathrm{d}T - \alpha V\mathrm{d}P$$

where α is the coefficient of thermal expansion (Eq. (2.2)). [*Hint*: Use derivatives of $G(T,P)$.]

7.7 Convection within the mantle or within a magma chamber causes the thermal gradient to approach the adiabatic gradient ($\text{đ}Q = 0$ and $\mathrm{d}S = 0$). Using the expression for dS in Problem 7.6b, along with the cyclical rule for partial derivatives, show that the adiabatic gradient is given by

$$\left(\frac{\partial T}{\partial P}\right)_S = \frac{T\bar{V}\alpha}{C_{\mathrm{p}}}$$

7.8 The granitic magma in Problem 7.2, which has a molar volume of 7×10^{-5} m^3, has a molecular weight of 0.168 kg ($\rho = 2.4$ Mg m^{-3}), a coefficient of thermal expansion, α, of 2×10^{-5} K^{-1}, a heat capacity, C_{p}, of 0.8 kJ kg^{-1} K^{-1}, and a temperature of 900 °C (1173 K).

- **(a)** If magmatic convection has established an adiabatic temperature gradient within the body, calculate the value of $(\partial T/\partial P)_S$ from the relation in Problem 7.7. (Note 1 J = 1 Pa m^3.)
- **(b)** Using the relation $\mathrm{d}P = \rho g\ \mathrm{d}z$ (Eq. (1.1)), calculate the adiabatic temperature gradient, $(\partial T/\partial z)_S$, in this convecting body of magma, and compare this value with a typical geothermal gradient in the upper continental crust.

7.9 From the data in Table 7.1 and using Eq. (7.21), calculate the entropies of kyanite and andalusite at 466 K and

10^5 Pa. What entropy change accompanies the change of kyanite to andalusite under these conditions?

7.10 Using the results for ΔH° and ΔS° from Problem 7.3 and 7.9 respectively, calculate the free energy change, ΔG_r (Eq. (7.27)), for the reaction of kyanite to andalusite at 466 K and 10^5 Pa. From the value of the $\Delta G_\mathrm{r}^\circ$, what can you conclude about this reaction under these conditions of temperature and pressure?

7.11 For the reaction kyanite → andalusite at 466 K and 10^5 Pa calculated in Problem 7.10, determine the change in the ΔG_r if the pressure is increased to 10^8 Pa at 466 K. Which of the minerals will be more stable under these new conditions? (Recall that $1\ \mathrm{J} = 1\ \mathrm{Pa\ m^3}$.)

7.12 For the reaction kyanite → andalusite calculated in Problem 7.11, calculate how much the temperature would have to be increased at a pressure of 10^8 Pa in order to return the value of ΔG_r to zero, that is, to reestablish equilibrium. Assume that the entropy change for the reaction remains constant and is equal to the value calculated in Problem 7.9.

7.13 Using Eq. (7.45), calculate the ΔG_f of calcite and of aragonite at 800 K and 10^5 Pa. Which phase is more stable under these conditions?

7.14 Using Eq. (7.45), calculate the ΔG_f of calcite and of aragonite at 298.15 K and 0.5 GPa. Which phase is more stable under these conditions?

8 Free energy and phase equilibria

8.1 INTRODUCTION

In Chapter 7, our experiences of the behavior of energy in nature were formalized into laws dealing with the conservation of energy, the natural direction of processes, and the absolute zero of temperature. Once given mathematical expression, these laws were used to relate important thermodynamic functions to easily measured properties. The thermodynamic functions, in turn, were used to determine the direction of reactions and the conditions necessary for equilibrium. In this chapter, we deal with the practical aspects of applying these thermodynamic principles to determining the equilibrium relations between minerals.

We have seen that energy can take various forms – mechanical, thermal, chemical, electrical. In each case, the energy can be expressed as the product of two terms, one *intensive* and the other *extensive*. Intensive ones can be measured at any point in a system and are independent of the extent of the system; they can be thought of as environmental factors. They include pressure, temperature, and chemical and electrical potentials. The extensive ones, on the other hand, must be integrated over the entire system. These include volume, entropy, number of moles, and electrical charge. Changes in energy can then be expressed as follows:

Mechanical	Thermal	Chemical	Electrical
$P\,\mathrm{d}V$	$T\,\mathrm{d}S$	$\mu_i\,\mathrm{d}n_i$	$\mathcal{E}\,\mathrm{d}Q$

where μ_i is the chemical potential of substance i (Chapter 9), n_i the number of moles of i, $\mathcal{E}$ the electrical potential, Q the electrical charge, and all other terms are as defined in Chapter 7.

Writing energies as the products of intensive and extensive variables provides insight into the meaning of the more abstract thermodynamic terms. Just as electrical charge is the capacity of a system for electrical energy, so entropy is the capacity of a system for thermal energy. Similarly, while electrical potential determines the level to which the electrical capacity is exploited, temperature determines how much the thermal capacity will be exploited. Similarly, volume is the capacity of a system for mechanical energy, and pressure is the mechanical potential that exploits that volume. The capacity for chemical energy is the number of moles of substance present in the system, and the chemical potential is the environmental factor that exploits that capacity.

Another important reason for distinguishing these two types of variable is that at equilibrium, any intensive variable must have the same value throughout a system; that is, no gradients in the potentials of mechanical (P), thermal (T), chemical (μ_i), and electrical ($\mathcal{E}$) energy can exist. Although perhaps self-evident, this principle is so important that it is commonly referred to as the *zeroth law* of thermodynamics.

In determining the equilibrium between minerals, we constantly have to evaluate how variations in the intensive and extensive variables affect equilibrium. One particularly important question is how many of these variables can be changed simultaneously without disturbing equilibrium. If a system can be described in terms of n variables, and r equations relate these variables, then $(n-r)$ variables can be changed independently while maintaining equilibrium. The number $(n-r)$ is called the *variance* or number of *degrees of freedom* of the system.

We can illustrate the significance of variance by examining a simple system consisting of 1 mol of ideal gas. Because only 1 mol is considered, the number of moles is not a variable. The system, then, can be described in terms of the two intensive variables P and T and the one extensive variable, V. But these three variables are related by the ideal gas law ($PV = RT$), so that the variance of the system must be 2 (i.e. $3-1$); that is, any two variables can be changed independently, but the value of the third is determined by the equation relating the variables. We will have occasion to refer often to the variance of systems. The numbers of variables and relations between variables may be considerably greater than those considered in this simple example, but the same principle will be involved (see Section 10.5).

8.2 FREE ENERGY SURFACE IN *G–T–P* SPACE

In Section 7.8 we saw that in a closed system, that is, one in which no matter is transferred in or out, the change in the Gibbs free energy, dG, is a function only of T and P; it can therefore be expressed as a total differential in terms of T and P (Eq. (7.30)). If we consider a system consisting only of a phase α, the free energy of the system at any T and P could be obtained by integrating Eq. (7.30); that is,

$$\int_{\bar{G}_0^\alpha}^{\bar{G}_{T,P}^\alpha} \mathrm{d}G = \int_{T_0}^{T} \left(\frac{\partial \bar{G}^\alpha}{\partial T}\right)_P \mathrm{d}T + \int_{P_0}^{P} \left(\frac{\partial \bar{G}^\alpha}{\partial P}\right)_T \mathrm{d}P \qquad (8.1)$$

where T_0 = 298 K, P_0 = 10^5 Pa, and $\bar{G}_0^\alpha$ = free energy per mole at 298 K and 10^5 Pa. From Eqs. (7.31) and (7.32), we know that $\left(\partial\bar{G}^{\alpha}/\partial T\right)_P$ is $-\bar{S}^{\alpha}$ and $\left(\partial\bar{G}^{\alpha}/\partial P\right)_T$ is $\bar{V}^{\alpha}$, respectively. If we assume that *S* and *V* remain constant over the range of conditions considered, integration of Eq. (8.1) gives

$$\bar{G}^{\alpha}_{T,P} = \bar{G}^{\alpha}_{T_0,P_0} - \bar{S}^{\alpha}(T - T_0) + \bar{V}^{\alpha}(P - P_0) \tag{8.2}$$

This relation is the equation of a surface in *G–T–P* space that has an intercept on the *G* axis of $\bar{G}^{\alpha}_{T_0,P_0}$ (Fig. 8.1). If *P* is held constant, the slope on the surface is $-\bar{S}^{\alpha}$; if *T* is held constant, the slope is $\bar{V}^{\alpha}$. In reality, *S* and *V* would be functions of *T* and *P*, so that slopes would change with *T* and *P*, and the surface would be curved. For simplicity of illustration, however, *S* and *V* have been taken as constants.

If the system contains a second phase *β*, a polymorph of phase *α*, for example, another surface in *G–T–P* space can be constructed (Fig. 8.2). Phase *β* will most likely have different values of *S* and *V* from those of phase *α*. Thus, the free-energy surfaces of the two phases will not be parallel and will intersect along a line where $\bar{G}^{\alpha} = \bar{G}^{\beta}$. But if $\bar{G}^{\alpha} = \bar{G}^{\beta}$ and phase *α* changes into phase *β*, the free energy would not change; that is, $\Delta\bar{G}_r = 0$. According to Eq. (7.25), this condition describes equilibrium. The line of intersection of the two surfaces therefore gives the values of *T* and *P* under which phases *α* and *β* are in equilibrium with each other; that is, they can coexist.

We know from Eq. (7.25) that a spontaneous reaction is one in which *dG* is negative; that is, the reaction proceeds in a direction that minimizes *G*. On the low-temperature/low-pressure side of the line of intersection of the two surfaces in the *G–T–P* plot of Figure 8.2, the free energy of the system can be lowered if phase *β* changes into phase *α* . In this region, then, phase *α* is stable relative to phase *β*. At pressures and temperatures above the line of intersection of the two surfaces, the *β* phase has the lower free energy and is therefore stable relative to the *α* phase. It is important to emphasize that these stabilities must be given relative to one another and not as absolute stabilities. For example, although the *α* phase has a lower free energy than the *β* phase at low temperature and is therefore the more stable of the two under these conditions, a third phase *γ* might exist that could have a lower free energy than either *α* or *β*. We would still be able to say that *α* was more stable than *β* under these conditions, but both phases would be unstable relative to the *γ* phase.

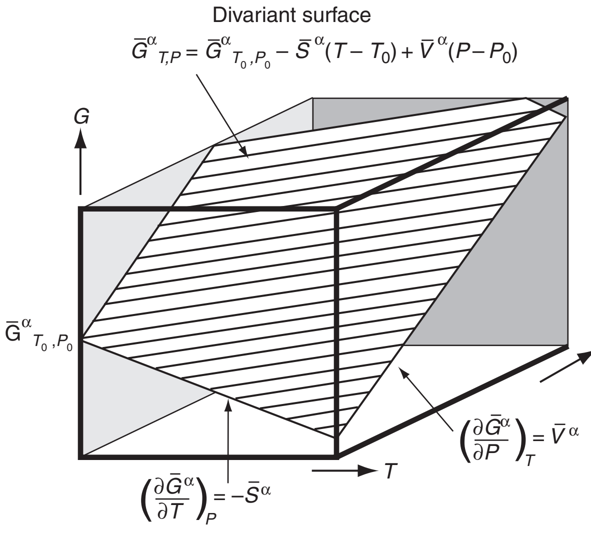


Fig. 8.1 Free-energy surface of phase *α* plotted in *G–T–P* space. The surface is ruled with lines of constant free energy.

Let us determine the variance along the line $\bar{G}^{\alpha} = \bar{G}^{\beta}$ ($\Delta G = 0$) in Figure 8.2. The system has three variables *G*, *T*, and *P*, which are related through Eq. (8.2) for phase *α*. A similar but independent equation can be written for phase *β*. Thus, the three variables are related through two independent equations, and the variance must be $3-2 = 1$. The assemblage of $\alpha+\beta$ is therefore said to be *univariant*. Any one of the three variables can be changed, but the values of the other two are then fixed by the equations. If only one of the phases is present, the three variables are related by only one equation, and the variance of the system would be 2. This *divariant* surface would be the surface for the *α* or the *β* phase in the *G–T–P* plot.

The three-dimensional plot of *G–T–P* space is commonly reduced to two dimensions by taking sections through it at constant *P* or constant *T*, or projecting the data parallel to the *G* axis onto the *P–T* plane (Fig. 8.2). In a section at constant *P* or *T*, the two surfaces for phases *α* and *β* appear as lines that intersect at a point where the univariant line $\bar{G}^{\alpha} = \bar{G}^{\beta}$ pierces the plane of the section. The variance of the assemblage $\alpha+\beta$ therefore becomes zero (*invariant*) under these conditions. This follows from the fact that the number of variables is reduced to two by holding one of them constant while two equations still relate the variables. With only two variables and two equations, there can be no variance; that is, only one set of conditions will permit phases *α* and *β* to coexist. Should a rock be found containing these phases, we would know precisely the conditions under which it formed.

Despite the simplicity of the sections at constant *T* and *P*, the most useful diagram is the *P–T* projection. In this plot, the variance of the system is not decreased by the projection (nothing is held constant). Thus, the univariant line $\bar{G}^{\alpha} = \bar{G}^{\beta}$ remains a univariant line. Even the divariant surfaces for phases *α* and *β* can be contoured with lines of constant *G*, but in most diagrams only the univariant line is shown. It is important when examining *P–T* diagrams illustrating phase equilibria to realize that they are projections from *G–T–P* space.

To plot univariant lines in *P–T* projections, we must know how to determine their slopes. The difference in free energy (ΔG) between the two surfaces for the *α* and *β* phases in *G–T–P* space is a function only of *T* and *P* in a closed system. According to the cyclic rule for total differentials (Eq. (7.39)), we can therefore write

$$\left(\frac{\partial\Delta G}{\partial T}\right)_P \left(\frac{\partial P}{\partial\Delta G}\right)_T \left(\frac{\partial T}{\partial P}\right)_{\Delta G} = -1$$

from which it follows that

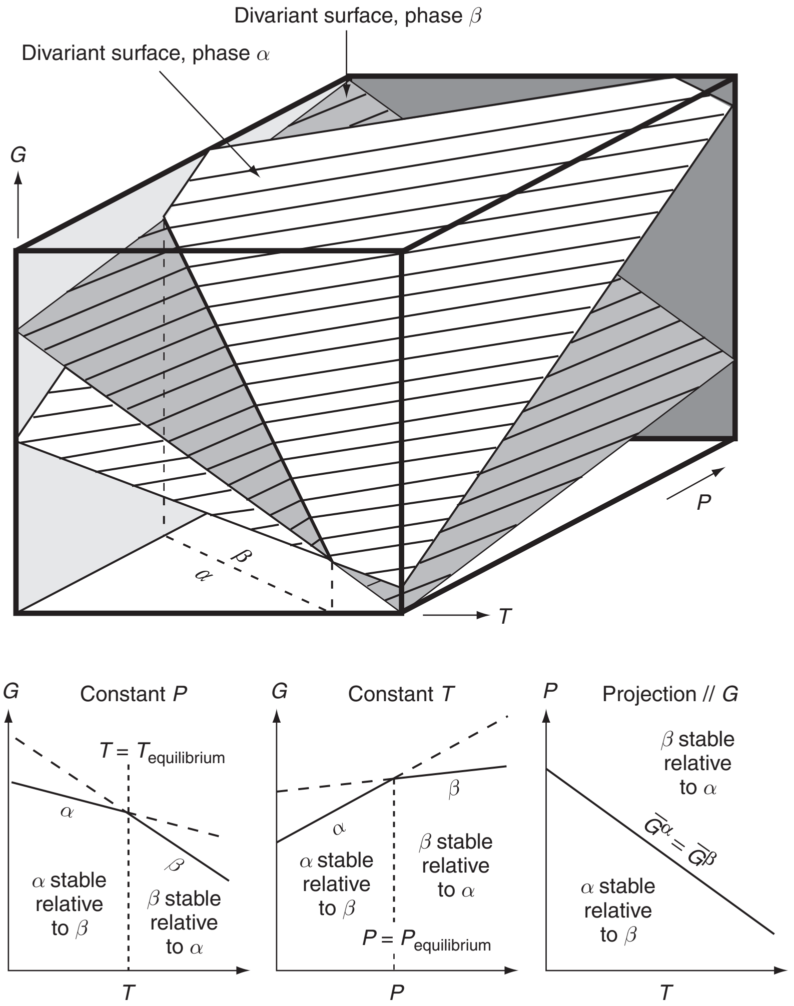


Fig. 8.2 Intersecting free-energy surfaces of phase α and phase β (shaded). Both surfaces are ruled with lines of constant *G*. The line of intersection of the two surfaces, where the free energies of the two phases are equal, indicates the pressure and temperature conditions under which the two phases are in equilibrium with each other. Sections through the *G–T–P* plot at constant *T* and constant *P* are shown below, along with a projection of the line of intersection of the two surfaces on the *P–T* plane.

$$\left(\frac{\partial P}{\partial T}\right)_{\Delta G} = -\frac{(\partial \Delta G/\partial T)_P}{(\partial \Delta G/\partial P)_T}$$

But according to Eqs. (7.33) and (7.34), $(\partial \Delta G/\partial T)_P = -\Delta S$ and $(\partial \Delta G/\partial P)_T = \Delta V$. Thus,

$$\left(\frac{\partial P}{\partial T}\right)_{\Delta G} = \frac{\Delta S_r}{\Delta V_r} \tag{8.3}$$

This is known as the *Clapeyron equation*. It shows that the slope of a univariant reaction ($\alpha \rightarrow \beta$) in a *P–T* projection is given by the entropy change ($S^{\beta} - S^{\alpha}$) divided by the volume change ($V^{\beta} - V^{\alpha}$) of the reaction. The univariant lines can have positive or negative slopes depending on the signs of ΔS and ΔV. A metamorphic reaction in which a gas is liberated will invariably have positive values of ΔS and ΔV, and thus its slope in a *P–T* diagram will be positive. Similarly, most melting is accompanied by positive values of ΔS and ΔV (ice is an exception), and thus most melting reactions also have positive slopes in *P–T* diagrams.

Standard units for ΔS and ΔV in Eq. (8.3) give units of J m^{-3} K^{-1} for $\partial P/\partial T$. But a joule is the amount of work done when a force of 1 newton causes a displacement of 1 meter (1 J = 1 N m). Thus the units of $\partial P/\partial T$ become N m^{-2} K^{-1} or Pa K^{-1}.

For purposes of illustration, the phases α and β have been taken to be polymorphs, such as α and β quartz. But the general principles outlined here are equally applicable when α and β represent assemblages of phases constituting the reactants and products of a reaction. They can be crystalline minerals, magmas, or gases. The entropy and volume changes in these cases are the total changes in these properties determined for the balanced chemical reactions.

8.3 PLOTTING UNIVARIANT LINES IN *P–T* DIAGRAMS

Although the Clapeyron equation (Eq. (8.3)) gives the slope of a univariant line in a *P–T* diagram, it does not locate the line with respect to absolute values of pressure and temperature. To determine this, we must know the equilibrium temperature of the reaction at some specified pressure. We might determine it at 10^5 Pa, for example, using the standard data in Table 7.1 and Eq. (7.28). Such a calculated temperature, however, might be far removed from the particular temperature (and pressure) range of interest. The known starting point might also be derived from experimental work carried out at high pressures and temperatures. In either case, it is necessary to be able to extend the standard thermodynamic data in Table 7.1 to conditions of high pressure and temperature.

We will illustrate the manipulation of thermodynamic data by considering a much-studied reaction, that of the polymorphic transformation of kyanite to andalusite. At low pressures and temperatures this reaction is too slow to be studied in the laboratory. Even at high pressures and temperatures the reaction is extremely sluggish, but careful experiments have shown that the two polymorphs are in equilibrium at 800 °C and 0.72 GPa (Richardson *et al.*, 1969). Is this result in agreement with the thermodynamic data in Table 7.1, and if so, what slope should the reaction have in a *P–T* diagram?

Using Eqs. (7.14), (7.21), and (7.27) and data from Table 7.1, the free energies of formation $(\Delta\bar{G}^{\text{o}}_{f,1073})$ of kyanite and andalusite at 800 °C (1073 K) and 10^5 Pa are determined (see Problems 7.4, 7.9, and 7.10). With these values, we calculate the ΔG_r of the transformation of kyanite to andalusite as follows:

$$\Delta G^{\text{o}}_{r,1073} = \Delta G^{\text{And}}_{f,1073} - \Delta G^{\text{Ky}}_{f,1073} = -5571 \quad \text{J mol}^{-1}$$

Because ΔG_r is negative, the reaction kyanite → andalusite proceeds spontaneously at 800 °C and 10^5 Pa. But the value of ΔG_r can be changed by increasing the pressure. If it is increased by 5571 J mol^{-1}, ΔG_r will become zero and the reaction will be at equilibrium. The change in ΔG_r with pressure is given by Eq. (7.34), which in its integrated form gives

$$\Delta G^{P}_{r,1073} = \Delta G^{\text{o}}_{r,1073} + \Delta V_r\left(P - 10^5\right)$$

From Table 7.1 we obtain $\Delta V^{\text{o}}_{r,298}$. Molar volumes of the minerals will increase slightly with increasing temperature, but because the coefficients of thermal expansion of most minerals are small and similar, the value of ΔV_r remains almost constant. The effect of the coefficient of compressibility is even less important. The value of $\Delta V^{\text{o}}_{r,298}$ can therefore be used at high pressures and temperatures. For the kyanite to andalusite reaction, it is $+7.39 \times 10^{-6}$ m^3 mol^{-1}.

If we are to reestablish equilibrium by raising the pressure at 800 °C, $\Delta G^{P}_{r,1073}$ will have to become zero (the condition for equilibrium). In that case,

$$\Delta G^{\text{o}}_{r,1073} = -\Delta V_r\left(P - 10^5\right)$$

from which it follows that *P* must be 0.75 GPa, a pressure in good agreement with the experimental data.

The slope of the univariant line at 800 °C and 0.75 GPa is given by $\Delta S^{P}_{r}/\Delta V^{P}_{r}$, according to the Clapeyron equation (Eq. (8.3)). As stated above, the volume change of the reaction at 298 K and 10^5 Pa can be used at high pressures and temperatures without introducing serious error. The value of ΔS^{o} can be calculated from the data in Table 7.1 for 800 °C and 10^5 Pa. According to Eq. (7.36), $(\partial S/\partial P)_T = -\alpha V$. But we are dealing with the difference in volumes (ΔV_r) between *solids* that have very small and similar values of α. We can therefore use the $\Delta S^{\text{o}}_{r,1073}$ for the high-pressure value (8.96 J mol^{-1} K^{-1}). The resulting slope of the univariant line at 800 °C and 0.75 GPa is

$$\left(\frac{\partial P}{\partial T}\right)_{\Delta G} = \frac{8.96}{7.39 \times 10^{-6}} = +1.2 \times 10^6 \quad \text{Pa K}^{-1}$$

which again is in good agreement with the experimentally determined slope. The reason that the thermodynamic calculation agrees so well with the experimental results is that the initial thermochemical data used for Table 7.1 were adjusted to conform with as many high-quality experimental studies as possible (Holland and Powell, 1998).

Univariant curves for reactions involving a gas are not as simply calculated as those involving condensed phases (solid or liquid), because ΔV_r and ΔS_r are strongly dependent on pressure and temperature. The free energy of a gas at high pressures can, however, be calculated easily from Eq. (7.32), which upon integration gives

$$\bar{G}^{P}_{T} = \bar{G}^{\text{o}}_{T} + \int_0^P \bar{V}\,\text{d}P \tag{8.4}$$

The free energy of the gas at temperature *T* and 10^5 Pa $(\bar{G}^{\text{o}}_{T})$ can be determined directly from the data in Table 7.1. Integration of the $\bar{V}\text{d}P$ term requires that we know the functional relation between *V* and *P*.

If the gas behaves ideally ($V = RT/P$), Eq. (8.4) reduces to

$$\bar{G}^{P}_{T} = \bar{G}^{\text{o}}_{T} + RT\ \ln\frac{P}{10^5} \tag{8.5}$$

where pressure is measured in pascal. If the pressure is measured in bars (1 bar = 10^5 Pa), Eq. (8.5) becomes

$$\bar{G}^{P}_{T} = \bar{G}^{\text{o}}_{T} + RT\ \ln P \quad \text{bars} \tag{8.6}$$

Thus, pressure is a measure of the free energy of an ideal gas.

Under the high pressures encountered in the Earth, most gases are not likely to behave ideally, so they will not obey Eq. (8.6). But the functional form of this equation is so convenient that it is preserved for real gases by inventing a function, *f*, known as the *fugacity*, such that

$$\bar{G}^{P}_{T} = \bar{G}^{\text{o}}_{T} + RT\ \ln f \tag{8.7}$$

The fugacity therefore measures the free energy of a real gas in the same way that pressure measures the free energy of an ideal gas. The fugacity of a gas is related to its pressure by the *fugacity coefficient*, γ, where

$$f = P\gamma \tag{8.8}$$

For an ideal gas $\gamma = 1$ and $f = P$. Tables of fugacities and fugacity coefficients for high pressures and temperatures are given for H_2O by Burnham *et al.* (1969) and for CO_2 by Mel'nik (1972).

For a reaction involving condensed phases and a gas, the free energy change can be obtained by integrating Eq. (8.4):

$$\Delta G^P_{r,T} = \Delta G^{\mathrm{o}}_{r,T} + \Delta V_{r,T,c}(P-1) + RT\ \ln f \tag{8.9}$$

where $\Delta V_{r,T,c}$ refers only to the volume change in the condensed phases; normally, the value of $\Delta V^{\mathrm{o}}_{298}$ can be used. The free-energy change resulting from the volume change of the gas is accounted for in the ($RT \ln f$) term, where the fugacity is given for the specified *P* and *T*. If we are interested in knowing at what pressure this reaction is at equilibrium, we set $\Delta G^P_{r,T}$ to zero, and Eq. (8.9) becomes

$$\Delta G^{\mathrm{o}}_{r,T} = -\Delta V_{r,T,c}(P-1) - RT\ \ln f \tag{8.10}$$

which can be solved for *P* by trial and error or by using, for example, the Solver function in Microsoft Excel (Problem 8.8). Note that *P* and *f* have units of bars (1 bar = 10^5 Pa = 10^5 J m^{-3}).

8.4 SCHREINEMAKERS RULES FOR INTERSECTING SURFACES IN *G–T–P* SPACE

So far, we have considered the location and orientation of only a single univariant reaction ($\alpha \rightarrow \beta$) in *G–T–P* space and its projection onto the *P–T* plane. The minerals involved in this reaction, however, might, under certain conditions of pressure and temperature, be involved in other reactions. More than one univariant line may, therefore, occur in the pressure and temperature range of interest. Because each line would most likely have a different slope, the lines will intersect at a point in the *P–T* diagram, and under these conditions of pressure and temperature, the assemblage of minerals represented by the two lines will have zero degrees of freedom. The possible ways in which univariant lines can intersect at an invariant point follow certain simple geometrical rules. These were first worked out by the Dutch physical chemist Schreinemakers and are thus known as the *Schreinemakers rules* (Zen, 1966).

A univariant line describing a reaction, such as the one illustrated in Figure 8.2, divides the *P–T* plane in two (Fig. 8.3). To the left of the line, phase *α* is more stable than phase *β*, because phase *α*'s free energy is lower than that of phase *β*. To the right of the line, phase *β* is more stable than phase *α*, because under these conditions phase *β* has the lower free energy. This condition would be equally true if *α* and *β* represented assemblages of minerals rather than single minerals. Also, note the deliberate use of the phrase "more stable than…" to indicate that we are discussing relative, not absolute, stabilities.

If a second univariant line were present, as would happen if *α*, for example, were able to transform into a third polymorph *γ*, the *P–T* plane would again be divided by this line, with *α* being more stable on one side of the line and *γ* being more stable on the other. Let us assume that this second reaction has a positive slope on the *P–T* diagram. The univariant line for the $\alpha \rightarrow \gamma$ reaction will intersect the univariant line for the β reaction as shown in Figure 8.4(A) and (B).

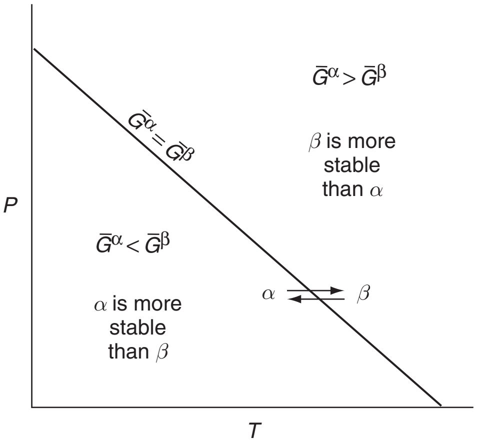


Fig. 8.3 Relative stabilities of phases *α* and *β* on either side of the line along which the two phases are in equilibrium because their free energies are the same.

However, for the reaction $\alpha \rightarrow \gamma$ to be stable, the free energies of *α* and *γ* must both be lower than that of *β*. This is satisfied only by the relations shown in Figure 8.4(A) . To the right of the *α* = *β* univariant line, *β* is more stable than *α*. So a reaction involving *α* on the right side of this line is said to be *metastable*. That is, the reaction can take place, but it does not involve the most stable phases under these conditions.

The distinction between metastable, stable, and unstable can be illustrated by likening a chemical reaction to the simple mechanical situation of a rectangular block resting on a surface (Fig. 8.5). When the block has its long axis parallel to the horizontal surface, the block's center of gravity is as low as it can possibly be. The potential energy of the block is therefore at a minimum, and this condition is described as *stable*. If, on the other hand, the block were resting with its long axis in a vertical position, the center of gravity would be higher than in the stable configuration. The potential energy would therefore be higher. By knocking the block over, the potential energy would be reduced. But energy is required to topple the block, because as the block tilts sideways, its center of gravity is raised slightly because of the pivoting action about the lower edge. The amount of energy necessary to raise the block during the pivoting is known as the *activation energy*; this amount of energy must be expended to topple the block. Once the center of gravity passes over the pivot point, both the activation energy introduced during the tilting and the initial excess potential energy are transformed to kinetic energy as the block falls. In this state the block is *unstable*. Had the activation energy not been introduced, however, the block could have remained in its vertical position indefinitely, even though in this position the block is not in its lowest-energy state. Under these conditions, the block is said to be *metastable*.

The univariant line *α* = *γ* in Figure 8.4 is metastable to the right of the univariant line *α* = *β*. This part of the line is therefore dashed. It will be noted that the line for the *α* = *β* reaction is dashed above the line for the *α* = *γ* reaction because, under these conditions, the β reaction must

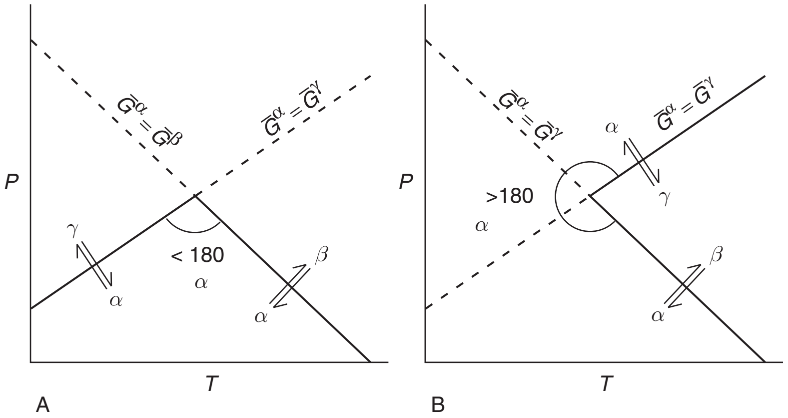


Fig. 8.4 Intersection of two univariant lines $\alpha = \beta$ and $\alpha = \gamma$. Because the $\alpha = \gamma$ reaction involves α, it must occur in a region where α is stable; that is, to the left of the $\alpha = \beta$ line. The configuration in (**B**) is therefore impossible. In (**A**), the univariant line $\alpha = \gamma$ is dashed to the right of the $\alpha = \beta$ line to indicate that here the reaction is metastable. The upper part of the $\alpha = \beta$ line is similarly dashed.

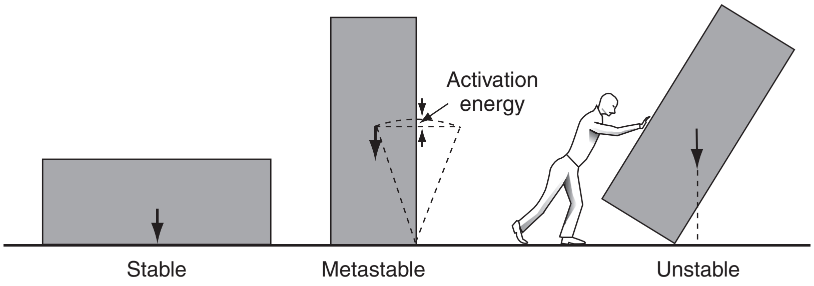


Fig. 8.5 Mechanical examples of stable, metastable, and unstable configurations of a block. The activation energy is the energy that has to be expended to make the metastable block become unstable.

also be metastable. It follows, then, that *the stable end of a univariant line must lie in the field in which the free energies of both the reactants and products are lower than the free energies of any other related phases in the system*.

At the pressure and temperature represented by the point of intersection of the two univariant curves in Figure 8.4, the system is invariant; that is, phases α, β, and γ can coexist only at this specific set of P–T conditions. Along the lines extending from this point of intersection, the system is univariant, with phases α and β or α and γ coexisting. Under conditions of pressure and temperature represented by the field between the univariant lines, the system is divariant, and only one phase exists (α in Fig. 8.4(A)). Because the stable end of a univariant line must be in a field where its reactants and products are more stable than any other phase (or phases), *the angle between univariant lines bounding a divariant field at an invariant point must be less than 180°*. The arrangement of univariant lines in Figure 8.6, for example, would be impossible because the α phase in both reactions would be made metastable by the other reaction.

Because the two univariant lines that intersect to give the invariant point represent coexisting α–β and α–γ, a third univariant line representing the coexistence of γ and β must also pass through the invariant point. Because this reaction involves γ, it must lie in the field where γ is stable; that is, within 180° in a clockwise direction from the stable end of the line $\alpha = \gamma$. But it also involves β and must therefore occur within 180° in a counterclockwise direction from the stable end of the line $\alpha = \beta$. These conditions are met in only one sector (Fig. 8.7). Note that the resulting divariant fields for β and γ both subtend angles of less than 180° around the invariant point.

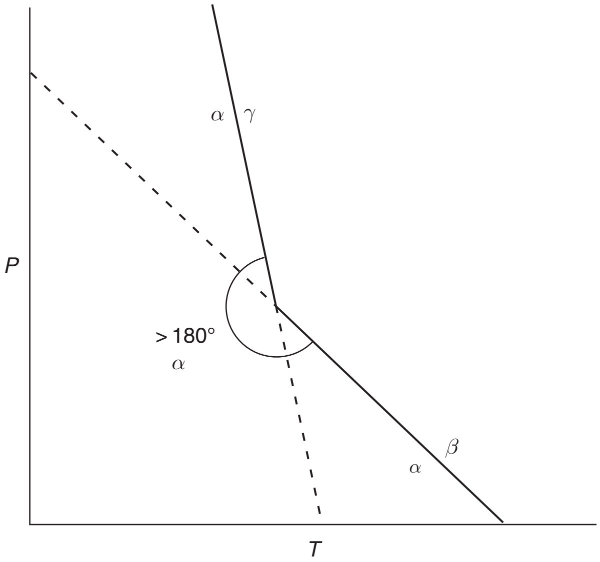


Fig. 8.6 Impossible arrangement of univariant lines, because both reactions make the phase α unstable in the other reaction. The sector in which phase α is stable must subtend an angle of less than 180° at the invariant point.

The set of univariant lines around the invariant point is now complete. Each line, while dividing the P–T plane into fields of different relative stabilities of the phases involved, is itself divided into stable and metastable parts by the other reactions. The various parts of the P T diagram can now be

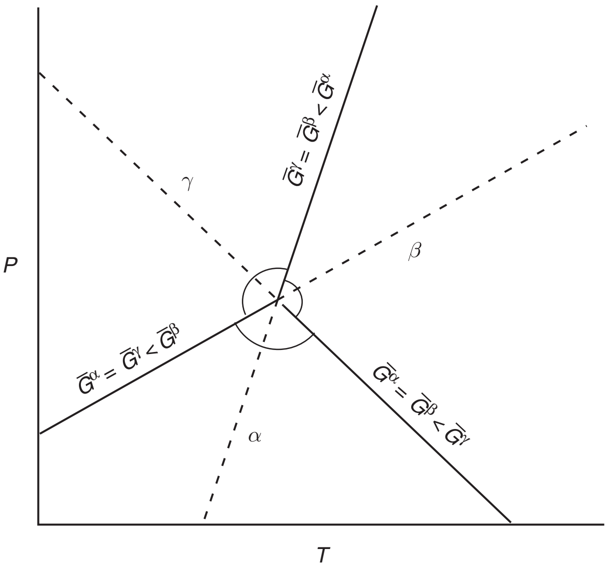


Fig. 8.7 Thermodynamically permissible arrangement of univariant lines separating the divariant fields of the phases α, β, and γ.

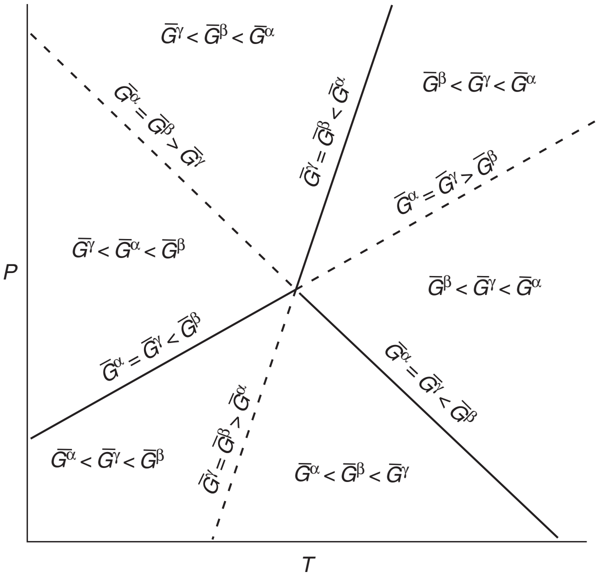


Fig. 8.8 Relative free energies of phases α, β, and γ in various parts of the P–T projection.

labeled in terms of the relative free energies of each of the phases (Fig. 8.8).

The relations between the free energies and the stability of the phases are most clearly seen in isothermal G–P plots (Fig. 8.9). At temperatures below the invariant temperature (T_{IP}), the trace of the free-energy surface of the β phase is above the point of intersection of the free-energy surfaces of the α and γ phases (Fig. 8.9(B)). Reactions with the β phase (1 and 2 in Fig. 8.9) are therefore metastable. At temperatures above the invariant point, the trace of the free-energy surface of the β phase lies below the intersection of the free-energy surfaces of the and phases (5 in Fig. 8.9(C)), making the $\alpha \rightarrow \gamma$ reaction metastable at these temperatures. The intersections of the α and β surfaces and the β and γ surfaces both become stable under these conditions. At the temperature of the invariant point (Fig. 8.9(D)), the free-energy surfaces for the three phases intersect at a point. Only under these conditions do all three phases have the same free energies; that is, they are equally stable (see Problem 8.1).

8.5 SCHREINEMAKERS RULES APPLIED TO MULTICOMPONENT SYSTEMS

In the example above, we considered a system consisting of three phases, α, β, and γ. The principles determining the arrangement of univariant lines about an invariant point, however, apply equally to systems containing more phases, the only difference being that as extra phases are added, the number of univariant lines increases. This number depends simply on how many independent reactions can be written between the minerals involved.

If we designate the number of *phases* (minerals, magmas, fluids) present by ϕ, and the *minimum* number of *components* necessary to describe the formulas of the phases by c, then the number of degrees of freedom (variance) of the system is given by

$$f = c - \phi + T + P + n \tag{8.11}$$

where n is the number of environmentally controlled (intensive) variables in addition to temperature (T) and pressure (P); the fugacity of water might, for example, play such a role. If temperature and pressure are the only intensive variables, Eq. (8.11) is commonly written as

$$f = c + 2 - \phi \tag{8.12}$$

where it is understood that the 2 refers to temperature and pressure. This relation, which is known as *Gibbs phase rule*, is developed more fully in Section 10.5.

If a system consists of $(c+2)$ phases, the variance will be zero. In a P–T diagram, $(c+2)$ univariant lines radiate from the invariant point, each line involving $(c+1)$ phases. These univariant lines divide P–T space into $(c+2)$ divariant fields. One or more assemblages of c phases occur in each field; one of these assemblages is unique to each field. Thus in the example worked out in Figure 8.7, the formulas of the phases α, β, and γ (polymorphs of silica, for example) can all be expressed in terms of one component (SiO_2). At the invariant point, three $(c+2)$ phases (α, β, and γ) coexist. If any one of the phases is absent, the system is univariant. For example, if γ is absent, the system consists of α and β, and the P–T conditions are given by the univariant line $\alpha = \beta$. It is convenient to describe each of the univariant lines by the phase that is absent. According to the symbolism used by Schreinemakers, this absent phase is placed between parentheses. Thus, the $\alpha = \beta$ reaction is designated by (γ). Similarly, the univariant lines for the $\alpha = \gamma$ and $\beta = \gamma$ reactions are designated by (β) and ( ), respectively.

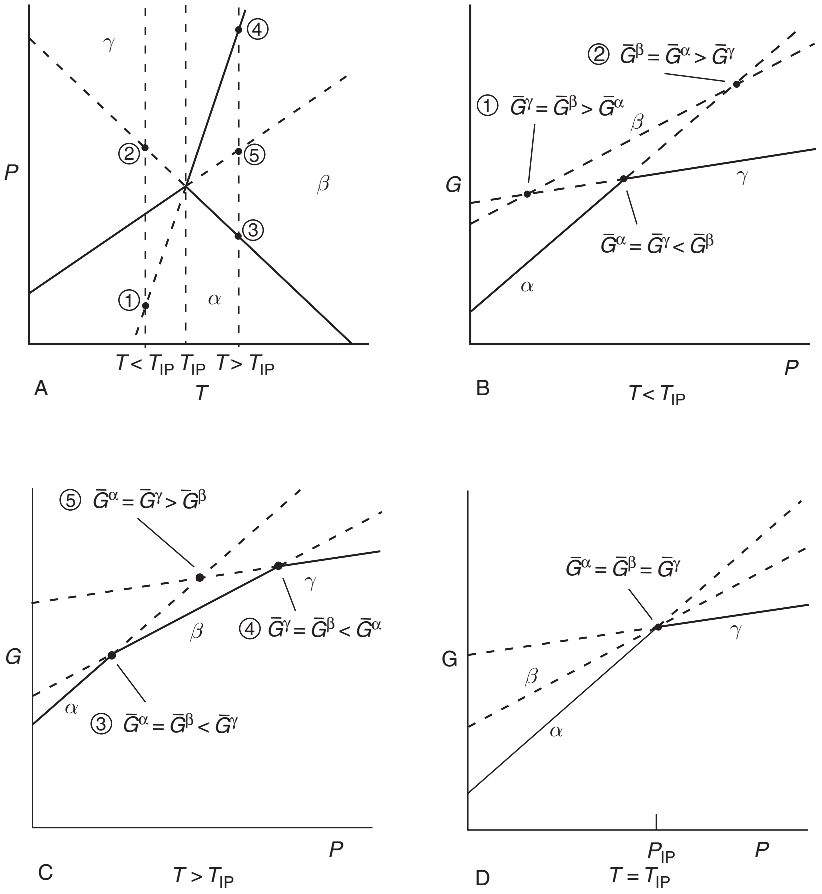


Fig. 8.9 Isothermal sections through *G*–*T*–*P* space showing the relative positions of the $\alpha = \beta$, $\alpha = \gamma$, and $\beta = \gamma$ univariant lines at a temperature below the invariant point (**B**), above the invariant point (**C**), and at the invariant point (**D**). The relative positions of these sections are shown in the *P*–*T* plot (**A**).

One reason for using this method of labeling reactions is that it leads to a simple mnemonic scheme for working out the sequence of univariant lines around an invariant point without having to resort to basic principles each time. We start by writing each reaction as a simple equation. The equation is then rewritten, placing each phase in parentheses and replacing the equals sign with the phase absent from that reaction. Thus, for the one-component system with phases α, β, and γ, we write the three reactions as

$$\begin{array}{ccc} \alpha = \beta & \alpha = \gamma & \beta = \gamma \\ (\alpha)|(\gamma)|(\beta) & (\alpha)|(\beta)|(\gamma) & (\beta)|(\alpha)|(\gamma) \end{array}$$

The phases in parentheses (phase absent) now refer to reactions. From the mnemonic for the first equation, we read that the stable part of the (α) reaction ($\beta = \gamma$) lies on one side of the univariant line (γ) and the (β) reaction lies on the other. Thus, either of the arrangements of lines in Figure 8.10(A) or (B) would satisfy this requirement. According to the mnemonic from the second equation, the (α) reaction must also lie on the opposite side of the (β) line from the (γ) reaction. Thus, if we choose the option in Figure 8.10(A), this condition can be satisfied only by making the high-pressure end of the (γ) reaction metastable (Fig. 8.10(C)). If the option in Figure 8.10(B) is chosen, the only way that reactions (α) and (γ) can be on opposite sides of reaction (β) is for the low-pressure end of the (γ) reaction to be metastable. The univariant lines are now all in their correct sectors, and the information from the third reaction, that is, that the (β) and (γ) univariant lines must be on opposite sides of the (α) reaction, serves simply to confirm that no mistakes were made in orienting the lines with the information from the first two equations.

This mnemonic scheme leads to two possible arrangements of univariant lines that are reverse images of each other. Both are thermodynamically consistent, and without additional information, selection of the correct one might be difficult. Of course, if thermodynamic data for the phases are available, the decision is simple. But even when this is lacking, knowledge of the field occurrence of the phases can provide the answer. If, for example, the mineral β were known to occur only in high-temperature rocks, the scheme in Figure 8.10(C) would be the correct one. Similarly, if one of the phases is a liquid or a gas, it must appear on the high-temperature side of a reaction with a solid.

The actual slopes of the univariant lines cannot be determined from the mnemonic scheme; this requires information on the ΔS and ΔV of the reactions or experimental data. Commonly, however, the slope of at least one of the reactions is known, and then the scheme provides some restrictions on the possible slopes of the remaining univariant lines. Two reactions must, of course, be known to determine the position of an invariant point in P T

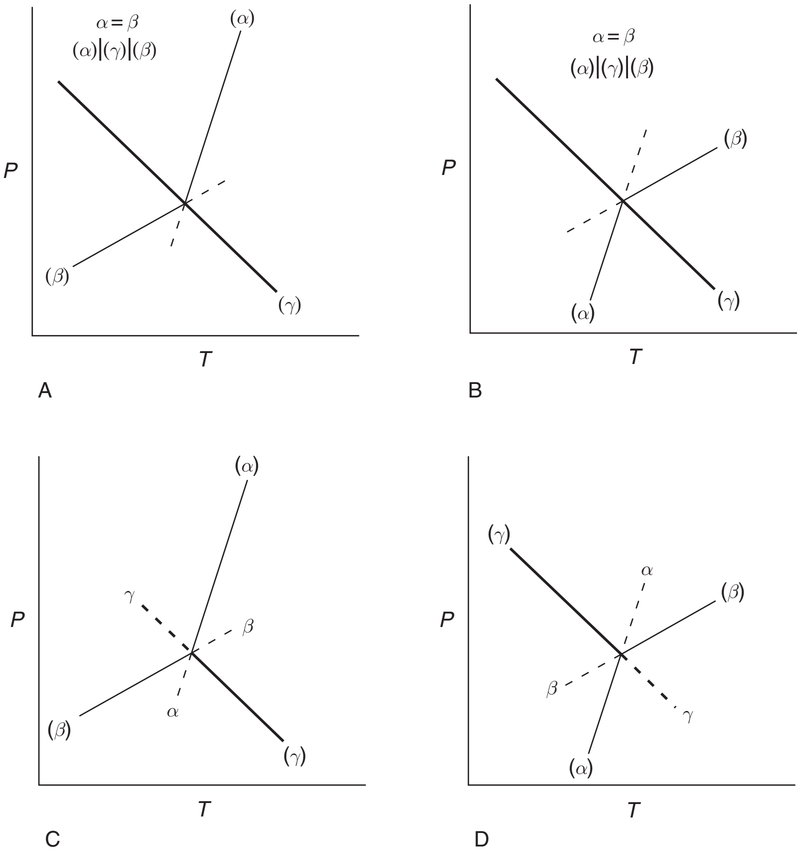


Fig. 8.10 Sequence of steps in arranging univariant lines around an invariant point in a one-component system. The line $\alpha = \beta$, designated by the phase absent (γ), is taken as given. The remaining two lines must plot on opposite sides of (γ) as shown in (**A**) or (**B**). Part of the (γ) reaction then becomes metastable as shown in (**C**) or (**D**).

Although the example worked out above is for a one-component system, the mnemonic scheme works equally well for systems containing more components. Consider, for example, the two-component system MgO–SiO$_2$. Common minerals in this system include periclase (MgO), forsterite (Mg_2SiO_4), enstatite ($MgSiO_3$), and quartz (SiO_2). Because the system has two components, it would be invariant if all four minerals were present together ($f = 0 = c+2-\phi$). Radiating from this invariant point would be four ($c+2$) univariant lines, each representing the coexistence of three minerals ($c+1$).

The compositions of the minerals and the reactions between them can be graphically represented by a simple line that at one end is pure MgO and at the other is pure SiO$_2$ (Fig. 8.11(E)). Enstatite ($1MgO + 1SiO_2$), for example, would plot halfway along this line, whereas forsterite ($2MgO + 1SiO_2$) would plot only one-third of the way toward SiO$_2$. Reactions between the minerals can be determined simply by inspecting the relative positions of the minerals on this compositional line. A reaction product must lie between the reactants. For example, periclase and enstatite could react to form forsterite. The four possible univariant reactions between periclase, forsterite, enstatite, and quartz and their mnemonic schemes are as follows:

1. MgO Periclase (Pe)	+	$MgSiO_3$ Enstatite (En)	= |Q|	Mg_2SiO_4 Forsterite (Fo)
2. Mg_2SiO_4 Forsterite (Fo)	+	SiO_2 Quartz (Q)	= |Pe|	$2MgSiO_3$ Enstatite (En)
3. MgO Periclase (Pe)	+	SiO_2 Quartz (Q)	= |Fo|	$MgSiO_3$ Enstatite (En)
4. $2MgO$ Periclase (Pe)	+	SiO_2 Quartz (Q)	= |En|	Mg_2SiO_4 Forsterite (Fo)

Thermodynamic data indicate that all of these reactions have positive slopes in *P*–*T* space and that quartz and periclase are stable on the high-temperature side of the (Fo) reaction (Fig. 8.11(A)). According to the mnemonic scheme for the first equation, the (Fo) reaction lies on the opposite side of the (Q) reaction from the univariant lines for (Pe) and (En). Starting with the (Fo) line as given, the (Q) line is drawn with a positive slope; it intersects the (Fo) line to create the invariant point (Fig. 8.11(A)). Because the (Q) reaction involves enstatite, the reaction is metastable below the (Fo) reaction. Similarly, because the (Fo) reaction involves enstatite and periclase, it must be metastable to the right of the (Q) reaction. The univariant lines for (Pe) and (En) can, according to the first reaction, be placed anywhere on the right side of the (Q) reaction. However, according to the second reaction, the (Fo) and (Q) lines must lie on one side of the (Pe) line and (En) on the other. Also, because the (Pe) reaction involves enstatite, only that part above the (Fo) line can be stable. The (Pe) line can therefore be placed only in the sector shown in

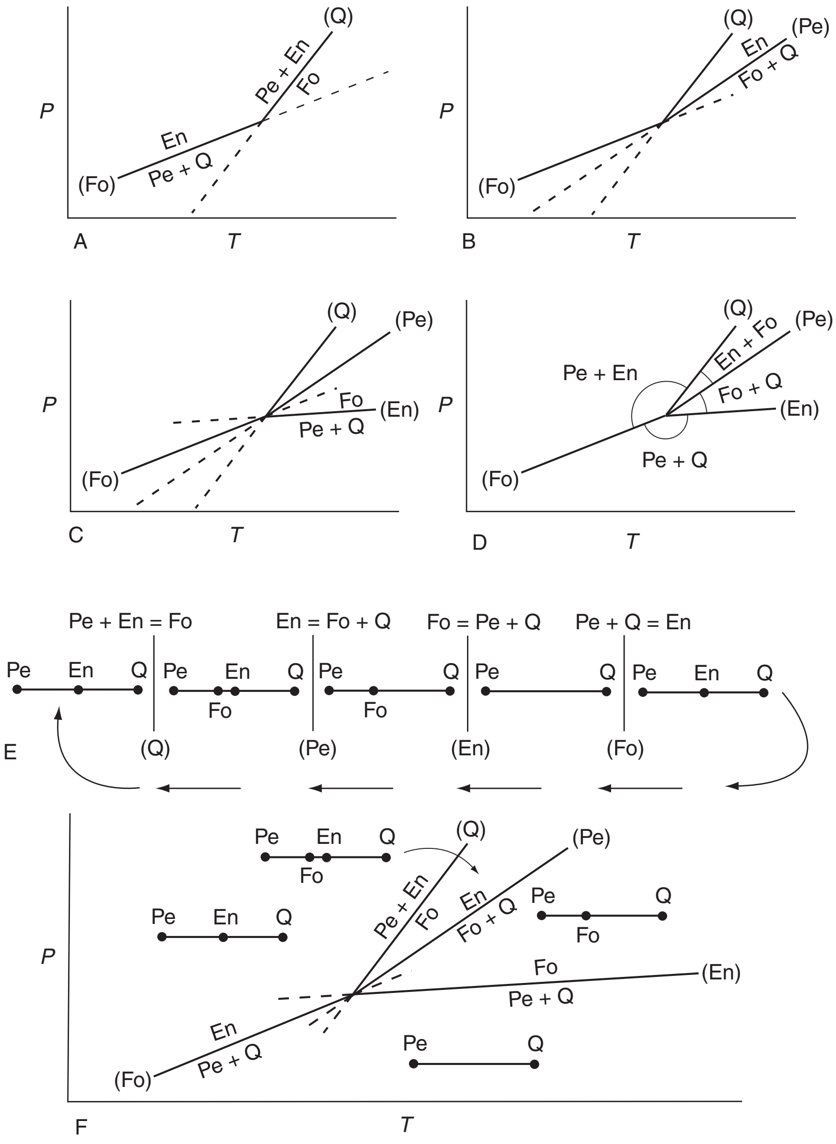


Fig. 8.11 Sequence of steps involved in arranging the univariant lines around the invariant point involving periclase (Pe), forsterite (Fo), enstatite (En), and quartz (Q) in the two-component system MgO–SiO$_2$. See text for explanation.

Figure 8.11(B). The third reaction requires that the (Pe) and (Q) lines lie on one side of the (Fo) line and (En) on the other. This can be satisfied only by placing the (En) line as shown in Figure 8.11(C). The diagram is now complete but can be checked by examining the fourth reaction. This indicates that the (Pe) and (Q) lines should lie on the opposite side of the (En) reaction from the (Fo) line, which indeed they do.

The four univariant lines in Figure 8.11(D) separate four divariant fields, each characterized by assemblages of two (c) phases. These fields each subtend angles of less than 180° at the invariant point. Some of these assemblages, such as that of Pe + Q in the lower sector, can be read directly from the phases involved in the reactions. Others, such as that of Pe + Fo, must be deduced. Consider the assemblage Pe + En being heated until it crosses the (Q) reaction to form Fo (Fig. 8.11(A)). If the initial assemblage contained an excess of En over that which was needed for the reaction, the reaction products would be Fo plus the excess En. Thus, En is stable to the right of the (Q) reaction as long as no Pe is present. Continued heating of the assemblage Fo + En would eventually result in the breakdown of En to Fo + Q when it crosses the (Pe) reaction (Fig. 8.11(B)). The assemblage En + Fo is therefore restricted to the sector between the (Q) and (Pe) univariant lines.

Divariant fields marking the stability limits of a single phase are also given by this diagram. Enstatite, for example (Fig. 8.11(D)), is stable only between the univariant lines (Fo) and (Pe). Here again, the sector in which a single phase is stable cannot exceed 180°.

The reactions given by the univariant lines radiating from the invariant point can be represented graphically as shown in Figure 8.11(E). We will start with the assemblage in the upper left of the diagram and proceed in a clockwise direction around the invariant point. The phases stable in the sector bounded by the (Q) and (Fo) univariant lines are Pe + En and En + Q. We see that the latter assemblage is possible by considering what the reaction products would be if a mixture of Pe + Q, containing an excess of Q, was raised above the

(Fo) reaction. Clearly En + Q would remain. The first compositional line in Figure 8.11(E) therefore has the minerals Pe, En, and Q marked on it. If a rock has a bulk composition that lies between En and Q, it would consist of the minerals En and Q. A rock with a composition between En and Pe would, on the other hand, consist of En and Pe.

On crossing the (Q) univariant line, Pe reacts with En to form Fo. Thus, on the compositional line, the mineral Fo is added. Now rocks having bulk compositions between En and Pe will consist of Pe + Fo or Fo + En, depending on whether they fall to the left or right of the Fo composition.

Continuing in a clockwise direction, the (Pe) reaction is encountered next, and En breaks down to form Fo + Q. The disappearance of En from the compositional line means that rocks of appropriate composition will contain Fo + Q. Note that the assemblage Pe + Fo is still stable in MgO-rich rocks.

Next, Fo breaks down to form Pe + Q on crossing the (En) reaction, leaving only Pe and Q as stable minerals on the composition line. Any composition in this system will therefore consist of some mixture of Pe and Q.

Finally, on crossing the (Fo) reaction, Pe and Q react to form En, and we then have the assemblage with which we started. It is important to note that each one of these steps involves changing only one mineral at a time. If you find that two minerals have been changed in going from one sector to the next, a mistake has been made in positioning one or more of the univariant lines.

We are now in a position to label all of the univariant lines and divariant fields with the appropriate mineral assemblages (Fig. 8.11(F)) to complete the *P–T* diagram. Such diagrams provide a convenient means of analyzing the significance of mineral assemblages found in rocks. Most rocks will fall in divariant fields in *P–T* diagrams, because they typically have existed over a range of pressures and temperatures. As a result, most rocks will contain *c* phases, which in this example is two. Rocks containing more than *c* phases must lie on univariant lines or at invariant points, or be out of equilibrium.

Univariant and invariant assemblages are not as uncommon as might at first be expected. If a change in intensive variables causes a divariant assemblage to reach a univariant reaction, the enthalpy change of the reaction may be large enough to buffer and control the intensive variables. For example, by analogy, it is not an unusual occurrence to find water boiling in a pot in a kitchen, yet the univariant reaction water → steam is actually invariant under the ambient pressure of 10^5 Pa; that is, you know that the temperature of the water is 100 °C if it is boiling. A stove may transmit considerable energy to the water, but this heat is used in vaporization rather than in raising temperature. Thus the boiling buffers the reaction at 100 °C as long as there is water to boil. In a similar way, reactions, such as the partial melting of the upper mantle or the growth of a metamorphic mineral in a contact aureole of an igneous intrusion, can produce univariant or invariant assemblages that will, through their enthalpies of reaction, maintain the intensive variables at the special values of univariant or invariant equilibria, as long as reactants remain.

As the number of components increases, so does the number of univariant lines radiating from an invariant point. The simple mnemonic scheme outlined above, however, can still be used to determine the sequence of univariant lines. Let us consider the system $MgO–Al_2O_3–SiO_2$. Because it contains three components, an assemblage of five minerals ($c + 2$) in this system will be invariant, and radiating from an invariant point will be five univariant lines. This system contains many minerals, any group of five of which will produce an invariant assemblage. We will consider the invariant point associated with the five minerals andalusite (An), corundum (Co), spinel (Sp), enstatite (En), and cordierite (Cd).

Because the system contains three components, the compositions of the minerals can be expressed conveniently in a triangular plot where each apex of the triangle represents one of the three components (Fig. 8.12). From such a plot, the reactions between the minerals can easily be determined. First, draw a set of lines joining the compositions of each of the minerals, but no two lines may intersect. The set of lines shown in triangle I in Figure 8.12 is one possible way of drawing these lines, but any of the arrangements shown in triangles I to V would be acceptable. These lines divide a triangle into a number of smaller triangles. A rock with a bulk composition plotting within any one of these subtriangles consists of the minerals at the apexes of the subtriangle. For example, a rock composed of 40% SiO_2, 30% MgO, and 30% Al_2O_3 (x in triangle I) would consist of Co + En + Cd.

Next, remove any one line and replace it with another line joining a different pair of minerals. For example, the line joining En and Co can be replaced with one joining Cd and Sp (II in Fig. 8.12). These switching lines constitute a reaction, which in this case is En + Co = Cd + Sp. Because this reaction does not involve andalusite, it is referred to as the andalusite absent reaction or (An). Note that when this one line is switched, the remainder of the lines do not change; that is, only one line is changed at a time. Continuing in this manner, the remainder of the reactions are easily determined (Fig. 8.12 (II) to (V)).

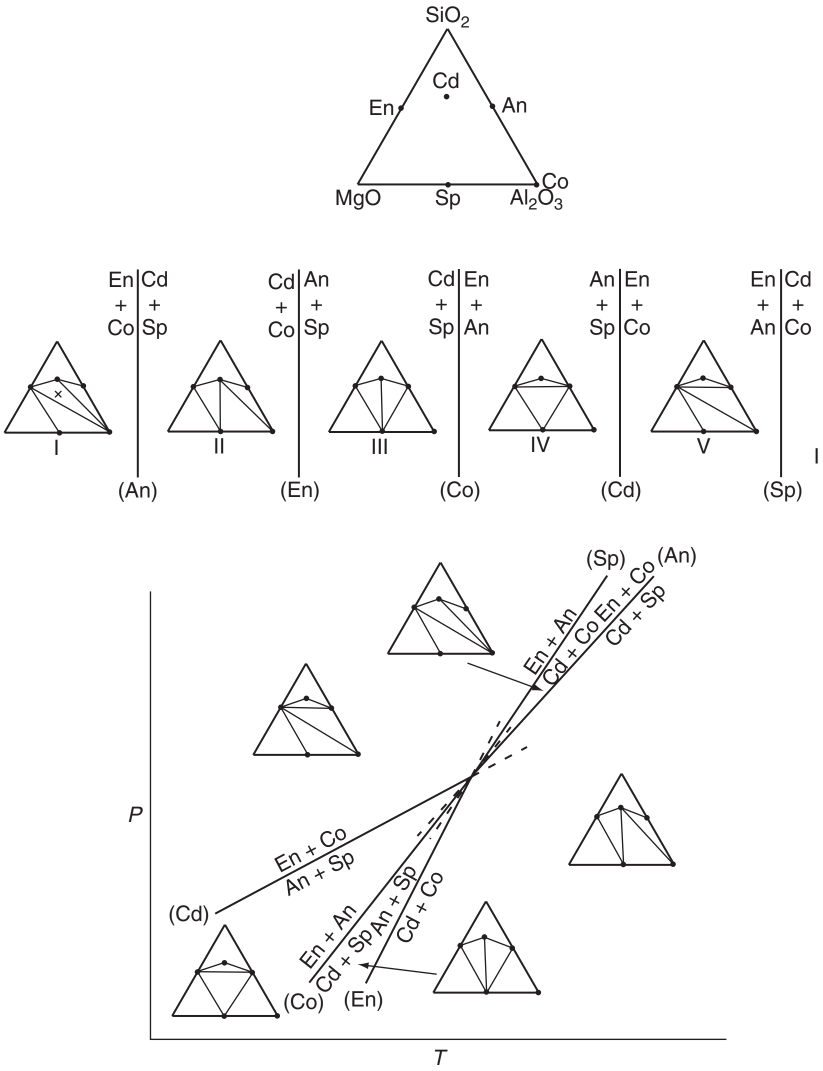
We now write out the balanced reactions and the mnemonic scheme for each of the five univariant lines radiating from the invariant point (Table 8.2). From the ΔS and ΔV of each reaction, the slopes of the univariant lines are determined. The ΔS of the reaction also indicates which assemblage is favored at high temperature. For example, with increasing temperature, reactions (An) and (Sp) proceed to the right, but the remainder proceed to the left.

Start by placing the (An) reaction on the *P–T* diagram with its appropriate slope (it is instructive to draw this yourself and then compare it with Fig. 8.12). From the mnemonic, we know that the (En) and (Co) lines must plot on one side and the (Cd) and (Sp) lines on the other side of the (An) reaction. Two complementary diagrams could be drawn if we did not know that the (An) reaction proceeds to the right with increasing temperature; that is, Cd and Sp are stable together on the high-temperature side of the (An) line. Thus the (Cd) and (Sp) lines must be placed on the low-temperature side of the (An) line, and (En) and (Co) on the high-temperature side. These lines can be plotted temporarily anywhere on each side

Table 8.2

Phase absent	Reaction	Mnemonic	ΔV ($m^3 \times 10^3$)	ΔS ($J\ mol^{-1}\ K^{-1}$)	dP/dT ($Pa\ K^{-1} \times 10^{-5}$)
(An)	5En + 5Co = Cd + 3Sp	(En)(Co)|(An)|(Cd)(Sp)	0.068	55.42	8.1
(En)	Cd + 5Co = 5An + 2Sp	(Cd)(Co)|(En)|(An)(Sp)	−0.024	−36.37	15.2
(En)	2Cd + Sp = 5En + 5An	(Cd)(Sp)|(Co)|(En)(An)	−0.092	−91.79	10.0
(Cd)	An + Sp = En + 2Co	(An)(Sp)|(Cd)|(En)(Co)	−0.009	−3.81	4.2
(Sp)	2En + 3An = Co + Cd	(En)(An)|(Sp)|(Co)(Cd)	0.042	43.99	10.5

Fig. 8.12 Construction of a *P–T* diagram for the univariant lines at the invariant point involving enstatite (En), cordierite (Cd), andalusite (An), corundum (Co), and spinel (Sp) in the three-component system MgO–Al_2O_3–SiO_2. See text for explanation.



of the (An) line. A short segment of the metastable extension of each line should be plotted beyond the invariant point; this will help in positioning the remainder of the univariant lines.

Next, consider the (En) reaction. According to the mnemonic, the (Cd) and (Co) reactions must lie on one side of the (En) line and the (An) and (Sp) reactions on the other. Depending on how these lines were drawn on either side of the (An) line in the first step, they may need to be switched or adjusted to be consistent with the mnemonic for the second reaction. Continuing in this manner, then, the positions of the univariant lines are adjusted until they are consistent with the mnemonic for each reaction. This leads to the unique solution shown in Figure 8.12.

Once the univariant lines have been positioned in their correct sectors the slopes of the lines can be adjusted to agree with the slopes determined from the thermodynamic data. This should not move a univariant line into a new sector unless a mistake has been made in the positioning of the univariant lines or if there is an error in the thermodynamic data. Each of the lines can now be labeled with the appropriate reaction. Care must be taken to get the correct assemblages on each side of each line. In doing this, it is useful to recall that the sector subtended at the invariant point by a phase assemblage cannot exceed 180°.

A final check on the correctness of the scheme of univariant lines can be obtained by constructing the compositional triangles in each divariant field based on the labeling of the univariant lines bounding the fields. Lines can be drawn between any pair of phases that a reaction indicates is stable in the particular sector. In the lower right sector of Figure 8.12, for example, Cd + Sp are stable together and can therefore be joined with a line. But this assemblage could also include Co, if Co had been present in excess when the (An) reaction formed the Cd + Sp. Consequently, Cd, Sp, and Co can all be joined with lines. Had En been in excess instead of Co, then Cd + Sp + En would also be stable. These phases, therefore, can also be joined with lines. The resulting compositional diagrams in each sector differ by only one line from the diagram in the adjoining sector. If they do not, univariant lines have been incorrectly positioned or labeled.

8.6 DEGENERATE SYSTEMS

In general, as many univariant lines radiate from an invariant point as there are phases in the system; this will be two more than the number of components. In some systems, however, reactions between certain phases may be described in terms of a smaller number of components than are needed to describe the other reactions being considered. In such cases, the number of univariant lines decreases and the system is said to be *degenerate*.

A system becomes degenerate in two ways. First, two or more phases may have the same composition (polymorphs, for example). Second, in systems consisting of three or more components, three or more phases may be collinear or four or more phases may be coplanar in ternary or quaternary compositional diagrams, respectively. The result is that some reactions will involve fewer than $c + 1$ phases. Those phases taking part in reactions involving fewer than $c + 1$ phases are termed *singular phases*, and the remainder are called *indifferent phases*. Two types of degeneracy occur depending on whether the indifferent phases plot on opposite or the same sides of the line joining the singular phases. This can be illustrated by considering a specific example.

The invariant point involving corundum, periclase, spinel, cordierite, and quartz in the system $MgO–Al_2O_3–SiO_2$ is degenerate for two reasons. First quartz, cordierite, and spinel plot on a straight line in the compositional diagram; second, periclase, spinel, and corundum also plot on a straight line (Fig. 8.13). In the first case Q, Cd, and Sp are considered the singular phases, whereas Pe and Co, which plot on opposite sides of the collinear group, are the indifferent phases. In the second case, Pe, Sp, and Co are the singular phases, and the indifferent ones are Q and Cd, which both plot on the same side of the collinear group.

As before, reactions between phases can be determined by inspection of the compositional triangle (Fig. 8.13). However, unlike the previous example, where five different reactions were obtained, only three occur at this invariant point. They are

(Co)(Pe)	Cd = 5Q + 2Sp
(Q)(Cd)	Sp = Co + Pe
(Sp)	5Q + 2Pe + 2Co = Cd

the mnemonics for which are

(Cd)	|(Co)(Pe)|	(Q)(Sp)
(Sp)	|(Q)(Cd)|	(Co)(Pe)
(Q)(Pe)(Co)	|(Sp)|	(Cd)

The first reaction does not involve Co or Pe; so it can be referred to as the (Co) or (Pe) line. Similarly the second reaction is both the (Q) and (Cd) reaction.

We will start with the (Sp) reaction on the *P–T* diagram. According to the mnemonic, the (Q), (Pe), and (Co) reactions lie on one side and the (Cd) reaction on the other side of the (Sp) line. Because Cd is stable at lower temperatures than Q + Pe + Co, the (Cd) reaction must lie on the high-temperature side of the (Sp) line and (Q), (Pe), and (Co) on the low-temperature side. But (Q) and (Cd) are the same reaction. Therefore, the reaction of Sp = Co + Pe passes straight through the invariant point with no change in slope. On the high-temperature side it is labeled (Cd) and on the low-temperature side, (Q).

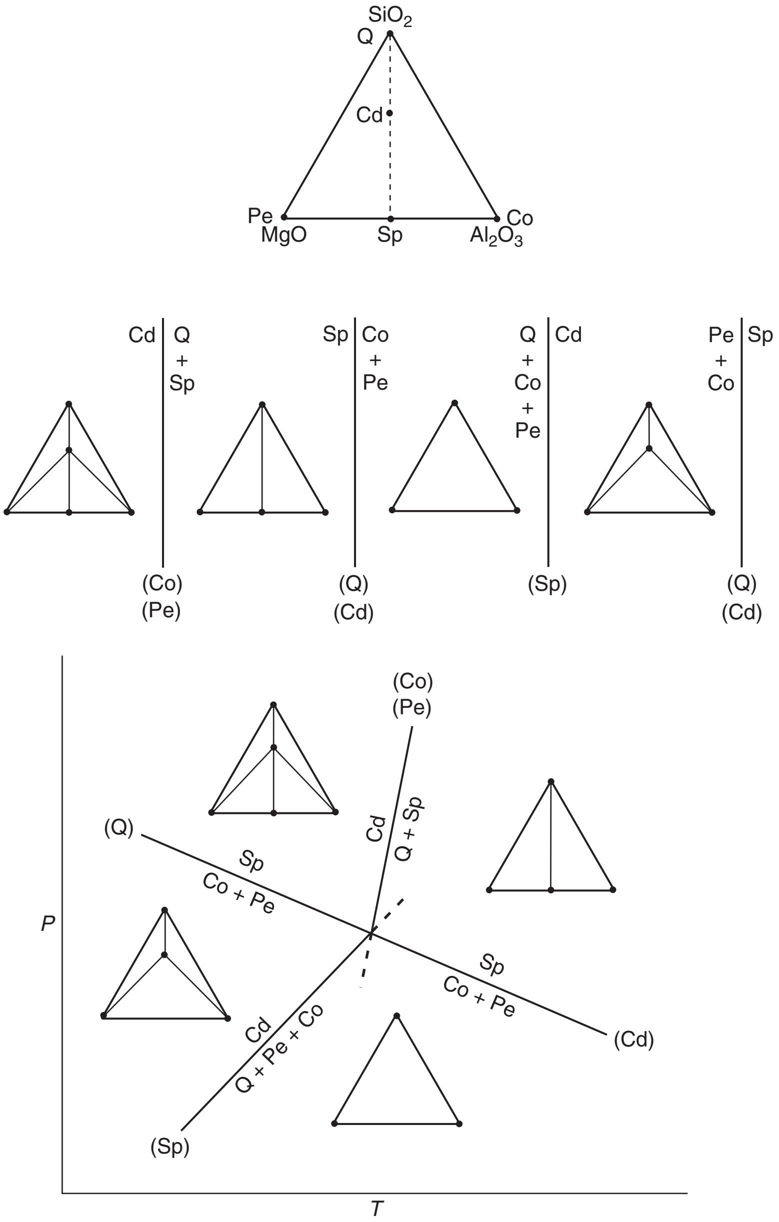
The mnemonic for the second reaction indicates that the (Sp) reaction lies on one side of the (Q) or (Cd) reaction and the (Co) and (Pe) reactions on the other. But the (Co) and (Pe) reactions are the same one; thus they must plot as a single line. Also, from the third reaction we see that the (Co) and (Pe) reactions must plot on the same side of the (Sp) reaction as the (Q) reaction. The univariant lines for (Co) and (Pe) therefore plot as a single line as shown in Figure. 8.13.

We see, then, that if the indifferent phases lie on the same side of the collinear or singular phases, a single univariant line passes straight through the invariant point ((Q) and (Cd) in Fig. 8.13). This degeneracy does not lead to a decrease in the number of univariant lines, but two of the lines will have identical slopes. If the indifferent phases lie on opposite sides of the singular phases, two univariant lines become one. This therefore leads to a decrease in the number of univariant lines about the invariant point.

8.7 SUMMARY AND CONCLUSIONS

We have seen that the Gibbs free energy of a phase (or group of phases) varies as a function of temperature ($-S$) and pressure (V). Because different phases are not likely to have identical molar entropies and volumes, free energy surfaces will not be parallel in *G–T–P* space. Where these surfaces intersect, the free energies of the two will be identical. Thus, if the phases are related by a reaction, the line of intersection of the free energy surfaces defines the pressure and temperature

Fig. 8.13 Construction of a *P–T* diagram for the univariant lines at the invariant point involving quartz (Q), cordierite (Cd), corundum (Co), spinel (Sp), and periclase (Pe) in the three-component system MgO–Al_2O_3–SiO_2. See text for explanation.



conditions under which the reactants and products are at equilibrium ($\Delta G = 0$).

The line of intersection of two free-energy surfaces is said to be univariant, because only one of the two intensive variables, T or P, can be independently varied. If a second univariant line is present involving the same components, the system is invariant at the point of intersection of the two lines. Only under this one set of *P–T* conditions are the assemblages of phases from both univariant lines stable together.

The position and slope of univariant lines in *P–T* space can be calculated from the thermodynamic data for the phases. In addition, Schreinemakers rules provide a simple geometrical means of correctly orienting univariant lines about an invariant point. The univariant lines divide *P–T* space into divariant fields; that is, regions in which both P and T can vary independently without modifying the mineral assemblage present in that field. In nature, rocks may form under conditions corresponding to those of divariant fields, univariant lines, or, under unusual conditions, invariant points. *P–T* diagrams therefore provide an invaluable tool for the analysis of the mineral assemblages found in rocks, and they will be used extensively throughout the remainder of this book.

8.8 PROBLEMS

8.1 For the system illustrated in Figure 8.9, draw three isobaric (constant pressure) G versus T diagrams for conditions where $P < P_{IP}$, $P > P_{IP}$, and $P = P_{IP}$, where P_{IP} is the pressure at the invariant point. Label all isobaric invariant points and indicate which parts of each diagram are stable and which are metastable.

8.2 The melting of a pure mineral can be written as a simple reaction. For diopside it is

$$\text{diopside}_{\text{crystal}} = \text{diopside}_{\text{liquid}}$$

At 10^5 Pa (1 bar), diopside melts at 1665 K. If the S^{o}_{1665} of crystalline and liquid diopside are 532.2 and 619.6 J mol^{-1} K^{-1}, respectively, and their volumes are, respectively, 0.06609×10^{-3} and 0.07609×10^{-3} m^3 mol^{-1}, calculate the melting point at 2 GPa (20 kbar) using the Clapeyron equation. (Recall that 1 J = 1 Pa m^3.)

8.3 The Al_2SiO_5 polymorphs andalusite, kyanite, and sillimanite can coexist at an invariant point at 500 °C and 0.376 GPa.

(a) Calculate the slopes of the three univariant lines radiating from the invariant point using values of S^{o}_{298} and V^{o}_{298} from Table 7.1.

(b) Plot the univariant lines in a *P–T* diagram (200 to 800 °C and 0 to 1 GPa).

(c) Compare your results with the experimentally determined values of Holdaway (1971), which are fitted by the equations

$$K = A \qquad P = 13.3(T-200) - 0.0026(T-200)^2$$
$$A = S \qquad P = 14.0(770 - T)$$
$$K = S \qquad P = 20.0(T-315) + 0.0009(T-315)^2$$

where *P* is in bars and *T* in °C.

(d) Discuss the possible causes for the differences.

8.4 The mineral jadeite, which is known to occur in metamorphic rocks formed at high pressures, can be related to two common low-pressure minerals by the reaction

$$\text{albite} = \text{jadeite} + \text{quartz}$$

(a) Balance the reaction, and using data in Table 7.1, calculate the ΔG of reaction at 298 K and 10^5 Pa. From the sign of $\Delta G^{\text{o}}_{r,298}$, comment on the feasibility of the reaction under these conditions.

(b) Repeat part (a), but for 600 °C and 10^5 Pa.

(c) Calculate what pressure increases would be needed at 298 K and 873 K to bring the reaction to equilibrium.

(d) How does the slope of this univariant line compare with the slope obtained from the Clapeyron equation and data for 298 K and 10^5 Pa?

(e) Compare your results with those obtained experimentally by Holland (1980). Plot Holland's results on your *P–T* diagram using his empirical equation for the univariant line.

8.5 At high pressure in the two-component system $NaAlSiO_4$–SiO_2, the four minerals nepheline, jadeite, albite, and quartz coexist at an invariant point.

(a) Write balanced reactions for the univariant lines that radiate from this invariant point.

(b) Using the S^{o}_{298} and V^{o}_{298} values from Table 7.1, calculate the slopes of these univariant lines.

(c) Using the mnemonic scheme and Schreinemakers rules, construct a *P–T* diagram for this set of univariant lines. While not changing the sectors in which the univariant lines plot, adjust the slopes of the lines to agree as closely as possible with the slopes determined from the Clapeyron equation.

8.6 Construct a *P–T* diagram for the set of univariant lines associated with the invariant point for the assemblage periclase, andalusite, spinel, enstatite, and cordierite in the system MgO–Al_2O_3–SiO_2. Use the values of S^{o}_{298} and V^{o}_{298} from Table 7.1 to determine the slopes of the lines. Note that the slopes of some of the lines calculated with the data for 298 K must be wrong, because they are not consistent with the arrangement of univariant lines required by Schreinemakers rules.

8.7 The invariant assemblage quartz, enstatite, sillimanite, cordierite, and sapphirine ($Mg_2Al_4SiO_{10}$) in the system MgO–Al_2O_3–SiO_2 is degenerate.

(a) Write balanced reactions for all univariant lines at this invariant point.

(b) Determine which phases are singular and which are indifferent. How many invariant lines do you expect around the invariant point?

(c) Determine the slope of the sapphirine-absent reaction from data in Table 7.1.

(d) Construct a *P–T* diagram for this set of univariant lines. Label all reactions, and draw triangular diagrams indicating the stable mineral assemblages in each divariant field.

8.8 Muscovite and quartz react to form sillimanite, potassium feldspar, and steam at high temperatures.

(a) From the data in Table 7.1, calculate the ΔG^{o}_r for this reaction at a range of temperatures from 400 K to 800 K, and plot the value of ΔG^{o}_r against *T*. Determine the equilibrium temperature for this reaction at 10^5 Pa. (This is most easily done using a spreadsheet.)

(b) Assuming steam behaves ideally, calculate the equilibrium pressure for this reaction at 600 K, 650 K, 700 K, 750 K, and 800 K using Eq. (8.10). This equation has to be solved by trial and error; guess a pressure and see if it is correct (or use the Solver function in Microsoft Excel). Plot a *P–T* diagram for the univariant line. (Recall that the gas constant R = 8.3144 J mol^{-1} K^{-1}.) Although steam behaves far from ideally under these conditions, the resulting *P–T* diagram is typical of all reactions in which a gas phase is liberated (this includes most important metamorphic reactions).

8.9 At high pressure, the reaction considered in Problem 8.7 will intersect the quartz–coesite polymorphic phase change. Without determining the precise nature of the distribution of univariant lines at this invariant point, make a qualitative prediction of what happens to the slope of this univariant line as it passes through the invariant point. The answer lies in the application of one of Schreinemakers rules.

9 Thermodynamics of solutions

9.1 INTRODUCTION

So far, we have dealt only with systems and minerals of fixed composition. Many systems in nature, however, are open; that is, material can be added or subtracted. In addition, most of the common rock-forming minerals belong to solid solution series in which compositions can vary widely (Becker and Prieto, 2006). Magmas and intergranular solutions in metamorphic rocks are other examples of natural materials of variable composition. Changes in composition bring about changes in energy. Thus, when systems strive for equilibrium, compositional adjustments must be made so as to minimize free energies.

Changes in energy resulting from changes in composition are of two types. First, there is the simple addition of material and the energy that it brings with it. For example, when you add 1 gallon of gasoline to your gas tank, you expect a certain amount of work in return, depending on the efficiency of the automobile. Similarly, a surge of new magma into a cooling body of magma introduces a certain amount of heat, which depends on the temperature, heat capacity, and volume of the added magma.

Material added to a system may react with the initial material and produce additional energy. The amount of energy created in this way depends on the nature of the reaction. Consider, for example, two systems, one consisting of oil and the other of sulfuric acid. To each is added an equal volume of water. Because oil and water do not mix, the change in energy in this system depends only on the heat capacity and temperature of the added water. In the other system, however, the water dissolves readily in the sulfuric acid, and, as anyone knows who has performed this experiment, so much heat is generated that the water–sulfuric acid solution may even boil. Clearly, the energy associated with the addition of material to a system depends on the nature in which that material is taken into the system.

Most petrologically important processes involve changes in composition. Magmas, for example, are complex solutions of seven major oxides and a number of minor and trace elements (Section 6.1). As a magma begins to crystallize, its composition changes depending on the minerals that form. Its composition can also change through additions of new magma or assimilation of xenoliths of country rock. During metamorphism, many important reactions evolve gas (H_2O or CO_2), which is added to the fluid phase that migrates along grain boundaries through the rock. The composition of this fluid is continually being modified by the various reactions that take place. In turn, the composition of this fluid in part determines what reactions can occur.

Igneous and metamorphic rock-forming minerals are mostly members of solid solution series; they, therefore, have a wide range of possible compositions. At the time of formation, however, the most favored composition is that which minimizes the free energy of the system. As conditions of temperature and pressure change, so will the equilibrium composition. But many minerals, once formed, lack the ability to change composition because of very low solid diffusion rates (Section 5.6). Thus, initial compositions may be preserved through a wide range of temperatures and pressures, especially in the cores of large grains. These initial compositions preserve a record of the conditions under which the minerals formed. The problem is how to read that record. Herein lies one of the most important applications of thermodynamics to petrology – the determination of the temperatures (*geothermometry*) and pressures (*geobarometry*) of formation of rocks from mineral compositions.

In this chapter, we deal with the basic principles governing the free energy of solutions. It introduces the important concept of chemical potential, which is the intensive measure of the chemical contribution of a component to the free energy of a solution. Through chemical potential, the free energy of a solution can be related to the free energies of its pure components, values of which can be obtained from thermodynamic tables. Once we know how the free energy of a solution varies with composition, we can determine how to minimize the free energy and thus determine equilibrium compositions. This, for example, is what the MELTS program does for magmas (Ghiorso and Sack, 1995) and the THERMOCALC program does for metamorphic rocks (Powell *et al*., 1998).

9.2 CONSERVATIVE AND NONCONSERVATIVE COMPONENTS OF A SOLUTION

The composition of a solution can be described in many different ways. Consider, for example, how we might describe the composition of a clinopyroxene that is a member of a solid solution whose general formula is

$$Ca^{VIII}(Mg_XFe_{1-X})^{VI}Si_2O_6$$

One mole of this pyroxene could be made from

$$1\,\text{mol Ca} + X\,\text{mol Mg} + (1 - X)\,\text{mol Fe} + 2\,\text{mol Si} + 6\,\text{mol O} = 10\,\text{mol atoms}$$

or

$$1\,\text{mol CaO} + X\,\text{mol MgO} + (1 - X)\,\text{mol FeO} + 2\,\text{mol SiO}_2 = 4\,\text{mol oxides}$$

where X is a mole fraction. In the first case 10 mol of atoms is needed to make 1 mol of pyroxene, whereas in the second case only 4 mol of oxides is needed. Clearly, the number of moles of the components defined in these ways is not a conservative measure of composition, that is, the numbers are not conserved.

Molar amounts can provide a conservative measure of composition if the components chosen to describe the system are selected carefully. In the case of the clinopyroxene we could form 1 mol of pyroxene from

$$X\ \text{mol CaMgSi}_2\text{O}_6 + (1{-}X)\ \text{mol CaFeSi}_2\text{O}_6 = 1\ \text{mol pyroxene}$$

These components, in addition to being conservative, have the advantage that they correspond to minerals for which thermodynamic data are available in Table 7.1.

The composition of a solution can also be expressed in terms of weight percent or fractions of its components, and regardless of how the components are defined, weight always provides a conservative measure of composition. For this reason, most of the diagrams in Chapter 10 relating the compositions and temperatures of silicate liquids to silicate minerals are plotted in terms of weight percent.

9.3 FREE ENERGY OF SOLUTIONS

In Section 7.8, the free energy in a closed system was seen to be a function only of temperature and pressure (review Eq. (7.30)). In an open system, however, the free energy is a function not only of temperature and pressure, but also of composition; that is, $G = G(T,P,n_a,n_b...n_i)$, where n_i is the number of moles of component i. Thus, in the example of the pyroxene above, we can write $G = G(T,P,n_{\text{Di}},n_{\text{Hd}})$. Any change in G can therefore be expressed in terms of the total differential, which for this pyroxene is

$$\mathrm{d}G = \left(\frac{\partial G}{\partial T}\right)_{P,n_{\text{Di}},n_{\text{Hd}}}\mathrm{d}T + \left(\frac{\partial G}{\partial P}\right)_{T,n_{\text{Di}},n_{\text{Hd}}}\mathrm{d}P + \left(\frac{\partial G}{\partial n_{\text{Di}}}\right)_{T,P,n_{\text{Hd}}}\mathrm{d}n_{\text{Di}} + \left(\frac{\partial G}{\partial n_{\text{Hd}}}\right)_{T,P,n_{\text{Di}}}\mathrm{d}n_{\text{Hd}} \qquad (9.1)$$

The partial derivative of G with respect to the number of moles of a particular component is given the name *chemical potential* and is defined as

$$\mu_i \equiv \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}} \qquad (9.2)$$

That is, the chemical potential of component i is the partial molar free energy of i T P, and number of moles of all other (j) components. Equation (9.1) can therefore be rewritten for the clinopyroxene as

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \mu_{\text{Di}}\mathrm{d}n_{\text{Di}} + \mu_{\text{Hd}}\mathrm{d}n_{\text{Hd}} \qquad (9.3)$$

or for the general case as

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \sum_i \mu_i \mathrm{d}n_i \qquad (9.4)$$

An important property of chemical potential is that when multiplied by the number of moles of the component, it gives the contribution of that component to the total free energy of the solution. Thus the total free energy of a solution is simply the sum of all the $n_i\,\mu_i$ terms; that is,

$$G = \sum_i n_i\mu_i \qquad (9.5)$$

If the solution were to contain just one component, diopside for example in the case of the pyroxene, Eq. (9.5) would reduce to $G^{\text{Px}} = n_{\text{Di}}\mu_{\text{Di}}$ or $\mu_{\text{Di}} = G^{\text{Px}}/n_{\text{Di}}$, which is $\bar{G}^{\text{Px}}$. In this limiting case of the solution becoming the pure end member, the chemical potential becomes the molar free energy of the pure substance.

If Eq. (9.5) is differentiated, we obtain

$$\mathrm{d}G = \sum_i n_i \mathrm{d}\mu_i + \sum_i \mu_i \mathrm{d}n_i \qquad (9.6)$$

which if combined with Eq. (9.4), yields

$$S\mathrm{d}T - V\mathrm{d}P + \sum_i n_i \mathrm{d}\mu_i = 0 \qquad (9.7)$$

This is known as the *Gibbs–Duhem* equation. Note that in this equation, the potentials or intensive variables all appear as derivatives, whereas in Eq. (9.4) only T and P appear as such.

In the special case where temperature and pressure are constant, the free energy of a solution can be changed only through changes in composition, and the Gibbs–Duhem equation reduces to

$$\sum_i n_i \mathrm{d}\mu_i = 0 \qquad (9.8)$$

This shows that the chemical potentials of the various components must vary in a related way. For the example of the pyroxene we can write

$$n_{\text{Di}}\mathrm{d}\mu_{\text{Di}} + n_{\text{Hd}}\mathrm{d}\mu_{\text{Hd}} = 0 \qquad (9.9)$$

which on rearranging gives

$$\mathrm{d}\mu_{\text{Hd}} = -\left(\frac{n_{\text{Di}}}{n_{\text{Hd}}}\right)\mathrm{d}\mu_{\text{Di}} \qquad (9.10)$$

Thus, if a variation in composition changes the chemical potential of diopside, the chemical potential of hedenbergite would change according to Eq. (9.10).

Let us consider what happens to the free energy of a system consisting of two crystals, one of augite and the other of pigeonite, if a small quantity of diopside component (d) is transferred from the pigeonite to the augite at constant temperature and pressure ($\mathrm{d}T = 0$, $\mathrm{d}P = 0$). Because G is an extensive property, its values are additive. Therefore,

$$\mathrm{d}G^{\mathrm{Sy}} = \mathrm{d}G^{\mathrm{A}} + \mathrm{d}G^{\mathrm{P}} \leq 0$$

where the superscripts Sy, A, and P stand for system, augite, and pigeonite, respectively. The $\leq$ sign indicates that if the change is to be spontaneous, dG must be negative, and at equilibrium dG is zero. From Eq. (9.3) we can write

$$\mathrm{d}G^{\mathrm{A}} = \mu^{\mathrm{A}}_{\mathrm{Di}}\mathrm{d}n^{\mathrm{A}}_{\mathrm{Di}} + \mu^{\mathrm{A}}_{\mathrm{Hd}}\mathrm{d}n^{\mathrm{A}}_{\mathrm{Hd}}$$

Since only the number of moles of diopside change, this equation reduces to

$$\mathrm{d}G^{\mathrm{A}} = \mu^{\mathrm{A}}_{\mathrm{Di}}\mathrm{d}n^{\mathrm{A}}_{\mathrm{Di}}$$

Similarly, we write for the pigeonite crystal

$$\mathrm{d}G^{\mathrm{P}} = \mu^{\mathrm{P}}_{\mathrm{Di}}\mathrm{d}n^{\mathrm{P}}_{\mathrm{Di}}$$

But $\mathrm{d}n^{\mathrm{A}}_{\mathrm{Di}} = -\mathrm{d}n^{\mathrm{P}}_{\mathrm{Di}}$, so that

$$\mathrm{d}G^{\mathrm{Sys}} = \mu^{\mathrm{A}}_{\mathrm{Di}}\mathrm{d}n^{\mathrm{A}}_{\mathrm{Di}} - \mu^{\mathrm{P}}_{\mathrm{Di}}\mathrm{d}n^{\mathrm{A}}_{\mathrm{Di}} \leq 0$$

or

$$\mathrm{d}G^{\mathrm{Sys}} = \left(\mu^{\mathrm{A}}_{\mathrm{Di}} - \mu^{\mathrm{P}}_{\mathrm{Di}}\right)\mathrm{d}n^{\mathrm{A}}_{\mathrm{Di}} \leq 0 \tag{9.11}$$

Thus, for a spontaneous change to occur, $\mathrm{d}G^{\mathrm{Sy}}$ must be less than zero. This can occur in one of two ways; either $\mathrm{d}n^{A}_{\mathrm{Di}} < 0$ or $\left(\mu^{\mathrm{A}}_{\mathrm{Di}} - \mu^{\mathrm{P}}_{\mathrm{Di}}\right) < 0$, which is to say that $\mu^{\mathrm{P}}_{\mathrm{Di}} > \mu^{\mathrm{A}}_{\mathrm{Di}}$. The first possibility does not satisfy the conditions of the problem, that is, the diopside component is transferred from pigeonite to augite $\left(\mathrm{d}n^{\mathrm{A}}_{\mathrm{Di}} > 0\right)$. We come to the important conclusion, therefore, that if the diopside component is to be transferred spontaneously from pigeonite to augite, the chemical potential of diopside in pigeonite must be greater than the chemical potential of diopside in augite. In other words, *material diffuses from regions of high chemical potential to regions of low chemical potential*. We can now see why we stated in the opening paragraph of Section 5.6 that there are some exceptions to the general tendency for material to diffuse from regions of high concentration to regions of low concentration. It is actually the chemical potential, and not the concentration, that drives diffusion. Although the chemical potential of a component is normally higher when its concentration is higher, there are exceptions.

It is also evident from Eq. (9.11) that at equilibrium $(\mathrm{d}G^{\mathrm{Sy}} = 0)$, $\mu^{\mathrm{A}}_{\mathrm{Di}} = \mu^{\mathrm{P}}_{\mathrm{Di}}$; that is, the transfer of diopside from pigeonite to augite proceeds until, at equilibrium, the chemical potentials of diopside in augite and in pigeonite are equal. This is analogous to heat being transferred from regions of high temperature to regions of low temperature until at equilibrium the temperatures become the same. *At equilibrium, then, potentials such as T, P, and μ_i must be everywhere the same throughout the system.*

9.4 FREE ENERGY OF IDEAL SOLUTIONS

If two end members of a solution are placed together, given time, they will mix and form a homogeneous solution. Obviously, this occurs more rapidly if the end members are gases than if they are solids, but this is a kinetic, not a thermodynamic, distinction. If mixing occurs spontaneously, the free energy of the system is lowered by an amount known as the *free energy of mixing*, ΔG_{mix}. From Eq. (7.24) we can write

$$\Delta G_{\mathrm{mix}} = \Delta H_{\mathrm{mix}} - T\Delta S_{\mathrm{mix}} \tag{9.12}$$

In general, there will be both an enthalpy and entropy of mixing. If, however, the enthalpy of mixing is zero, we have what is known as an *ideal solution*, where

$$\Delta G_{\mathrm{mix}} = -T\Delta S_{\mathrm{mix}} \tag{9.13}$$

To calculate the free energy of mixing of an ideal solution, we must be able to determine the ΔS_{mix} (nonideal solutions will be examined in Section 9.5).

For illustrative purposes we will determine the entropy of mixing of pure diopside and hedenbergite to form an intermediate composition pyroxene with the formula $\mathrm{Ca(Mg}_X\mathrm{Fe}_{1-X}\mathrm{)SiO_2O_6}$, which in terms of end members would be $\mathrm{Di}_X\mathrm{Hd}_{1-X}$. The solution will be assumed to behave ideally (see Grover, 1980, for review of pyroxene thermodynamics). The initial state of the system is represented in Figure 9.1(A) by pure diopside and hedenbergite separated by a partition. Once the partition is removed, the two pyroxenes mix to form the homogeneous pyroxene of the final state.

The entropy in the initial state of the system is simply the weighted average of the entropies of the two end members; that is,

$$S_{\mathrm{initial}} = X_{\mathrm{Di}}S^{\mathrm{Di}} + (1 - X_{\mathrm{Di}})S^{\mathrm{Hd}} \tag{9.14}$$

Graphically, this is obtained by drawing a straight line between the entropies of the end members on an S versus mole fraction plot, and then selecting the value on this line corresponding to the given value of X_{Di} (Fig. 9.1(B)). When the partition is removed, the initial entropies of the two pyroxenes are retained, but to these are added an entropy resulting from the spontaneous mixing of the pyroxenes. The final entropy of the system will be the sum of the initial entropies plus the configurational entropy resulting from mixing. That is,

$$S_{\mathrm{final}} = X_{\mathrm{Di}}\overline{S}^{\mathrm{Di}} + (1 - X_{\mathrm{Di}})\overline{S}^{\mathrm{Hd}} + S_{\mathrm{config}} \tag{9.15}$$

If the system eventually contains 1 mol of homogeneous pyroxene, N_0 (Avogadro's number) sites are available on which to place N_{Di} ions of Mg^{2+} and N_{Hd} ions of Fe^{2+}. It is important at this point to check the formula of the pyroxene to see that 1 mol will indeed have N_0 sites for the magnesium and iron (this depends on how the formula is written). The number of possible ways of arranging these ions (Section 7.6) is

$$\Omega = \frac{N_0!}{N_{\mathrm{Di}}!N_{\mathrm{Hd}}!} \tag{9.16}$$

According to Eq. (7.18), then, the configurational entropy is

$$S_{\mathrm{config}} = k\ln\frac{N_0!}{N_{\mathrm{Di}}!N_{\mathrm{Hd}}!} \tag{9.17}$$

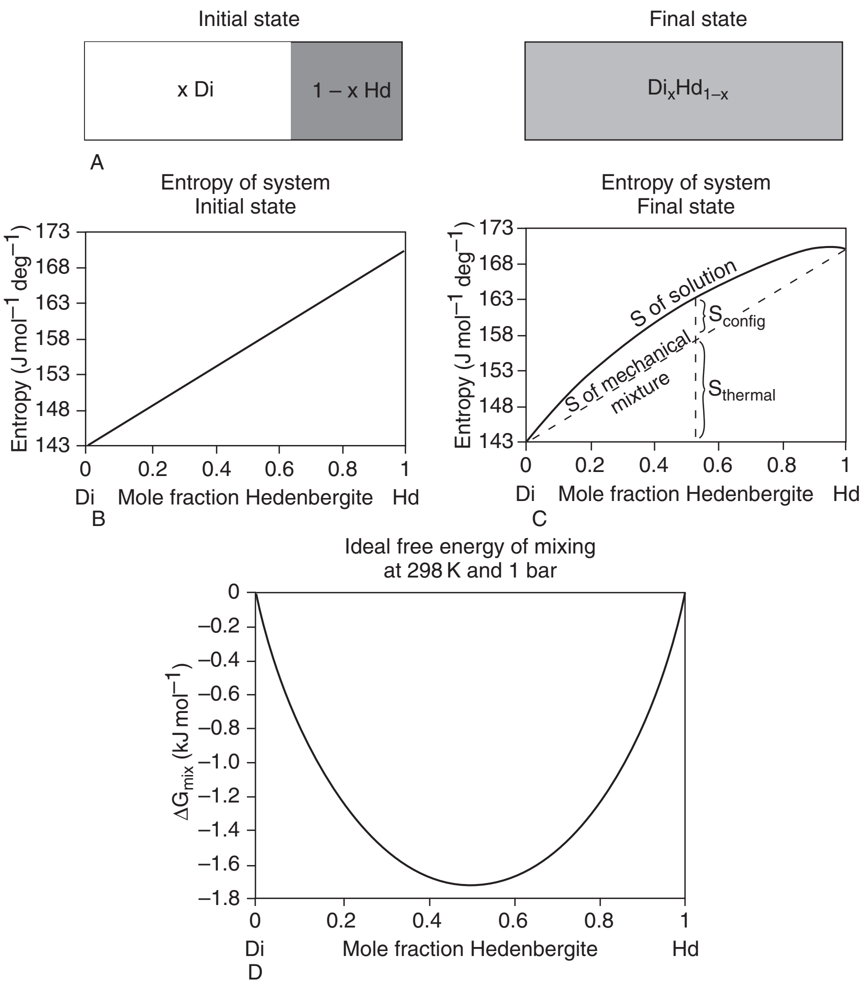


Fig. 9.1 Mixing of diopside and hedenbergite to form a clinopyroxene with composition $CaMg_xFe_{1-x}Si_2O_6$. (**A**) In the initial state of the system, diopside and hedenbergite are separated by a partition, which is then removed and the two pyroxenes mix to form a homogeneous solid solution. (**B**) Initial and (**C**) final entropies of the system, assuming ideal mixing. (**D**) Ideal free energy of mixing at 298 K and 1 bar.

where k is the Boltzmann constant, which is equal to R/N_0. Thus

$$S_{\text{config}} = \frac{R}{N_0}[\ln N_0! - (\ln N_{\text{Di}}! + \ln N_{\text{Hd}}!)] \tag{9.18}$$

Using Stirling's approximation ($\ln N! = N \ln N - N$ when N is large) and recalling that $N_{\text{Di}} + N_{\text{Hd}} = N_0$, Eq. (9.18) becomes

$$S_{\text{config}} = R\left(\frac{N_0}{N_0}\ln N_0 - \frac{N_{\text{Di}}}{N_0}\ln N_{\text{Di}} - \frac{N_{\text{Hd}}}{N_0}\ln N_{\text{Hd}}\right) \tag{9.19}$$

But $N_{\text{Di}}/N_0 = X_{\text{Di}}$ and $N_{\text{Hi}}/N_0 = X_{\text{Hi}}$ and $X_{\text{Di}} + X_{\text{Hd}} = 1$. Therefore,

$$S_{\text{config}} = R\left(\frac{N_{\text{Di}} + N_{\text{Hd}}}{N_0}\ln N_0 - X_{\text{Di}}\ln N_{\text{Di}} - X_{\text{Hd}}\ln N_{\text{Hd}}\right)$$

which on rearranging, gives

$$S_{\text{config}} = -R(X_{\text{Di}}\ln N_{\text{Di}} - X_{\text{Di}}\ln N_0 + X_{\text{Hd}}\ln N_{\text{Hd}} - X_{\text{Hd}}\ln N_0)$$

which simplifies to

$$S_{\text{config}} = -R\left(X_{\text{Di}}\ln\frac{N_{\text{Di}}}{N_0} + X_{\text{Hd}}\ln\frac{N_{\text{Hd}}}{N_0}\right)$$

or

$$S_{\text{config}} = -R(X_{\text{Di}}\ln X_{\text{Di}} + X_{\text{Hd}}\ln X_{\text{Hd}})$$

But $X_{\text{Hd}} = 1 - X_{\text{Di}}$, so

$$S_{\text{config}} = -nR[X_{\text{Di}}\ln X_{\text{Di}} + (1 - X_{\text{Di}})\ln(1 - X_{\text{Di}})] \tag{9.20}$$

The factor n, the number of equivalent sites on which mixing takes place in each formula unit, has been added to this equation to make it generally applicable. In the case of pyroxene [$Ca(Mg,Fe)Si_2O_6$], n is 1, but 1 mol of olivine contains $2 \times N_0$ mixing sites [$(Mg,Fe)_2SiO_4$], so n would be 2; and in garnet [$(Mg,Fe,Ca)_3Al_2Si_3O_{12}$], n would be 3.

The entropy of any homogeneous pyroxene can be found by combining Eqs. (9.20) and (9.15); that is

$$S_{\text{final}} = X_{\text{Di}}\bar{S}^{\text{Di}} + (1 - X_{\text{Di}})\bar{S}^{\text{Hd}} - nR[X_{\text{Di}}\ln X_{\text{Di}} + (1 - X_{\text{Di}})\ln(1 - X_{\text{Di}})] \tag{9.21}$$

Assuming an ideal solution, the variation in the entropy of clinopyroxene as a function of composition at 298 K and 10^5 Pa is given in Figure 9.1(C). The entropy of intermediate pyroxenes is clearly higher than that which would result from a simple admixture of end members. It is higher by the amount contributed by the configurational entropy. That is, the ΔS_{mix} in an ideal solution is the configurational entropy (Eq. (9.20)).

If diopside and hedenbergite mix ideally, ΔG_{mix} (Eq. (9.13)) is given by

$$\Delta G_{\text{mix}} = nRT[X_{\text{Di}} \ln X_{\text{Di}} + (1 - X_{\text{Di}}) \ln(1 - X_{\text{Di}})] \quad (9.22)$$

The graph of this equation is given in Figure 9.1(D). Note that ΔG is always negative, that is, mixing takes place spontaneously, and the greatest lowering of free energy occurs where the mole fraction is 0.5 (Problem 9.1).

We are now in a position to calculate the free energy of formation of intermediate members of an ideal solution, given the free energies of formation of the pure end members. We proceed in the same way as with entropy; a weighted average of the free energies of formation of the end members is added to the ΔG_{mix}. Again, for the example of a clinopyroxene,

$$\begin{aligned}\Delta \bar{G}_f = {} & X_{\text{Di}} \bar{G}_f^{\text{Di}} + (1 - X_{\text{Di}}) \bar{G}_f^{\text{Hd}} \\ & + nRT[X_{\text{Di}} \ln X_{\text{Di}} + (1 - X_{\text{Di}}) \ln(1 - X_{\text{Di}})]\end{aligned} \quad (9.23)$$

The graph of Eq. (9.23) is given in Figure 9.2. The first two terms of the equation give the dashed line; this is equivalent to a simple mechanical mixture (no solution) of end members. Actual free energies are lower than this by the amount contributed by the ΔG_{mix}, which is negative.

Let us consider what significance the slope of the $\Delta \bar{G}_f$ versus X_{Di} line has in Figure 9.2. We can define the slope of this line at any point in terms of the intercepts of the tangent with the G axes. We will define the intercept on the G axis at the diopside end member as μ_{Di} and at the hedenbergite end member as μ_{Hd}. The slope at any mole fraction X_{Di} is then

$$\left(\frac{\partial G}{\partial X_{\text{Di}}}\right)_{T,P} = \frac{\mu_{\text{Di}} - \mu_{\text{Hd}}}{1} = \mu_{\text{Di}} - \mu_{\text{Hd}} \quad (9.24)$$

But the slope can also be obtained by differentiating Eq. (9.23):

$$\left(\frac{\partial G}{\partial X_{\text{Di}}}\right)_{T,P} = \left(\Delta \bar{G}_f^{\text{Di}} + nRTX_{\text{Di}}\right) - \left(\Delta \bar{G}_f^{\text{Hd}} + nRTX_{\text{Hd}}\right) \quad (9.25)$$

from which it follows that

$$\mu_{\text{Di}} = \Delta \bar{G}_f^{\text{Di}} + nRT \ln X_{\text{Di}} \quad (9.26)$$

and

$$\mu_{\text{Hd}} = \Delta \bar{G}_f^{\text{Hd}} + nRT \ln X_{\text{Hd}} \quad (9.27)$$

The intercepts of the tangent line at $X_{\text{Hd}} = 0$ and $X_{\text{Hd}} = 1$ are, in fact, the chemical potentials of diopside and hedenbergite, respectively, in the pyroxene at the point of tangency on the free-energy curve. Note that as the point of tangency approaches pure diopside the intercept, μ_{Di}, approaches the free energy of formation of pure diopside; that is, as $X_{\text{Di}} \rightarrow 1$, $\mu_{\text{Di}} \rightarrow \Delta \bar{G}_f^{\text{Di}}$. Similarly, as $X_{\text{Hd}} \rightarrow 1$, $\mu_{\text{Hd}} \rightarrow \Delta \bar{G}_f^{\text{Hd}}$. This same conclusion was previously deduced from Eq. (9.5). Indeed, the molar free energy of formation of a pure substance $(\Delta \bar{G}_f^{\text{Di}})$ is commonly replaced by the symbol μ_{Di}^*, the * indicating that this is the chemical potential of diopside in the pure end member at the specified temperature and pressure. Equation (9.26) can therefore be written in a general form as

$$\mu_i = \mu_i^* + nRT \ln X_i \quad (9.28)$$

Thus as X_i approaches 1, ln X_i approaches zero and μ_i approaches the chemical potential of the pure end member (see Problem 9.3).

The relation between the chemical potential and mole fraction of a component in an ideal solution (Eq. (9.28)) can also be derived from the functional relation between free energy and the pressure of an ideal gas. From Eq. (8.6) we have

$$\bar{G}_{i,T,P} = \bar{G}_{i,T}^0 + RT \ln P_i \quad \text{bars} \quad (9.29)$$

where $\bar{G}_{i,T}^o$ is the molar free energy of pure gas i at temperature T and a pressure of 1 bar. Although this equation was introduced to evaluate how changes in the intensive parameter P affect the free energy of an ideal gas, it is equally applicable to situations where pressure is decreased by the gas being diluted with another gas; that is, it applies to partial pressures. For ideal gases, the partial pressure (P_i) is related to the pressure of the pure gas (P_i^*) by *Raoult's law*

$$P_i = X_i P_i^* \quad (9.30)$$

Substituting $X_i P_i^*$ for P_i in Eq. (9.29) gives

$$\bar{G}_{i,T,P} = \bar{G}_{i,T}^0 + RT \ln P_i^* + RT \ln X_i \quad (9.31)$$

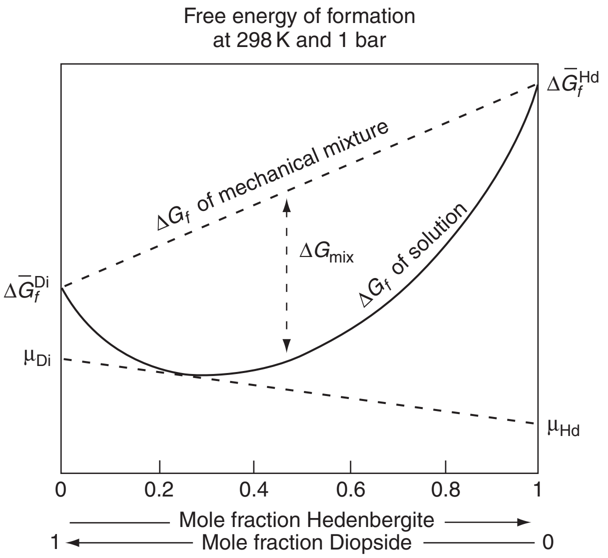


Fig. 9.2 Free energy of formation of diopside–hedenbergite solutions at 298 K and 1 bar, assuming ideal mixing. Intercepts on the ΔG axes of the tangent to the free energy curve give the chemical potentials of diopside and hedenbergite in the solution at the point of tangency.

But according to Eq. (8.6),

$$\bar{G}_{i,T} + RT\ln P_i^* = \bar{G}_{i,T,P}^*$$

where $\bar{G}_{i,T,P}^*$ is the free energy of the pure gas at pressure P^* and temperature T. Thus Eq. (9.31) becomes

$$\bar{G}_{i,T,P} = \bar{G}_{i,T,P}^* + RT\ln X_i \qquad (9.32)$$

But the free energy per mole is the chemical potential. Thus the chemical potential of a component i in an ideal gas solution is

$$\mu_i = \mu_i^* + RT\ln X_i \qquad (9.33)$$

where μ_i^* is the molar free energy of the pure i gas at specified pressure (bars) and temperature. This, then, is identical to Eq. (9.28).

9.5 FREE ENERGY OF NONIDEAL SOLUTIONS

In nonideal solutions, ΔH_{mix} is not zero and the relation between chemical potential and mole fraction no longer holds. To retain the simple functional form of Eq. (9.33), a term known as the *activity*, a_i, is created, such that for a nonideal solution,

$$\mu_i = \mu_i^* + RT\ln a_i \qquad (9.34)$$

The activity can be thought of as the effective mole fraction of a component in a nonideal solution, just as the fugacity is the effective pressure of a nonideal gas (Eq. (8.7)). The activity measures the chemical potential of a component in a nonideal solution in the same way that mole fraction measures the chemical potential of a component in an ideal solution.

The activity of a component is related to its mole fraction by an *activity coefficient*, γ_i;

$$a_i = \gamma_i X_i \qquad (9.35)$$

The activity coefficient is a function of temperature, pressure, and composition; it can be greater than or less than 1, and when equal to 1, $a_i = X_i$, and the solution is ideal. The value of γ_i cannot be determined from thermodynamic properties but must be determined experimentally from measurements of a_i and X_i. Typical relations of a_i and X_i are shown in Figure 9.3. Note that as $X_i \rightarrow 1$, $a_i \rightarrow 1$, and therefore $\gamma_i \rightarrow 1$. At low concentrations, a_i and X_i commonly are linearly related, making it possible to express the activity as

$$a_i = K_i X_i \qquad (9.36)$$

where K_i is known as the *Henry's law constant*. This formulation of the activity is useful in evaluating the distribution of trace elements between coexisting phases.

The free energy of a two-component (1, 2) nonideal solution can be expressed in a form similar to that for an ideal solution (Eq. (9.23)),

$$\Delta\bar{G}_f = X_1\Delta\bar{G}_f^1 + (1 - X_1)\Delta\bar{G}_f^2 + \Delta\bar{G}_{\text{mix}} \qquad (9.37)$$

Because the $\Delta\bar{G}_{\text{mix}}$ may contain a ΔH_{mix} term, and ΔS_{mix} may differ from the ideal entropy of mixing, the nonideal $\Delta\bar{G}$ can be greater or less than the ideal $\Delta\bar{G}_{\text{mix}}$. Indeed, under certain conditions of pressure, temperature, and composition, the $\Delta\bar{G}_{\text{mix}}$ in some solutions can become strongly positive. Before considering the variation of free energy with composition in nonideal solutions, it is necessary to consider the stability of a phase as a function of $G(T,P,X)$.

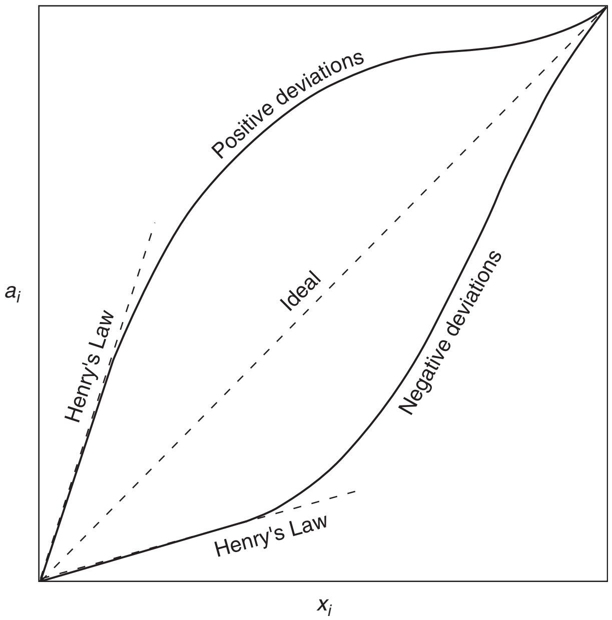


Fig. 9.3 Relation between activity, a_i, and mole fraction, X_i, of a component i in solution. Both positive and negative deviations from ideal behavior are shown. The linear relation between activity and mole fraction at low concentrations of i is referred to as Henry's law.

For a binary solution (ideal or nonideal) we can write from Eq. (9.4)

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \mu_1\mathrm{d}n_1 + \mu_2\mathrm{d}n_2 \qquad (9.38)$$

To determine if this solution is thermally stable, we take first and second derivatives of G with respect to T, keeping the other variables constant:

$$\frac{\partial G}{\partial T} = -S \quad \text{and} \quad \frac{\partial^2 G}{\partial T^2} = -\frac{\partial S}{\partial T}$$

But $\partial S/\partial T = C_p/T$, and because $C_p > 0$ and $T > 0$, $-\partial S/\partial T < 0$. Therefore, $\partial^2 G/\partial T^2 < 0$. Thus for a phase to be thermally stable, it must not only have a negative slope on a G–T plot ($\partial G/\partial T = -S$), but it must be concave downward ($\partial^2 G/\partial T^2 < 0$) (Fig. 9.4(A)).

To determine whether a phase has mechanical stability, we take first and second derivatives of Eq. (9.38) with respect to P:

$$\frac{\partial G}{\partial P} = V \quad \text{and} \quad \frac{\partial^2 G}{\partial P^2} = \frac{\partial V}{\partial P} = -V\beta < 0$$

The second derivative is negative because the volume and coefficient of compressibility (β) are both positive. On a G–P plot, a mechanically stable phase must have a positive slope (V) that is concave downward (Fig. 9.4(B)).

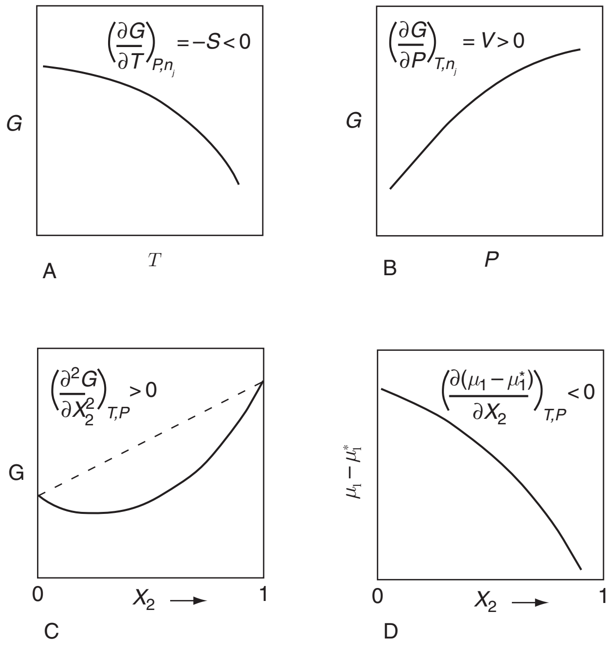


Fig. 9.4 Necessary conditions for a solution to be stable. (**A**) The *G* versus *T* plot must have a negative slope and be concave downward for thermal stability. (**B**) The *G* versus *P* plot must have a positive slope and be concave downward for mechanical stability. (**C**) The *G* versus *X* plot must be concave upward for chemical stability. (**D**) The *μ–μ** versus *X* plot must have a negative slope for chemical stability.

Finally, to determine whether a phase is chemically stable, we take first and second derivatives of Eq. (9.38) with respect to the mole fraction of, for example, component 2, recognizing of course that $X_2 = 1 - X_1$:

$$\frac{\partial G}{\partial X_2} = \mu_2 - \mu_1 \quad \text{and} \quad \frac{\partial^2 G}{\partial X_2^2} = \frac{\partial \mu_2}{\partial X_2} > 0$$

According to Eq. (9.34), addition of component 2 to the solution increases the chemical potential of component 2; therefore, $\partial\mu_2/\partial X_2$ must be positive. On a *G*−*X* plot, then, a stable phase must plot as a concave upward line. If it becomes convex upward, the phase will become unstable (Fig. 9.4(C)). To find out how the chemical potential of one of the components varies with composition in a chemically stable solution, we take first and second derivatives of Eq. (9.28):

$$\frac{\partial(\mu_1 - \mu_1^*)}{\partial X_2} = \frac{-RT}{1 - X_2} < 0 \quad \text{and} \quad \frac{\partial^2(\mu_1 - \mu_1^*)}{\partial X_2^2} = \frac{-RT}{(1 - X_2)^2} < 0$$

Because the first and second derivatives are both negative, a phase must plot as a negatively sloping concave downward line in a $\mu_1 - \mu_1^*$ versus *X* plot if it is to be stable (Fig. 9.4(D)). If this condition is not met, the phase becomes unstable and the solution unmixes.

Examination of Figure 9.2 reveals that the variation of free energy with composition in an ideal solution satisfies the requirements for chemical stability. Because the $\Delta\bar{G}_{\text{mix}}$ in an ideal solution is always negative, the free energy versus mole fraction plot is always concave upward. The same, however, is not always true for nonideal solutions. The $\Delta\bar{G}$ in nonideal solutions can even become positive, which makes the solution chemically unstable. We will now examine how the $\Delta\bar{G}_{\text{mix}}$ in nonideal solutions can vary.

9.6 NONIDEAL SOLUTION: THE REGULAR SOLUTION MODEL

The $\Delta\bar{G}_{\text{mix}}$ of nonideal solutions cannot be deduced from thermodynamic properties of the end members of a solution; it must be measured. Once the measured values are obtained, a number of physical models of solutions can be used to account for the variation of $\Delta\bar{G}_{\text{mix}}$ with composition. We will examine two of these, the *symmetric regular solution model* and the *asymmetric regular solution model* (Thompson, 1967).

In an ideal solution, the $\Delta\bar{G}_{\text{mix}}$ results only from the entropy of mixing (Eq. (9.22)), which is calculated from the number of possible ways of arranging the components on the available sites in the solution (Eq. (9.16)). It was assumed that the occupying of a site by one component did not affect the statistical probability of what component might occupy an adjoining site. If, however, different pairs of molecules have different interaction energies, this assumption would not be valid. The interaction between components in a regular solution is accounted for in the expression for the $\Delta\bar{G}_{\text{mix}}$ by adding an excess energy term to the ideal mixing term; that is,

$$\Delta\bar{G}_{\text{mix}} = \Delta\bar{G}_{\text{ideal}} + \Delta\bar{G}_{\text{excess}} \tag{9.39}$$

Prior to mixing in a two-component system (*A* and *B*) there would be separate groups of *A*–*A* and *B*–*B* molecules. After mixing, juxtaposed *A*–*B* molecules will be present. We let W_{AA} represent the increase in potential energy in bringing two molecules of *A* from infinity to adjoining sites in the pure *A* end member, and W_{BB} represents the equivalent increase in potential energy for the *B* end member. If the increase in potential energy in bringing one molecule of *A* and one of *B* from infinity to adjoining sites in the solution is W_{AB}, the change in potential energy resulting from mixing pure *A* and *B* would be

$$W = W_{AB} - \frac{W_{AA}}{2} - \frac{W_{BB}}{2} \tag{9.40}$$

W, which is known as the *interchange energy*, can be positive or negative. The $\Delta\bar{G}_{\text{excess}}$ is then given by

$$\Delta\bar{G}_{\text{excess}} = nWX_AX_B \tag{9.41}$$

The factor *n* is the number of equivalent sites on which mixing takes place (see Eq. (9.20)). The $\Delta\bar{G}_{\text{mix}}$ of a regular binary solution is therefore

$$\Delta\bar{G}_{\text{mix}} = \underbrace{nRT(X_A \ln X_A + X_B \ln X_B)}_{\text{ideal}} + \underbrace{nWX_AX_B}_{\text{excess}} \tag{9.42}$$

Note that as $X_A \rightarrow 1$, $X_B \rightarrow 0$, and therefore $\Delta\bar{G}_{\text{excess}} \rightarrow 0$. The value of $\Delta\bar{G}_{\text{excess}}$ is a maximum at a mole fraction of 0.5, and it decreases symmetrically above and below this value. For this reason this particular model is known as the *symmetric regular solution model*

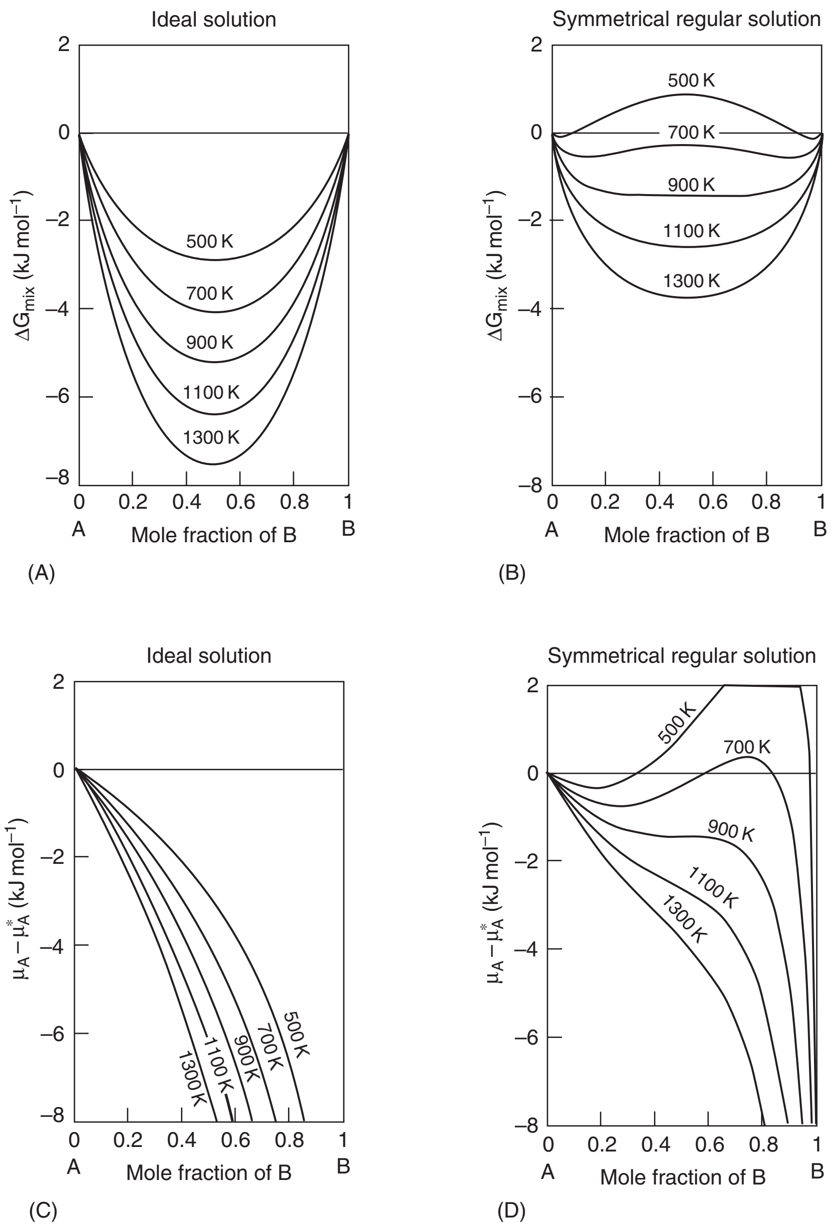


Fig. 9.5 (**A**) The ideal component of the free energy of mixing of a binary solution at temperatures ranging from 500 to 1300 K. (**B**) Actual free energies of mixing (ideal + excess) of a symmetric regular solution having an interchange energy of 15 kJ mol^{-1} at temperatures ranging from 500 to 1300 K. Below 900 K, part of the ΔG curve is concave downward and hence unstable. (**C**) Variation in the ideal component of the chemical potential of component A as a function of X_B. (**D**) Actual variation (ideal + excess) in chemical potential of component A at the same temperatures as in (B).

The ideal term in Eq. (9.42) is always negative and will always dominate the expression for $\Delta\bar{G}_{\text{mix}}$ at high temperatures. The excess term, on the other hand, can be positive or negative. It can, therefore, increase or decrease the $\Delta\bar{G}_{\text{mix}}$ over that which would occur from ideal mixing alone. At low temperatures, the excess term dominates the expression for $\Delta\bar{G}_{\text{mix}}$. These relations are illustrated in Figure 9.5. The ideal component of the $\Delta\bar{G}_{\text{mix}}$ is plotted separately for temperatures ranging from 500 to 1300 K (Fig. 9.5(A)) so that it can be compared with the actual $\Delta\bar{G}_{\text{mix}}$ that would result from an interchange energy of +15 kJ mol^{-1}. At temperatures in excess of 900 K, the $\Delta\bar{G}_{\text{mix}}$ versus X plot for this regular solution is concave upward, but the values of $\Delta\bar{G}_{\text{mix}}$ are far less negative than they would be if the solution behaved ideally. Below 900 K, the excess term causes part of the $\Delta\bar{G}$ X plot to become concave downward, and over this compositional range the solution is unstable and must unmix; that is, exsolution will occur.

The variation in the chemical potential of one component $\left(\mu_A - \mu_A^*\right)$ as a function of composition in a symmetrical regular solution is given by

$$\left(\mu_A - \mu_A^*\right) = \underbrace{nRT\ln(1 - X_B)}_{\text{ideal}} + \underbrace{nWX_B^2}_{\text{excess}} \tag{9.43}$$

This also consists of ideal and excess terms. A graph of the ideal part of this equation is plotted in Figure 9.5(C). At all temperatures, the plots have negative slopes and are concave downward. But when the excess term is added, the slopes are greatly diminished, and at temperatures below 900 K part of each curve has a positive slope. Solutions having compositions in the range where the slopes are positive are unstable.

Because of the form of the expression for the $\Delta\bar{G}_{\text{mix}}$ of the symmetrical regular solution model, instabilities will first occur with falling temperature at a mole fraction of 0.5. With falling temperature, the region of instability expands, but it does so symmetrically about the mole fraction of 0.5 (Fig. 9.5(B)). In some solutions, however, deviations from ideality are greater near one end member than they are near the other. For these solutions the *asymmetric regular solution model* (Margules' formulation) is used. Here a separate value of W is used for each of the components, and then the $\Delta\bar{G}_{\text{mix}}$ is given by

$$\begin{aligned}\Delta\bar{G}_{\text{mix}} &= nRT(X_A \ln X_A + X_B \ln X_B)\\ &\quad + nX_AX_B(X_BW_A + X_AW_B)\end{aligned} \tag{9.44}$$

or

$$\begin{aligned}\Delta\bar{G}_{\text{mix}} &= nRT[(1-X_B)\ln(1-X_B) + X_B \ln X_B]\\ &\quad + n(1-X_B)X_B[X_BW_A + (1-X_B)W_B]\end{aligned} \tag{9.45}$$

With the asymmetric regular solution model, variation in the chemical potential of one component as a function of composition is given by

$$\begin{aligned}\mu_A - \mu_A^* &= nRT\ln(1-X_B)\\ &\quad + nX_B^2[W_A + 2(W_B - W_A)(1-X_B)]\end{aligned} \tag{9.46}$$

Note that if $W_A = W_B$, then Eq. (9.46) reduces to Eq. (9.43).

9.7 UNMIXING OF NONIDEAL SOLUTIONS: EXSOLUTION

Examples of unmixing of nonideal solutions can be found in a number of important petrologic systems. For example, alkali feldspars (perthite), plagioclase (peristerite), pyroxenes, amphiboles, spinels, and numerous sulfides all form solutions that may unmix at low temperatures. The unmixing produces intergrowths of the two stable phases, forming what are known as *exsolution textures*. These textures commonly exhibit striking geometrical patterns that result from exsolving phases orienting themselves crystallographically so as to minimize the stress caused by having one phase in another (Fig. 9.6(A)). Unmixing also occurs in magmas to form silicate liquids of strikingly different compositions (Fig. 9.6(B)).

Using the symmetric regular solution model as an example of a nonideal solution, unmixing is seen to occur only if the interchange energy, W, is positive. If W is positive, below some critical temperature, T_c, the solution has a higher free energy than it would if it unmixed. Below this temperature the solution is unstable, which, as shown in Section 9.5, occurs when $d\mu_A/dX_B$ becomes positive. The critical temperature, therefore, can be determined by differentiating Eq. (9.43) and setting the derivative equal to zero:

$$\frac{d\mu_A}{dX_B} = \frac{nRT_c}{1-X_B} + 2nWX_B = 0 \tag{9.47}$$

Because this critical temperature occurs at a mole fraction of 0.5 in the symmetric regular solution model, it follows that

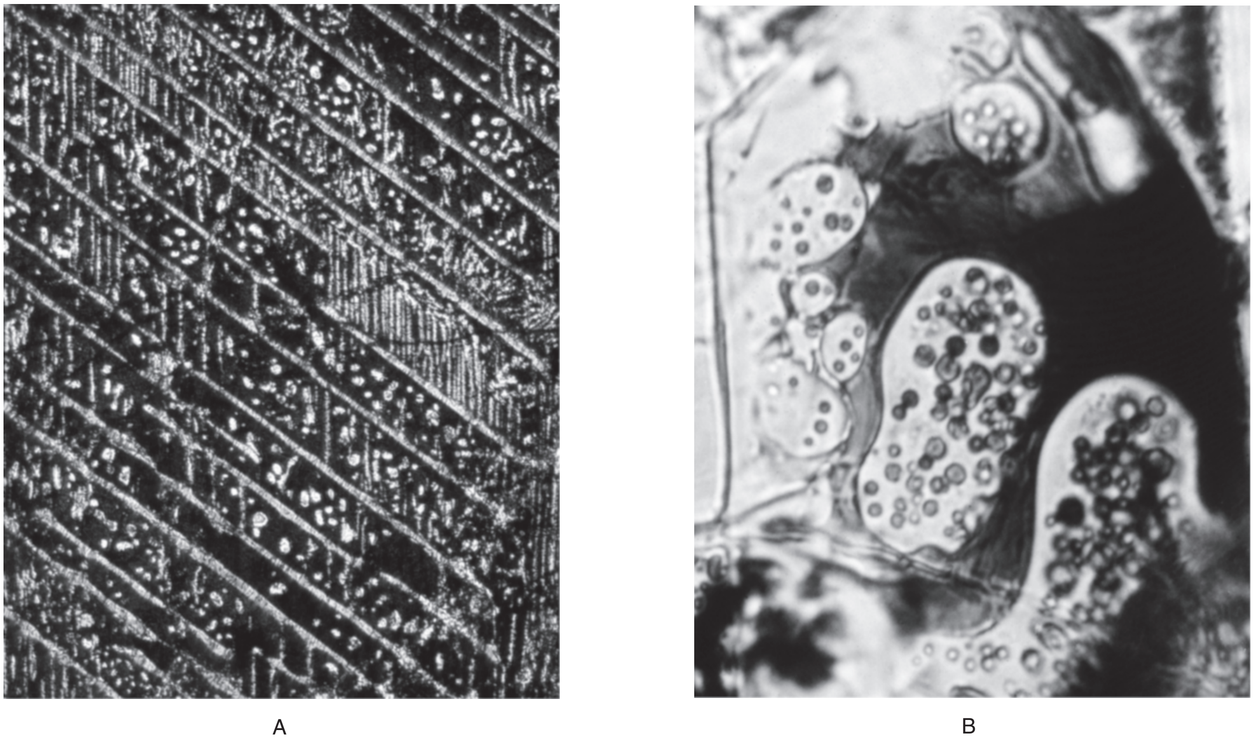


Fig. 9.6 Exsolution textures developed in solid (**A**) and liquid (**B**) solutions. (**A**) Two directions of augite exsolution lamellae in a crystal of orthopyroxene. (**B**) Exsolution of an iron-rich silicate liquid (dark) from an iron-poor silicate liquid (clear) in the residual liquid in a basalt from Mauna Loa, Hawaii.

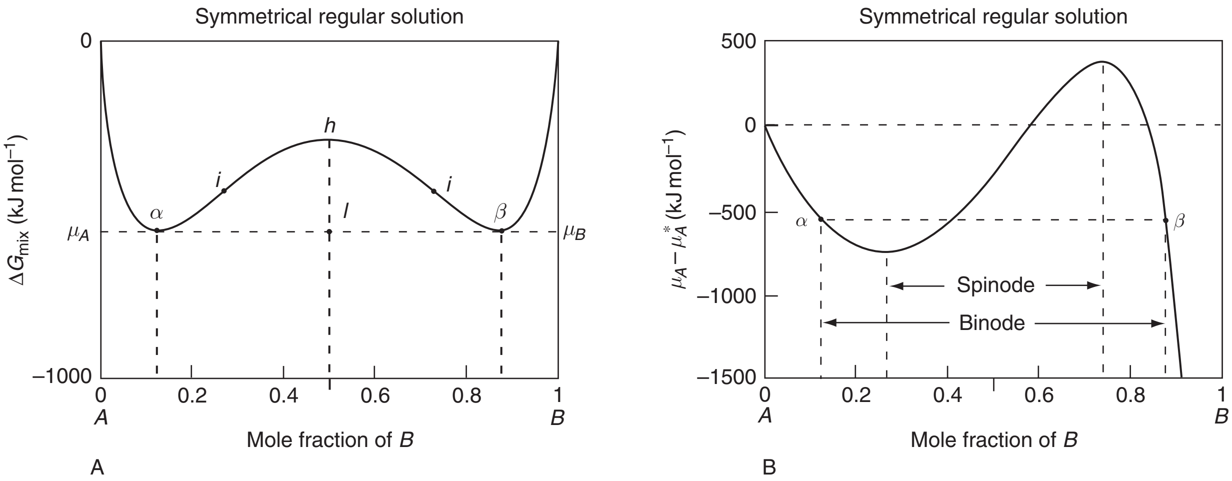


Fig. 9.7 (**A**) Variation in the ΔG_{mix} of a symmetric regular solution at 700 K in which the interchange energy is 15 kJ mol^{-1}. Points α and β define the compositions of coexisting phases on the solvus at this temperature – α and β both have the same chemical potentials as *A* and *B*. The inflection points, *i*, on the free-energy curve mark the compositions of the spinode at this temperature. (**B**) Variations in the chemical potential of *A* ($\mu_A - \mu_A^*$) as a function of X_B. Compositions α and β mark the binode, and the points where the slope on the curve is zero mark the spinode.

$$T_c = \frac{2WX_B(1-X_B)}{R} = \frac{W}{2R} \tag{9.48}$$

The critical temperature in the example worked out in Figure 9.5, where $W = 15\,\text{kJ mol}^{-1}$, would therefore be 15 000/(2 × 8.3144) or 902 K. Note that in Figure 9.5(D), the plot for $\mu - \mu_i^*$ versus X_B at 900 K has a short segment in which the slope is positive; that is, the solution is unstable.

At any given temperature below T_c, the $\Delta\bar{G}_f$ versus X plot for a nonideal binary solution is of the general form shown in Figure 9.7(A). In this diagram, the $\Delta\bar{G}_f$ of the two pure end members have been assigned equal values, to simplify the illustration. The following argument, however, is just as valid if the two end members have different values.

From Figure 9.7(A) we see that a solution having a mole fraction of 0.5*A* and 0.5*B* has a higher free energy (*h*) than if this bulk composition were composed of two solutions, one richer in *A* and the other richer in *B*. Indeed, from inspection of Figure 9.7(A), the free energy of the system is seen to be minimized if the solution unmixes into the two compositions indicated by α and β. If the solution unmixes to these two compositions, the free energy of the system would be lowered to the point *l*, which is simply the sum of $\Delta\bar{G}_f^\alpha + \Delta\bar{G}_f^\beta$. It will be noted that a tangent to the ΔG versus X curve at point α is also tangent at point β, indicating that $\mu_A^\alpha = \mu_A^\beta$ and $\mu_B^\alpha = \mu_B^\beta$. At the particular temperature and pressure represented in Figure 9.7(A), solutions having compositions of α and β are the equilibrium phases that should form (given appropriate kinetics) for bulk compositions between α and β.

The precise compositions of α and β at any given temperature and pressure can be determined by differentiating Eq. (9.42) and setting the derivative equal to zero (minimum) (see Problem 9.7). From this it follows that

$$T_{\text{binode}} = \frac{W\left(1 - 2X_B^\alpha\right)}{R\ln\left[\left(1 - X_B^\alpha\right)/X_B^\alpha\right]} \tag{9.49}$$

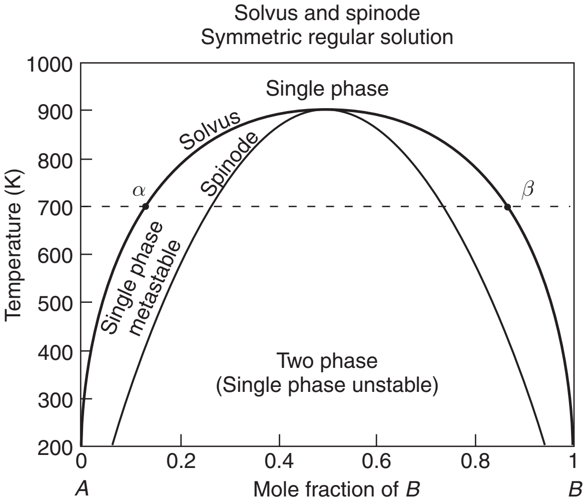


Fig. 9.8 Plot of the solvus and spinode in a binary (*A* – *B*) symmetric regular solution model having an interchange energy of 15 kJ mol^{-1}. Compositions α and β at 700 K are the same points as in Figure 9.7. Between the solvus and spinode, single-phase solutions are metastable and, given time, should exsolve. Below the spinode, all single-phase solutions are unstable and must unmix.

The term *binode* indicates that, at this temperature, the two nodes on the free-energy curve indicate that two phases coexist and have lower free energies than intermediate compositions. Also note that because we are treating a symmetrical regular solution, $X_B^\alpha = X_A^\beta$.

The compositions α and β in Figure 9.7(A) are for one particular temperature. We know from Fig. 9.5(B) that if the temperature is lowered, the compositional difference between α and β will increase. The locus of points tracing the binodes in a *T*–*X* plot is known as a *solvus* (Fig. 9.8). A solvus in a binary *T*–*X* diagram is a line at temperatures above which a homogeneous solution is stable but below which two separate phases have a lower free energy and hence are more stable than a single phase. At any temperature and bulk composition below the solvus, the equilibrium compositions of coexisting phases are those indicated by the compositions on the solvus on opposite sides of the two-phase region (α and β, for example). Clearly, these compositions change with temperature; if the compositions are preserved, for example by rapid eruption and quenching of magma, they can serve as an indicator of temperature. For example, a pair of alkali feldspars having compositions of 0.45 and 0.55 X_{Ab} must have formed at a higher temperature than a pair having compositions of 0.35 and 0.65 X_{Ab}.

We can also examine the unmixing of nonideal solutions in terms of the variation of the chemical potential of a component as a function of composition. In Figure 9.7(B), the variation in the chemical potential of A $\left(\mu_A - \mu_A^*\right)$ is plotted versus the mole fraction of the other component, X_B. We know that once the slope becomes positive in such a plot, the solution is no longer stable. The two phases that form by unmixing must have the same chemical potential as a component in both. In addition, because of the symmetric regular solution model, if the mole fraction of component B in one phase is X_B^α, the mole fraction of B in the other phase must be $1 - X_B^\alpha$, that is, $X_B^\alpha = X_A^\beta$. These conditions are satisfied by only one pair of compositions for α and β at this temperature, and this then defines the solvus, or binode, at this temperature.

Despite the thermodynamic arguments presented above, when solutions cool below a solvus, exsolution occurs only if there is time to nucleate the appropriate disparate phases and to diffuse the necessary components to these phases for growth. In many silicates, especially at lower temperatures, diffusion rates are so low that solutions formed above the solvus may remain homogeneous, but metastable, at temperatures below the solvus. But over a certain composition range, even rapidly cooled high-temperature solutions are not able to remain homogeneous.

Consider a composition on the ΔG versus X curve in Figure 9.7(A) that falls between the two inflection points (i). Any fluctuation in composition, no matter how small, in a solution in this range would lower the free energy of the system, because the curve is concave downward; that is, a line joining any two compositions on the free-energy curve in this interval falls below the curve. Beyond the inflection points, however, small fluctuations of composition cause the free energy of the system to rise, because the free-energy curve is concave upward; in this range it is necessary to nucleate phases of very different composition to bring about a lowering of free energy – phases on either side of the solvus. Between the inflection points, phase separation cannot be prevented, but because it involves only small fluctuations in composition, the separation typically takes place on a submicroscopic scale; hence X-ray diffraction or electron microscopy is normally required to detect this type of exsolution. The locus of points marking the inflection points in a T–X diagram is known as a *spinode* (Fig. 9.8).

To determine the composition of the spinode at any temperature and pressure we need to find where the inflection points are on the ΔG versus X curve. This is done by setting the second derivative of Eq. (9.42) equal to zero (Problem 9.8). From this we obtain

$$T_{\text{spinode}} = \frac{2WX_B(1 - X_B)}{R} \tag{9.50}$$

In Eq. (9.48), the critical temperature on the solvus, T_{c}, was shown to be $W/2R$. Therefore $W/R = 2T_{\mathrm{c}}$, which when substituted into Eq. (9.50) gives

$$T_{\text{spinode}} = 4T_{\mathrm{c}}X_B(1 - X_B) \tag{9.51}$$

The inflection points on the ΔG versus X plot correspond to the two points on the $\mu_A - \mu_A^*$ versus X plot having zero slope (Fig. 9.7(B)). Thus, T_{spinode} could equally well have been determined by setting the first derivative of Eq. (9.43) equal to zero (Problem 9.6c).

Figure 9.8 shows both the binode, or solvus, and the spinode. With slow cooling, exsolution will strive to achieve the equilibrium compositions given by the solvus. Because of the kinetics of diffusion, especially in solid solutions, metastable high-temperature solutions are commonly preserved to low temperatures. Within the spinode, however, even metastable solutions become unstable and break down into small domains (mostly submicroscopic) of solutions having compositions given by the spinode.

Although the model presented here for the development of the spinode is satisfactory for symmetric regular *liquid* solutions, it tends to give somewhat too high temperatures for spinodes in *solid* solutions. Compositional fluctuations induce strains in crystals that must be taken into account in considering the energetics of spinodal exsolution. However, this is beyond the scope of this book.

9.8 EQUILIBRIUM CONSTANT OF A REACTION

Many reactions involve phases that are members of solutions. Commonly, such reactions do not involve the elimination of a phase but simply the adjusting of compositions of the solutions so as to minimize free energies. Consider, for example, an olivine crystal reacting with silica in a magma to form an orthopyroxene crystal. Each of these phases is a member of a solution, and as the reaction proceeds, the compositions, and thus the chemical potentials of the components in these solutions, change. The silica content of the magma decreases as the olivine reacts to form orthopyroxene; this in turn decreases the chemical potential of silica in the magma. The reaction does not proceed until the magma contains no silica, but only until the chemical potential of silica in the magma is decreased to a level where it is in equilibrium with the olivine.

Let us consider the simple reaction that might occur between olivine and orthopyroxene in a peridotite during metamorphism. Both minerals belong to solid solution series in which iron can be exchanged for magnesium. Equilibrium between these minerals requires that the iron and magnesium be distributed in such a way as to minimize the free energy of the system. This equilibrium can be achieved through a simple exchange reaction:

$$\begin{array}{cccc} \mathrm{Mg_2SiO_4} + & \mathrm{2FeSiO_3} \rightleftharpoons & \mathrm{Fe_2SiO_4} + & \mathrm{2MgSiO_3} \\ \text{Forsterite} & \text{Ferrosilite} & \text{Fayalite} & \text{Enstatite} \end{array} \tag{9.52}$$

This reaction, of course, does not proceed all the way to the right or to the left. Instead, at equilibrium, the olivine and the orthopyroxene both have specific Mg/Fe ratios. The question is: what determines these ratios?

As this reaction proceeds toward equilibrium, the reactants are not pure forsterite, ferrosilite, fayalite, and enstatite. Instead, the forsterite, for example, is in solution in an olivine and the ferrosilite is in solution in an orthopyroxene. Thus, if we are to write an expression for the ΔG of this reaction as it approaches equilibrium, it must be done in terms of the free energies of the reactants in their respective solutions; that is,

$$\Delta G_r = n_{\mathrm{Fa}}^{\mathrm{Ol}}\mu_{\mathrm{Fa}}^{\mathrm{Ol}} + n_{\mathrm{En}}^{\mathrm{Opx}}\mu_{\mathrm{En}}^{\mathrm{Opx}} - n_{\mathrm{Fo}}^{\mathrm{Ol}}\mu_{\mathrm{Fo}}^{\mathrm{Ol}} - n_{\mathrm{Fs}}^{\mathrm{Opx}}\mu_{\mathrm{Fs}}^{\mathrm{Opx}}$$

For the reaction to balance, the number of moles of fayalite must equal the number of moles of forsterite, and the numbers of moles of ferrosilite and enstatite must be twice that of fayalite. If we let there be 1 mol of fayalite, and allow the reaction to proceed to equilibrium ($\Delta G = 0$), we can write

$$\mu_{\mathrm{Fa}}^{\mathrm{Ol}} + 2\mu_{\mathrm{En}}^{\mathrm{Opx}} - \mu_{\mathrm{Fo}}^{\mathrm{Ol}} - 2\mu_{\mathrm{Fs}}^{\mathrm{Opx}} = 0 \tag{9.53}$$

If we assume that both olivine and orthopyroxene form ideal solid solutions, each of the chemical potential terms can be replaced with the expression from Eq. (9.28); for example,

$$\mu_{\mathrm{Fa}}^{\mathrm{Ol}} = \mu_{\mathrm{Fa}}^{*\mathrm{Fa}} + RT\ln X_{\mathrm{Fa}}^{\mathrm{Ol}}$$

where $\mu_{\mathrm{Fa}}^{*\mathrm{Fa}}$ is the chemical potential or molar free energy of pure fayalite at the specified temperature and pressure. If the solutions did not behave ideally, we would make use of Eq. (9.34) and the compositions would be expressed in terms of activities. Assuming ideal behavior, and making this type of substitution for each term in Eq. (9.53) and rearranging, yields

$$\underbrace{\mu_{\mathrm{Fa}}^{*\mathrm{Fa}} + 2\mu_{\mathrm{En}}^{*\mathrm{En}} - \mu_{\mathrm{Fo}}^{*\mathrm{Fo}} - 2\mu_{\mathrm{Fs}}^{*\mathrm{Fs}}}_{\Delta G_r} + RT\ln\underbrace{\frac{\left(X_{\mathrm{Fa}}^{\mathrm{Ol}}\right)\left(X_{\mathrm{En}}^{\mathrm{Opx}}\right)^2}{\left(\mathrm{X}_{\mathrm{Fo}}^{\mathrm{Ol}}\right)\left(X_{\mathrm{Fs}}^{\mathrm{Opx}}\right)^2}}_{K} = 0 \tag{9.54}$$

The first group of terms is simply the free-energy change of reaction between the pure end members as written in Eq. (9.52). We can therefore write

$$-\frac{\Delta G_r}{RT} = \ln\frac{\left(X_{\mathrm{Fa}}^{\mathrm{Ol}}\right)\left(X_{\mathrm{En}}^{\mathrm{Opx}}\right)^2}{\left(\mathrm{X}_{\mathrm{Fo}}^{\mathrm{Ol}}\right)\left(X_{\mathrm{Fs}}^{\mathrm{Opx}}\right)^2} \tag{9.55}$$

The group of terms in the right-hand term is the familiar *equilibrium constant* of a reaction, K, which is the product of the molecular concentrations of the resultants of a reaction divided by the product of the molecular concentrations of the reactants; if any coefficient in the reaction is not 1, the concentration of that substance is raised to the power of that coefficient. If the solutions did not behave ideally, activities would be used instead of mole fractions. In this reaction, the X_{En} and $\mathrm{X_{Fs}}$ in the orthopyroxene are raised to the power 2. Equation (9.55) then becomes

$$\ln K = -\frac{\Delta G_r}{RT} \tag{9.56}$$

Equation (9.56) indicates that for a reaction where the reactants and products are in solution in phases, the equilibrium constant, K, is a function of the ΔG of the reaction between the pure end members. If the solutions behave ideally, the equilibrium constant is a direct measure of the compositions of the phases. If the solutions behave nonideally, the equilibrium constant is expressed in terms of activities; compositions can then be determined only after the activity coefficients are known.

Because the ΔG_r in Eq. (9.56) is likely to change with changes in temperature, and to a lesser extent, pressure, the equilibrium constant will also change. This means that the compositions of the minerals involved in the reaction may provide a measure of the temperatures and pressures at the time of reaction. A reaction that provides a sensitive measure of temperature can be used as a *geothermometer*, whereas one that is sensitive to changes in pressure can be used as a *geobarometer* (Essene, 1982).

To analyze how the equilibrium constant varies with temperature and pressure, we write Eq. (9.54) for a general case involving i different chemical potentials

$$R\ln K = -\frac{\sum v_i\mu_i^*}{T} \tag{9.57}$$

where v_i is the coefficient of the ith component in the reaction ($v = 1$ for fayalite and forsterite and 2 for enstatite and ferrosilite in Eq. (9.54)). The μ_i^* are functions of T and P; we therefore can express them in differential form by their total differentials:

$$\begin{aligned} R\,\mathrm{d}\ln K &= -\sum v_i\left[\left(\frac{\partial\mu_i^*/T}{\partial T}\right)_P \mathrm{d}T + \left(\frac{\partial\mu_i^*/T}{\partial P}\right)_T \mathrm{d}P\right] \\ &= -\sum v_i\left[\left(\frac{\partial\mu_i^*/T}{\partial T}\right)_P \mathrm{d}T + \frac{1}{T}\left(\frac{\partial\mu_i^*}{\partial P}\right)_T \mathrm{d}P\right] \end{aligned} \tag{9.58}$$

The group of terms in the first set of parentheses is the same as $\left(\partial\bar{G}_i/T/\partial T\right)_P$, which, from the rules of differentiation of a product, is

$$\frac{1}{T}\left(\frac{\partial\bar{G}_i}{\partial T}\right)_P - \frac{1}{T^2}\bar{G}_i$$

But $\left(\partial\bar{G}_i/\partial T\right)_P = -\bar{S}$ (Eq. (7.31)), so that these terms can be expressed as

$$-\frac{\bar{S}_i}{T} - \frac{\bar{G}_i}{T^2} = -\frac{T\bar{S}_i + \bar{G}_i}{T^2} = -\frac{\bar{H}_i}{T^2}$$

When multiplied by v_i and summed over the i different pure end members, this becomes $-\Delta H_r/T^2$. The group of terms in the second set of parentheses, $\left(\partial\mu_i^*/\partial P\right)_T$, can, according to Eq. (7.32), be replaced by $\bar{V}_i$, which when summed over the i different pure end members becomes ΔV_r. Equation (9.58) can then be rewritten as

$$R\mathrm{d}\ln K = \left(\frac{\Delta H_r}{T^2}\right)_P \mathrm{d}T - \left(\frac{\Delta V_r}{T}\right)_T \mathrm{d}P \qquad (9.59)$$

Equation (9.59) allows us to determine how the equilibrium constant of a reaction varies as a function of temperature and pressure. If pressure is held constant ($\mathrm{d}P=0$), Eq. (9.59) becomes

$$\left(\frac{\partial \ln K}{\partial T}\right)_P = \frac{\Delta H_r}{RT^2} \qquad (9.60)$$

This is known as the *van't Hoff* equation or *Gibbs–Helmholtz* equation. It indicates that if a reaction is endothermic ($\Delta H_r > 0$), $\ln K$, and thus K, increases with increasing temperature; that is, the reaction, as written, proceeds further to the right (increased products). If the reaction is exothermic ($\Delta H_r < 0$), K decreases with increasing temperature, which is to say that the reaction proceeds more to the left, thus increasing the concentrations of the reactants.

It also follows from Eq. (9.60) that a good geothermometer should have a large value of ΔH_r. In the example given in Eq. (9.52) involving olivine and orthopyroxene, ΔH_r is small, and thus the exchange of iron and magnesium between these minerals is rather insensitive to temperature changes. By contrast, the exchange of iron and magnesium between coexisting orthopyroxene and augite is sensitive to temperature changes and therefore makes a good geothermometer (see, for example, Saxena, 1973).

If the temperature is held constant ($\mathrm{d}T=0$) in Eq. (9.59), we have

$$\left(\frac{\partial \ln K}{\partial P}\right)_T = -\frac{\Delta V_r}{RT} \qquad (9.61)$$

Most petrologic reactions involving condensed phases (solids and liquids) have very small values of ΔV_r. Consequently, changes in the equilibrium constant for these reactions are insensitive to changes in pressure. Very few reactions are suitable for geobarometry (Brey and Köhler, 1990). Most reactions involving gases have large ΔV_r, but the gas phase commonly escapes during reaction or we may not know its composition. In either case, we are unable to evaluate the equilibrium constant.

Let us close this chapter by examining one new and potentially very useful geothermometer, which is based on the titanium content of quartz (Wark and Watson, 2006). Although we think of quartz as being one of the purest of minerals, simply SiO_2, small amounts of Ti^{+4} (up to ~100 ppm) can substitute for Si^{+4} in its structure, especially at high temperature. New techniques of electron microprobe and ion microprobe analysis allow for accurate determination of these low concentrations and allow the distribution of Ti in quartz grains to be measured. In addition, Ti in quartz causes cathodoluminescence, which allows detailed patterns of Ti distribution to be seen (Fig. 9.9).

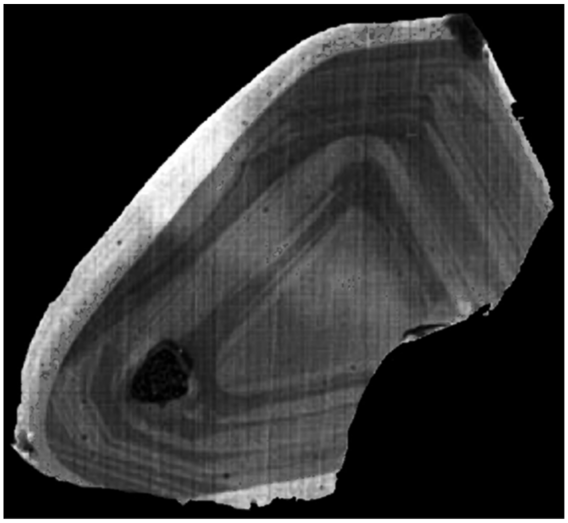


Fig. 9.9 Cathodoluminescence image of a ~1.5-mm-diameter quartz phenocryst in the Bishop Tuff, showing complex oscillatory zoning. (Photograph from Peppard *et al*., 2001; published with permission of the *American Mineralogist*.) Wark *et al*. (2007) have shown that the bright rim in the cathodoluminescent image is due to increased Ti by as much as ~50 ppm, which according to the Ti-in-quartz geothermometer indicates an increase in the magma temperature of ~70 °C just prior to eruption. This was probably caused by an influx of more mafic magma into the base of the chamber.

We start by writing a simple titanium exchange reaction between the minerals rutile and quartz,

$$\mathrm{TiO}_2^{\mathrm{rutile}} \rightleftharpoons \mathrm{TiO}_2^{\mathrm{quartz}} \qquad (9.62)$$

A chemical equilibrium constant can be written for this reaction, but because the solution of titanium in quartz is not expected to be ideal, this constant must be expressed in terms of activities; that is,

$$K = \frac{a_{\mathrm{TiO_2}}^{\mathrm{quartz}}}{a_{\mathrm{TiO_2}}^{\mathrm{rutile}}} \qquad (9.63)$$

According to Eq. (9.56), the equilibrium constant can be related to the free energy change of the reaction and temperature by

$$\ln K = \ln\left(\frac{a_{\mathrm{TiO_2}}^{\mathrm{quartz}}}{a_{\mathrm{TiO_2}}^{\mathrm{rutile}}}\right) = -\frac{\Delta G}{RT} \qquad (9.64)$$

If the rutile is relatively pure, which is generally the case in many rocks, the $a_{\mathrm{TiO_2}}^{\mathrm{rutile}}$ will be very nearly 1 (Fig. 9.3), and if we replace the $a_{\mathrm{TiO_2}}^{\mathrm{quartz}}$ with $\gamma_{\mathrm{TiO_2}}^{\mathrm{quartz}} \cdot X_{\mathrm{TiO_2}}^{\mathrm{quartz}}$ (Eq. (9.35)), Eq. (9.64) reduces to

$$\ln\left(\gamma_{\mathrm{TiO_2}}^{\mathrm{quartz}} \cdot X_{\mathrm{TiO_2}}^{\mathrm{quartz}}\right) = -\frac{\Delta G}{RT}$$

which on rearranging gives

$$\ln X_{\mathrm{TiO_2}}^{\mathrm{quartz}} = -\ln \gamma_{\mathrm{TiO_2}}^{\mathrm{quartz}} - \frac{\Delta G}{RT} \tag{9.65}$$

Because the concentration of TiO_2 in quartz is so low (ppm), the $a_{\mathrm{TiO_2}}^{\mathrm{quartz}}$ and $X_{\mathrm{TiO_2}}^{\mathrm{quartz}}$ are expected to be linearly related (Fig. 9.3), and the activity coefficient, $\gamma_{\mathrm{TiO_2}}^{\mathrm{quartz}}$, will remain constant; that is, the solution will obey Henry's law (Eq. (9.36)). Equation (9.65) therefore indicates that the natural log of the mole fraction of TiO_2 in quartz is expected to be a linear function of the reciprocal absolute temperature, with the slope being $-\Delta G/R$ and the intercept being minus the natural log of the activity coefficient.

Wark and Watson (2006) did careful experiments in which they equilibrated quartz with TiO_2 at temperatures between 600 and 1000 °C and found that the solubility of TiO_2 in quartz does, indeed, behave as predicted by Eq. (9.65). They were able to show that the temperature of quartz crystallization, or reequilibration, in equilibrium with rutile can be determined to within ±5 °C (2 sigma) by the following relation

$$T\ (^{\circ}\mathrm{C}) = \frac{-3765}{\log_{10}\left(X_{\mathrm{Ti}}^{\mathrm{quartz}}\right) - 5.69} - 273 \tag{9.66}$$

where Ti concentration in quartz ($X_{\mathrm{Ti}}^{\mathrm{quartz}}$) is in ppm by weight.

To be a useful geothermometer, a mineral must be able to achieve its equilibrium composition in a geologically reasonable time and then to be able to retain a record of that composition. For example, if quartz crystallized with the equilibrium concentration of Ti from a magma, would a grain be able to retain this composition as the rock cooled, or would the Ti diffuse out of the grain? If a metamorphic rock were formed during a period of heating, how long would it take the Ti to diffuse into quartz grains to establish the equilibrium composition? These are kinetic rather than thermodynamic questions, and the answers depend on the diffusion rate of titanium in quartz.

Cherniak *et al*. (2007) have determined the diffusion coefficient of Ti in quartz (Fig. 5.13). For example, at 500 °C, D_{Ti} is $\sim 3 \times 10^{-26}\ \mathrm{m^2\ s^{-1}}$ and at 700 °C it is $\sim 1.3 \times 10^{-22}\ \mathrm{m^2\ s^{-1}}$. If we use $2\sqrt{Dt}$ as the length scale for significant diffusion (see Fig. 5.15) then in one million years, Ti would diffuse ~2 μm at 500 °C and ~125 μm at 700 °C. Thus, in metamorphic rocks, the core of quartz grains coarser than 1 mm are not likely to be affected by Ti diffusing in from the surroundings unless the period of heating is significantly more than 1 million years or the temperature is higher. Only the margins of these grains would have the potential to equilibrate with Ti in the surroundings. During cooling, Ti will try to diffuse out of quartz grains and again the same typical length scales will be involved. The rims of crystals will tend to record lower temperatures than deeper parts of grains. The problems of interpreting the temperature record in metamorphic rocks are discussed in more detail in Chapter 22.

Magmatic quartz is likely to have the equilibrium concentration of Ti at the time of crystallization. Because magmas are typically intruded to shallower depths, their rates of cooling are faster than those of regional metamorphic rocks, and their grains are generally coarser than metamorphic quartz grains. As a result, the cores of igneous quartz grains are likely to preserve initial Ti concentrations. In rapidly quenched volcanic rocks, quartz phenocrysts can preserve complex records of temperature fluctuations resulting from periodic recharge of the magma chamber. For example, Peppard *et al*. (2001) show that quartz phenocrysts in the Bishop tuff (Fig. 9.9), which erupted during the Long Valley caldera collapse in eastern California (Hildreth and Wilson, 2007), have complex oscillatory zoning patterns that are visible in cathodoluminescent images (light given off when sample is bombarded with electrons). This brightness of the cathodoluminescent image is due in part to the abundance of titanium. Using the Ti-in-quartz geothermometer, Wark *et al*. (2007) showed that these phenocrysts record a period of heating just prior to the eruption of the Bishop tuff, which was presumably caused by an influx of hotter magma into the base of the silicic magma chamber. The preservation of steep Ti concentration gradients at magmatic temperatures (~800 °C) indicates that diffusion did not have time to smooth or eradicate these profiles. The heating is therefore interpreted to have occurred within less than 100 years of the eruption. Indeed, the recharge of the magma chamber may have been responsible for the eruption of the Bishop tuff (Hildreth and Wilson, 2007).

In deriving Eq. (9.65), it was assumed that rutile was present and that it was almost pure TiO_2, so that the $a_{\mathrm{TiO_2}}^{\mathrm{rutile}}$ could be set to one. However, if rutile is not pure, for example due to the solubility of ferric iron, then the $a_{\mathrm{TiO_2}}^{\mathrm{rutile}} < 1$. Similarly, in a magma, rutile may not be present, and again, the $a_{\mathrm{TiO_2}}^{\mathrm{magma}} < 1$. This does not negate the use of the Ti-in-quartz geothermometer, but it does require that we have some other means of determining the TiO_2 activity. The activity, for example, might be buffered by the presence of other titanium-bearing minerals such as ilmenite and ulvospinel. If the $a_{\mathrm{TiO_2}} < 1$, then Eq. (9.66) becomes

$$T\ (^{\circ}\mathrm{C}) = \frac{-3765}{\log_{10}\left(\frac{X_{\mathrm{Ti}}^{\mathrm{quartz}}}{a_{\mathrm{TiO_2}}}\right) - 5.69} - 273 \tag{9.67}$$

where $a_{\mathrm{TiO_2}}$ must be determined independently. For example, in the magma that erupted to form the Bishop tuff, Wark *et al*. (2007) estimated that the $a_{\mathrm{TiO_2}}$ was ~0.6. Even when an accurate determination of the $a_{\mathrm{TiO_2}}$ cannot be made, the geothermometer may still give useful relative temperatures.

One final point to be considered is whether the Ti-in-quartz geothermometer is pressure sensitive, which would negate its use if the pressure at which the quartz crystallized was unknown. The volume change in the exchange reaction (Eq. (9.62)) is likely to be small, which would suggest that pressure would have little effect on the equilibrium distribution (Eq. (9.61)). The Ti-in-quartz diffusion experiments of Cherniak *et al*. (2007), which were done at one atmosphere, and those of Wark *et al*. (2006), which were done at 1 GPa, indicate no pressure effect on the solubility of Ti in quartz. The solubility of Ti in quartz is therefore a useful geothermometer.

We conclude that the equilibrium constant provides a simple means of studying reactions in petrology. Even when a lack of thermodynamic data makes quantitative evaluation of the constant impossible, it is still feasible to use the constant in a qualitative way to order observations on the compositions of minerals. For example, many useful studies have been made of the variation in composition of metamorphic minerals with increasing temperature of metamorphism, even though no thermodynamic data were available at the time of these studies. Knowledge of how the equilibrium constant is related to fundamental thermodynamic properties, however, allowed these compositional variations to be analyzed properly. We will have many occasions in the remainder of this book to refer to the equilibrium constant.

9.9 PROBLEMS

9.1 Prove that the maximum lowering of free energy resulting from mixing at constant temperature and pressure in an ideal binary solution occurs for a mole fraction of 0.5.

9.2 Forsterite and fayalite form the olivine solid solution series, which can be considered ideal. For 298 K and 10^5 Pa, calculate the free energy of 3 mol of forsterite and 2 mol of fayalite both before and after mixing. (Take care to consider the number of equivalent sites on which mixing occurs in the olivine formula.)

9.3 Plot a graph of $\mu_i - \mu_i^*$ versus X_i for a component *i* in an ideal solution at 298 K and 1000 K.

9.4 At high temperatures, the alkali feldspars (Ab–Or) form a complete solid solution series. The solution, however, is far from ideal, and at low temperatures it unmixes into Na-rich and K-rich feldspars (perthite). The critical temperature on the solvus is given by Waldbaum and Thompson (1969) as $T_c\,(\mathrm{K}) = 921.23 + 13.4607P_c\,(\mathrm{kbar})$. Assuming a symmetric regular solution model, calculate the interchange energy, *W*, for this solid solution series at 1 bar.

9.5 Using an interchange energy of 15.5 kJ mol^{-1} in a symmetric regular solution model of alkali feldspars, calculate (**a**) the expected compositions of coexisting alkali feldspars on the solvus at 700 K and 1 bar, and (**b**) the compositions of coexisting feldspars on the spinode at 700 K, ignoring any strain energies in the crystal.

9.6 The thermodynamic properties of the alkali feldspar solid solution series are better accounted for by an asymmetric rather than symmetric regular solution model (Waldbaum and Thompson, 1969). The two interchange energies, W_{Ab} and W_{Or}, are found to be functions of *T* and *P*

$$W_{\mathrm{Ab}} = 26471 + 0.3870P - 19.381T \quad \mathrm{J\ mol^{-1}}$$
$$W_{\mathrm{Or}} = 32099 + 0.4690P - 16.136T \quad \mathrm{J\ mol^{-1}}$$

where *P* is measured in bars (10^5 Pa) and *T* is in kelvin.

- **(a)** Plot ΔG_{mix} for the alkali feldspars as a function of mole fraction of orthoclase, X_{Or}, at 700 K and 1 bar. From the graph, determine the compositions of the coexisting alkali feldspars on the solvus, and compare these compositions with those obtained in Problem 9.5a.
- **(b)** From the plot in part (a), determine μ_{Ab} and μ_{Or} in the two feldspars coexisting on the solvus at 700 K and 1 bar relative to their values in the pure end members.
- **(c)** From a plot of $\mu_{Ab} - \mu_{Ab}^*$ versus X_{Or}, determine the composition of the coexisting feldspars on the spinode at 700 K and 1 bar.

9.7 Derive Eq. (9.49) from Eq. (9.42).

9.8 The spinode is determined by the inflection points on the ΔG versus *X* plot. From Eq. (9.42), determine the positions of the inflection points and develop the expression in Eq. (9.50) relating temperature, mole fraction, and interchange energy.

9.9 Many rocks contain coexisting plagioclase and alkali feldspar. Equilibrium between these two minerals involves the exchange of the albite component between the plagioclase and alkali feldspar structures; that is, $\mathrm{Ab^{Plag}} \rightleftarrows \mathrm{Ab^{Alk}}$. Assuming ideal mixing in both feldspar types at high temperature, show through fundamental thermodynamic relations how the equilibrium constant for this exchange reaction could be used as a geothermometer.

9.10 Phenocrysts of quartz toward the top of the Bishop tuff, which erupted from the Long Valley caldera in California, are enriched in Ti towards their rims, with concentrations increasing from 55 to 95 ppm by weight (Wark *et al*., 2007). If the activity of $\mathrm{TiO_2}$ in the magma at the time of eruption was 0.6, what increase in temperature is indicated by these Ti-enriched rims?

10 Phase equilibria in igneous systems

10.1 INTRODUCTION

When first encountered, the great variety of igneous rocks and the large number of names and textural terms used to describe them can be bewildering. Diagrams showing the relations between mineralogical parameters, such as the composition of feldspar, abundance of ferromagnesian minerals, and so on, or the variation between major chemical constituents, provide a means of attaching names to rocks, but they do not, in themselves, provide an explanation for the compositions and frequency of occurrence of rock types. These diagrams can be likened to the periodic table of the elements; they provide a means of classifying, but the underlying important factors justifying such a classification have to be sought elsewhere. Thermodynamics provides the rationale for rock compositions.

Ideally, when a magma crystallizes at a given pressure and temperature, an assemblage of minerals is formed that provides the minimum free energy possible for that particular bulk composition under the existing conditions. A close approach to this ideal situation is common in igneous rocks, particularly plutonic ones, and even when equilibrium is not achieved, mineral assemblages can still be understood in terms of reactions that are striving to bring about this minimization of free energy. Thermodynamics, therefore, provides a simple explanation for the mineralogical composition of igneous rocks, and as will be seen later, it can also be used to account for the composition of magmas.

Just as a magma, when crystallizing, strives to create an assemblage of minerals that has the minimum free energy for the given conditions of pressure and temperature, the MELTS program calculates such an assemblage for any given composition of magma using an internally consistent set of thermodynamic data (Ghiorso and Sack, 1995; Asimow and Ghiorso, 1998; Ghiorso, *et al*., 2002). The program can also calculate these assemblages for given conditions of temperature and volume (isochoric), enthalpy and pressure (isenthalpic), or entropy and pressure (isentropic), each of which allows us to investigate the crystallization of magma under special conditions. These special conditions will be discussed later, but first it is important to have a clear understanding of the possible phase relations during the crystallization of magma. Because these relations can be complex, it is convenient to have a simple graphical way of representing them. This is done through the construction of phase diagrams. Although graphical representations are limited to three dimensions (usually only two on a sheet of paper), which limits us to being able to represent only two- and three-component systems, we are able to illustrate graphically all of the possible phase relations that occur during the crystallization of more complex systems such as magmas.

An excellent explanation of phase diagrams and many examples of diagrams that are of petrologic interest can be found in *Phase Diagrams for Ceramists*, published by the American Ceramic Society (Levin *et al*., 1964, and later supplements). General discussions of petrologic phase diagrams can be found in the books by Ernst (1976) and Morse (1980).

10.2 TWO-COMPONENT SYSTEMS

Magmas are simply high-temperature solutions of the Earth's most abundant elements. The thermodynamic properties of solutions (Chapter 9) should therefore provide a basis for analyzing the stability and composition of magmas. This task, at first, may appear formidable, because the typical magma contains 10 major oxides and many minor elements. But most processes involved in the crystallization of magma can be described in terms of only two or three components thus greatly simplifying the task. We shall first consider two-component systems.

In Problem 8.2, you were asked to calculate the melting point of diopside from the entropies and enthalpies of the solid and liquid forms of this mineral. The solution, which involves determining the temperature at which the free energy of the liquid and solid are equal, is illustrated in Figure 10.1. In a plot of free energy versus temperature, both crystalline and liquid diopside have negative slopes $[(\partial G/\partial T)_P = -S]$, but that of the liquid is steeper because it has the larger entropy. Consequently, at high temperatures the liquid has the lower free energy and is stable, whereas at low temperatures the solid has the lower free energy and is stable. The two curves intersect at the melting point (T_m) of diopside.

At some temperature T below the melting point, the free energy of liquid diopside is higher than that of solid crystalline diopside. Although rapid cooling to this temperature might temporarily preserve liquid diopside in a metastable, supercooled (glassy) state, crystallization brings about a lowering of free energy by an amount ΔG. Under equilibrium conditions, therefore, liquid diopside can never be stable at temperatures below the melting point, unless some way exists

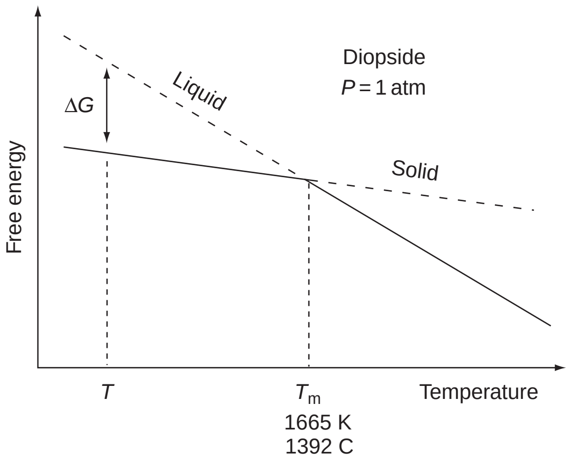


Fig. 10.1 Free energy versus temperature plot at constant pressure for solid and liquid diopside. T_m is the melting point of diopside, where the free energies of liquid and solid diopside are equal.

of lowering the free energy of the liquid to the value of the solid. Solutions provide a means of doing this.

If we assume, for the moment, that magma containing diopside in solution behaves ideally, Eq. (9.28) indicates that the chemical potential of diopside in that magma is given by

$$\mu_{\mathrm{Di}}^{\mathrm{L}} = \mu_{\mathrm{Di}}^{*\mathrm{L}} + RT \ln X_{\mathrm{Di}}^{\mathrm{L}} \tag{10.1}$$

where $\mu_{\mathrm{Di}}^{*\mathrm{L}}$ is the chemical potential of diopside in a pure diopside liquid at the specified temperature and pressure. Note that if the solution becomes pure diopside, the mole fraction of diopside in the liquid ($X_{\mathrm{Di}}^{\mathrm{L}}$) becomes 1, its logarithm becomes zero, and $\mu_{\mathrm{Di}}^{\mathrm{L}} = \mu_{\mathrm{Di}}^{*\mathrm{L}}$. In any other solution, however, X is a fraction, its logarithm is negative, and $\mu_{\mathrm{Di}}^{\mathrm{L}} < \mu_{\mathrm{Di}}^{*\mathrm{L}}$. As long as the solution remains ideal, the lowering of chemical potential is independent of the substance causing the dilution. For instance, molten diopside could be diluted with molten olivine or molten anorthite, and the chemical potential of diopside would depend only on the mole fraction to which diopside was diluted.

At temperature T below the melting point of pure diopside (Fig. 10.1), it should be possible to dilute molten diopside so that its chemical potential is ΔG lower than that of pure molten diopside at this temperature. The chemical potential of molten diopside would then be the same as that of pure crystalline diopside ($\mu_{\mathrm{Di}}^{\mathrm{S}}$) at this temperature, and equilibrium would be established. Note that we are considering only the case where the crystalline phase is pure; that is, crystals of diopside exhibit no solid solution with the component that dissolves in the magma. Many important rock-forming minerals, of course, do form solid solutions; these will be considered in Section 10.9. The equilibrium between pure crystalline diopside and molten diopside in solution can be expressed as

$$\mu_{\mathrm{Di}}^{\mathrm{S}} = \mu_{\mathrm{Di}}^{\mathrm{L}} = \mu_{\mathrm{Di}}^{*\mathrm{L}} + RT \ln X_{\mathrm{Di}}^{\mathrm{L}} \tag{10.2}$$

This indicates that although pure, molten diopside can stably coexist with crystalline diopside at only one temperature (the melting point of pure diopside at a specified pressure), diopside-bearing solutions can be in equilibrium with diopside over a range of temperatures, as long as there is appropriate dilution of diopside in the melt.

To determine the temperature range over which this solid–liquid equilibrium can exist, we first rearrange Eq. (10.2) as follows:

$$RT \ln X_{\mathrm{Di}}^{\mathrm{L}} = \mu_{\mathrm{Di}}^{\mathrm{S}} - \mu_{\mathrm{Di}}^{*\mathrm{L}} \tag{10.3}$$

But $\mu_{\mathrm{Di}}^{\mathrm{S}} - \mu_{\mathrm{Di}}^{*\mathrm{L}}$ is simply the negative of the free energy change per mole accompanying the fusion of pure diopside at this temperature; that is, ΔG in Figure 10.1. Hence we can write

$$RT \ln X_{\mathrm{Di}}^{\mathrm{L}} = -\Delta G_{\mathrm{m}} = -(\Delta H_{\mathrm{m}} - T\Delta S_{\mathrm{m}}) \tag{10.4}$$

where the subscript m stands for melting. On rearranging and simplifying, we obtain

$$\ln X_{\mathrm{Di}}^{\mathrm{L}} = -\frac{\Delta H_{\mathrm{m}}}{RT} + \frac{\Delta S_{\mathrm{m}}}{R} \tag{10.5}$$

At the melting point of pure diopside the ΔG of melting is, of course, zero, and thus $\Delta S_{\mathrm{m}} = \Delta H_{\mathrm{m}}/T_{\mathrm{m}}$, where T_{m} is the melting point of pure diopside. For many reactions, such as the melting of a substance, the value of ΔS varies little with temperature. Thus, over a limited temperature range, $\Delta H_{\mathrm{m}}/T_{\mathrm{m}}$ can be substituted for ΔS_{m} without introducing serious error, and Eq. (10.5) becomes

$$\ln X_{\mathrm{Di}}^{\mathrm{L}} = \frac{\Delta H_{\mathrm{m}}}{R}\left(\frac{1}{T_{\mathrm{m}}} - \frac{1}{T}\right) \tag{10.6}$$

This is known as the *cryoscopic equation*, because it describes the freezing-point (melting-point) depression of a substance brought about by its dilution in a solution.

A plot of the cryoscopic equation is given for diopside in Figure 10.2(A). The melting point of pure diopside is 1392 °C (1665 K), and its latent heat of fusion is 142.6 kJ mol^{-1} (Navrotsky *et al.*, 1980). Dilution of diopside to a mole fraction of 0.75 lowers its melting point to 1347 °C. By diluting the diopside more, still lower temperatures are achieved. The resulting line in this plot is referred to as the *liquidus* of diopside. At temperatures and compositions above the liquidus of a particular phase, that phase must be entirely in solution (molten). At the liquidus, however, the solution is saturated in the phase, and precipitation (crystallization) can occur with further cooling.

The latent heat of fusion of a phase determines the slope of the liquidus of that phase. For example, diopside (ΔH_{m} = 142.6 kJ mol^{-1}) has its melting point lowered 45 °C by dilution to a mole fraction of 0.75. Cristobalite, on the other hand, with a latent heat of fusion of only 8.2 kJ mol^{-1}, has its melting point lowered 737 °C by this same dilution. It should be noted, however, that these minerals have very different molecular weights (Di = 0.2166 kg; Cr = 0.0601 kg). Had

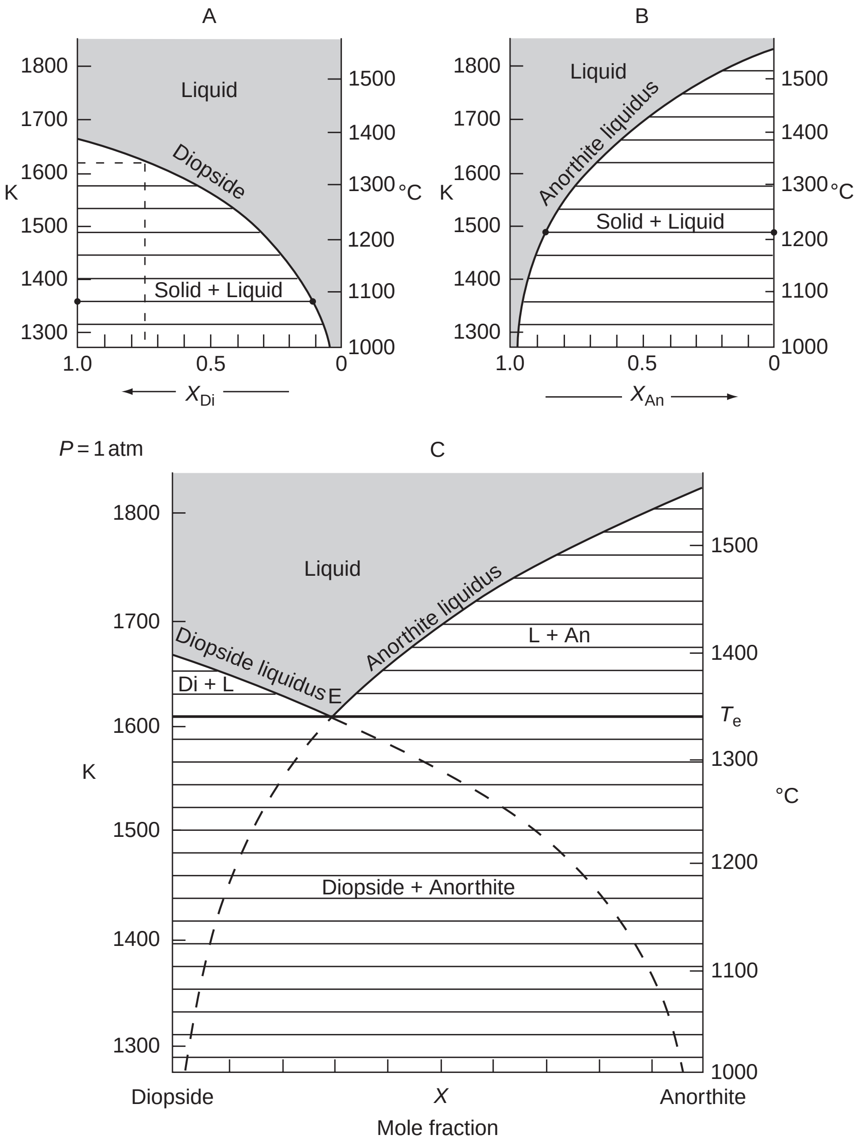


Fig. 10.2 (**A**) Diopside liquidus plotted as a function of mole fraction of diopside in melt. (**B**) Similar plot for liquidus of anorthite. (**C**) Calculated phase diagram for system diopside–anorthite at atmospheric pressure. The lowest possible melting mixture, the eutectic, is marked by the letter E.

equivalent weight fractions been considered, the difference in melting-point depression would have been very small.

So far, the material causing the dilution of the diopside in the melt has not entered into consideration, except for the stipulation that it cannot form a solid solution with diopside. As the melting point of diopside is lowered, the concentration of the solvent increases, and a stage is reached where the solvent becomes so concentrated that it too starts to precipitate (crystallize).

Let us consider, for illustrative purposes, diluting diopside with molten anorthite. Other minerals could be chosen, but this combination forms an important two-component (diopside–anorthite) solution that approximates the compositions of basalts. The lowering of the melting point of diopside, then, requires increasing the concentration of anorthite in the melt. This is feasible only if anorthite is capable of forming melts at these temperatures. Pure anorthite, however, can form a liquid only above 1557 °C, but its melting point can be lowered, according to the cryoscopic equation, by diluting the anorthite in the liquid. In our example, this can be done by dissolving diopside in the melt. The latent heat of fusion of anorthite is poorly known, but lies between 125 and 140 kJ mol^{-1} (Navrotsky *et al*., 1980). The liquidus for anorthite in Figure 10.2(B) is obtained using a value of 136 kJ mol^{-1}.

The abscissas in Figure 10.2(A) and (B) are, in fact, the same. Because we are dealing with a two-component solution, a decrease in the mole fraction of diopside must be accompanied by a corresponding increase in the mole fraction of anorthite; that is, $X^{\mathrm{L}}_{\mathrm{Di}} = 1 - X^{\mathrm{L}}_{\mathrm{An}}$. We, therefore, can combine these diagrams to form what is known as a two-component or *binary T–X phase diagram* (Fig. 10.2(C)).

In this diagram, the lowering of the melting point of diopside is seen to be accompanied by enrichment of the melt in anorthite. Indeed, when the liquidus of diopside has decreased to 1337 °C, the melt has become so enriched in this component that the liquidus of anorthite is encountered. Any further attempt to descend the diopside liquidus would be accompanied by crystallization of anorthite. This would prevent the liquid becoming more enriched in this component and thus prevent any further lowering of the melting point of diopside. These arguments apply equally to the lowering of the melting point of anorthite due to dilution by diopside. Our conclusion, therefore, must be that under equilibrium conditions and at the pressure specified for the diagram, the temperature of the liquid cannot be lowered beyond the point of intersection of the liquidus lines for the phases involved. This point of intersection is known as the *eutectic* and is marked with an E in Figure 10.2(C). Below the eutectic, the liquidus of a phase is metastable and can be followed only if the other phase fails to crystallize due to kinetic factors. For this reason the liquidus lines of diopside and anorthite are dashed below the eutectic.

In deriving the cryoscopic equation, the solution was taken to behave ideally. This, of course, may not always be valid. Indeed, many petrologically important systems exhibit strong deviations from ideality. The cryoscopic equation, however, can still be used if we substitute activity (a_i) for mole fraction (X_i). It will be recalled from Section 9.5 that the activity can be thought of as the effective mole fraction of a component in solution. Activity is related to the mole fraction by the activity coefficient ($\gamma_i X_i = a_i$). For a nonideal solution the cryoscopic equation can be written as

$$\ln a_{\mathrm{Di}}^{\mathrm{L}} = \ln \gamma_{\mathrm{Di}} X_{\mathrm{Di}}^{\mathrm{L}} = \frac{\Delta H_{\mathrm{m}}}{R}\left(\frac{1}{T_{\mathrm{m}}} - \frac{1}{T}\right) \tag{10.7}$$

Figure 10.4 presents the experimentally determined phase diagram for the system diopside–anorthite. Compositions in this diagram are plotted in terms of weight percent rather than mole fractions. Once this difference is accounted for (Problem 10.1), comparison with Figure 10.2(C) reveals differences that indicate diopside–anorthite solutions are not strictly ideal. Differences between the calculated and experimental data can be used to evaluate the activity coefficients of diopside and anorthite in these melts (see Problem 10.2).

10.3 LEVER RULE

Before continuing with the discussion of phase diagrams, it is necessary to describe how compositions are read from such diagrams. In Figure 10.2(C), compositions are plotted along the abscissa, with diopside at one end and anorthite at the other. Any point in the diagram, regardless of the temperature, represents a composition that can be read directly from the abscissa. Other compositional information, such as the proportion of liquid to crystals, can also be read from these diagrams, if they are plotted in weight rather than mole fractions. Weight is a conservative measure of composition and is therefore used in most phase diagrams.

If a line AB (Fig. 10.3) represents compositions of mixtures of two components A and B, any point f along this line represents specific weight fractions of A and of B. Clearly, if f were placed at the midpoint of the line, it would indicate equal amounts of A and B. If, on the other hand, f were closer to A, as shown in Figure 10.3, it would represent a composition richer in A, and conversely, if it were closer to B, it would be richer in B. The actual fractions of A and B in any composition f are determined from the relative lengths of the lines fB and fA, respectively. The fraction of A is given by the ratio fB/AB, and the fraction of B, by the ratio fA/AB. Thus the longer the line fB, the greater is the amount of A present, and the longer the line fA, the greater is the amount of B present. Because of the analogy of this relation to the moments of a lever about a fulcrum (Fig. 10.3), this way of reading compositions is commonly referred to as the *lever rule*. for example, for a lever we could state that the weight A is to the length L_2 as the weight B is to the length L_1. By analogy, then, the weight of A in composition f is to the line fB as the weight of B in f is to the line fA.

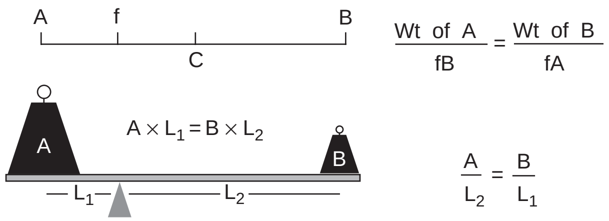


Fig. 10.3 Graphical representations of the abundance of components in a binary mixture is analogous to the position of a fulcrum on a lever to the weights at each end of the lever. See text for discussion.

It should be noted that the absolute length of a line in such a diagram is not important, for it is the ratio of lengths that conveys the compositional information. Consider the point f on the line AB; the ratio fA/AB is 0.25, and thus the composition represented by point f is composed of 25% B and 75% A. But this same composition f could be expressed in terms of A and C, the latter being some compound with a composition between A and B. In this case, the ratio fA/AC is 0.5, and therefore, composition f contains 50% C and 50% A. In both cases, the absolute length of line fA remains the same, but the percentage of A is different.

Before continuing with the discussion of phase diagrams, the lever rule must be fully appreciated, and you should become proficient at graphically manipulating compositions. It is strongly recommended that Problem 10.3 be done before proceeding to the next section.

10.4 SIMPLE BINARY SYSTEMS WITH NO SOLID SOLUTION

Let us return to a consideration of the binary-phase diagram for the system diopside–anorthite (Fig. 10.4). Along the liquidus of anorthite, melts are saturated in anorthite, and cooling brings about crystallization of this mineral. Similarly, the diopside liquidus is the saturation curve for diopside. At the eutectic, where the two liquidus lines intersect, both anorthite and diopside crystallize simultaneously upon withdrawal of heat. The heat removed comes from the crystallization of diopside and anorthite (latent heat of crystallization = $-\Delta H_{\mathrm{m}}$) and does not cause a lowering of the eutectic temperature. Indeed, as will be shown later, the temperature of a

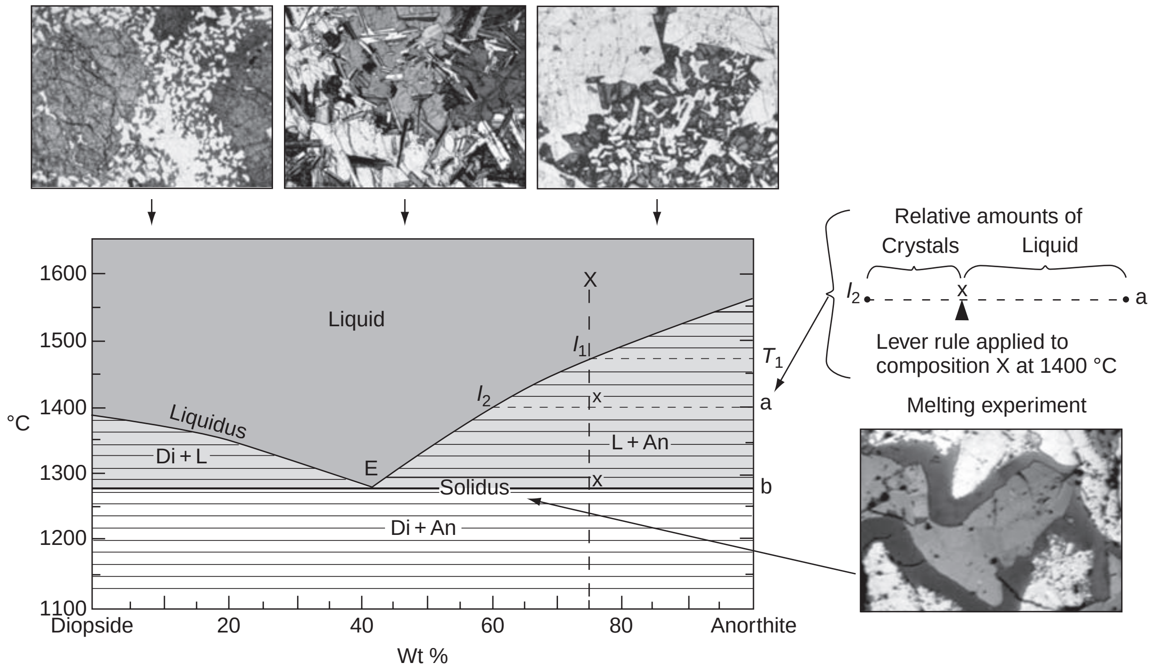


Fig. 10.4 System diopside ($CaMgSi_2O_6$)–anorthite ($CaAl_2Si_2O_8$) at 1 atm pressure. Slight deviations from a binary system for diopside-rich compositions have been ignored. (After Weill *et al.*, 1980.) Photomicrographs illustrate typical textures in gabbroic rocks resulting from crystallization in different regions of this system. Clockwise from top left: augite phenocrysts surrounded by ophitic (eutectic) intergrowth of augite and plagioclase; ophitic intergrowth of simultaneously intergrown augite and plagioclase; plagioclase phenocrysts surrounded by ophitic intergrowth of plagioclase and pyroxene; polished section in reflected light showing melt (dark) formed between plagioclase (gray) and augite (white) in an experimentally heated sample of basalt.

crystallizing liquid at a eutectic, at a specified pressure, remains constant until all of the liquid has crystallized. The eutectic liquid, then, is converted into crystals of anorthite and diopside, both of which must, if there is equilibrium, be at the same temperature as the liquid. Through the point E in Figure 10.4, we, therefore, can draw a horizontal line (constant temperature) that joins the composition of the eutectic liquid to the two phases crystallizing from it. This line is known as the *solidus*.

The liquidus and solidus divide the phase diagram into a number of regions. Above the liquidus the system is entirely liquid, and below the solidus it is completely solid. Between these, however, both liquid and crystals coexist. The eutectic in the diopside–anorthite system divides this region in two; on one side, liquid coexists with anorthite and on the other, it coexists with diopside. These various regions consist of either one or two phases and are labeled with the phases present. In one-phase regions, such as that of the liquid in this diagram, any bulk composition forms a single phase with a composition identical to that of the bulk composition. In two-phase regions, on the other hand, a bulk composition consists of two phases, the compositions of which are indicated by the ends of the horizontal (isothermal) lines drawn in that region. Thus, within the region marked liquid plus anorthite (L + An), any bulk composition consists of liquid and anorthite, with the composition of the liquid being given by the liquidus of anorthite at the temperature of interest. Below the solidus, bulk compositions consist of crystals of diopside and anorthite, the compositions of which are represented by the two sides of the diagram.

Let us consider the cooling and crystallization of a liquid consisting of 75% anorthite and 25% diopside (x in Fig. 10.4). First, we construct a vertical line to mark the bulk composition. Such a line is known as an *isopleth*, a line of constant composition. Once the isopleth is drawn, the crystallization sequence can be read directly from the phase assemblages encountered by the isopleth as it descends through each phase region. This, of course, assumes that the bulk composition remains unchanged during the cooling; that is, nothing is added to or subtracted from the system.

On cooling, the liquid first encounters the liquidus of anorthite at l_1 and temperature T_1. Further cooling, under equilibrium conditions, brings about crystallization of anorthite, which in turn enriches the melt in diopside. But for the liquid to remain in equilibrium with crystals of anorthite, the enrichment in diopside must be accompanied by cooling at a rate that keeps the liquid on the anorthite liquidus.

When the temperature has fallen to 1400 °C, the composition of the liquid will have changed to l_2. It is instructive to consider what fraction of the original liquid will have crystallized by this temperature. We know that the bulk composition of the system has remained unchanged (the isopleth), and that at 1400 °C liquid with composition l_2 coexists with anorthite crystals. The lever rule indicates that the fraction of liquid in this system is xa/l_2a and that the fraction of anorthite crystals is xl /l a. The ratio of liquid to crystals at this temperature is, therefore, xa/xl_2. This proportion is readily apparent from inspection of the phase diagram.

With falling temperature, the proportion of liquid to crystals decreases as the liquid changes its composition along the liquidus, eventually reaching a ratio of xb/xE at the eutectic. At this point, however, the onset of diopside crystallization prevents the composition of the liquid from changing further. Because the eutectic liquid crystallizes both diopside and anorthite and yet does not change its composition, these phases must crystallize in precisely the same proportions as they are present in the melt, that is 58% diopside and 42% anorthite. Once all the eutectic liquid has crystallized, the solids, which then consist of 75% anorthite and 25% diopside, are free to cool.

In the treatment above, equilibrium was maintained between crystals and liquid at all times, and thus cooling of the melt gave rise to what is referred to as *equilibrium crystallization*. If, on the other hand, crystals are removed from or otherwise prevented from being in equilibrium with the melt from which they form, *fractional crystallization* results. This could happen, for example, if crystals were much denser than the liquid; they could then separate by sinking. Such fractionation effectively changes the bulk composition, depleting it in the phase that is separating. Of course, equilibrium and complete fractional crystallization are extremes, and in nature, we can expect to find all possible gradations between these.

Fractional crystallization need not take place continuously but can occur in stages. For example, melt with a composition of 75% anorthite and 25% diopside could cool, under equilibrium conditions, to 1400 °C at which temperature the melt would have composition l_2. At this stage, all of the crystals could be removed, leaving only liquid l_2. This would provide a new starting composition through which could be constructed another isopleth, which is closer to the eutectic than the original. Regardless of whether fractionation is stepwise or continuous, its effect is always to produce liquids that are closer in composition to the eutectic.

The phase assemblages encountered on heating a solid mixture of 75% anorthite and 25% diopside until completely molten are the reverse of those encountered on cooling a liquid of this composition. Solids heated to the eutectic temperature begin to melt with the formation of a eutectic liquid. Only liquid of this composition can exist at this low temperature. Because diopside and anorthite both contribute to the liquid, melting occurs only where these different grains are in contact (see photo in lower right of Fig. 10.4). The solids are consumed in the same proportions as they are present in the eutectic (58% Di, 42% An). Because the starting composition is rich in anorthite, there comes a time when the last grain of diopside melts, and the system then consists of eutectic liquid and anorthite crystals. The temperature has remained constant at the eutectic temperature throughout this melting episode, but with the disappearance of diopside, it is again free to rise. Continued heating causes the remaining anorthite to dissolve, changing the melt's composition along the liquidus. When the melt reaches composition l , all anorthite has dissolved, and there can be no further enrichment in this component. The melt, consequently, leaves the liquidus and rises along the isopleth with constant composition.

Let us now examine this phase diagram in terms of what it can tell us about the origin of igneous rocks. First, and perhaps most important, although separate minerals have relatively high melting points, mixtures of minerals melt at considerably lower temperatures, especially when present in eutectic proportions. It is not surprising, then, to find that in the largely solid outer part of the Earth, most igneous rocks have compositions close to eutectics (e.g. the melt in the experimentally melted basalt in the lower right of Fig. 10.4). They are the products of an Earth whose limited thermal energy is sufficient to melt only the lowest-melting fractions. Basalts, for example, have compositions close to the diopside–anorthite eutectic, and granites, as we will see later, have compositions close to the eutectic between quartz and alkali feldspar (Fig. 10.5). Note that regardless of the proportion of anorthite to diopside in a region undergoing partial melting, the first-formed melt must be of eutectic composition. Even if there is sufficient heat to produce liquids of noneutectic composition, these liquids, on cooling, will descend the liquidus toward the eutectic and fractionation can again produce eutectic liquids. The first important conclusion from phase diagrams, then, is that *many common igneous rocks have near-eutectic compositions*.

Any liquid with a composition on the anorthite side of the eutectic, such as the one discussed previously, first crystallizes anorthite. Indeed, if the composition is far from the eutectic, anorthite crystallizes over a considerable temperature interval before diopside appears at the eutectic. The early-formed anorthite crystals are consequently larger than later-formed crystals, and a *porphyritic texture* results, with *phenocrysts* of anorthite (Fig. 10.4). Phenocrysts are formed from the phase whose liquidus is first encountered by the cooling magma. If volcanic eruption of a melt on the anorthite liquidus results in rapid quenching, the liquid may never crystallize diopside, and the phenocrysts may instead be surrounded by glass. The texture is then said to be *vitrophyric*. Had the melt first encountered the diopside liquidus, phenocrysts of this mineral would have formed (Fig. 10.4). Regardless of which phase forms the phenocryst, the surrounding liquid is always enriched in the eutectic composition.

An interesting situation develops at the eutectic, for here both minerals crystallize together. Unlike phenocrysts, which are free to develop their own crystal faces, minerals forming at the eutectic must accommodate each other's growth habits. This commonly results in characteristic textures known as *eutectic intergrowths*. Plagioclase and pyroxene, for example, form what is referred to as an *ophitic texture* (Fig. 10.4), with plagioclase laths embedded in large single crystals of pyroxene. A particularly striking texture involving quartz and alkali feldspar is known as. *graphic granite* (Fig. 10.5), because the grains of quartz, which are embedded in large single crystals of feldspar, have shapes that resemble the characters of cuneiform (wedge-shaped) writing.

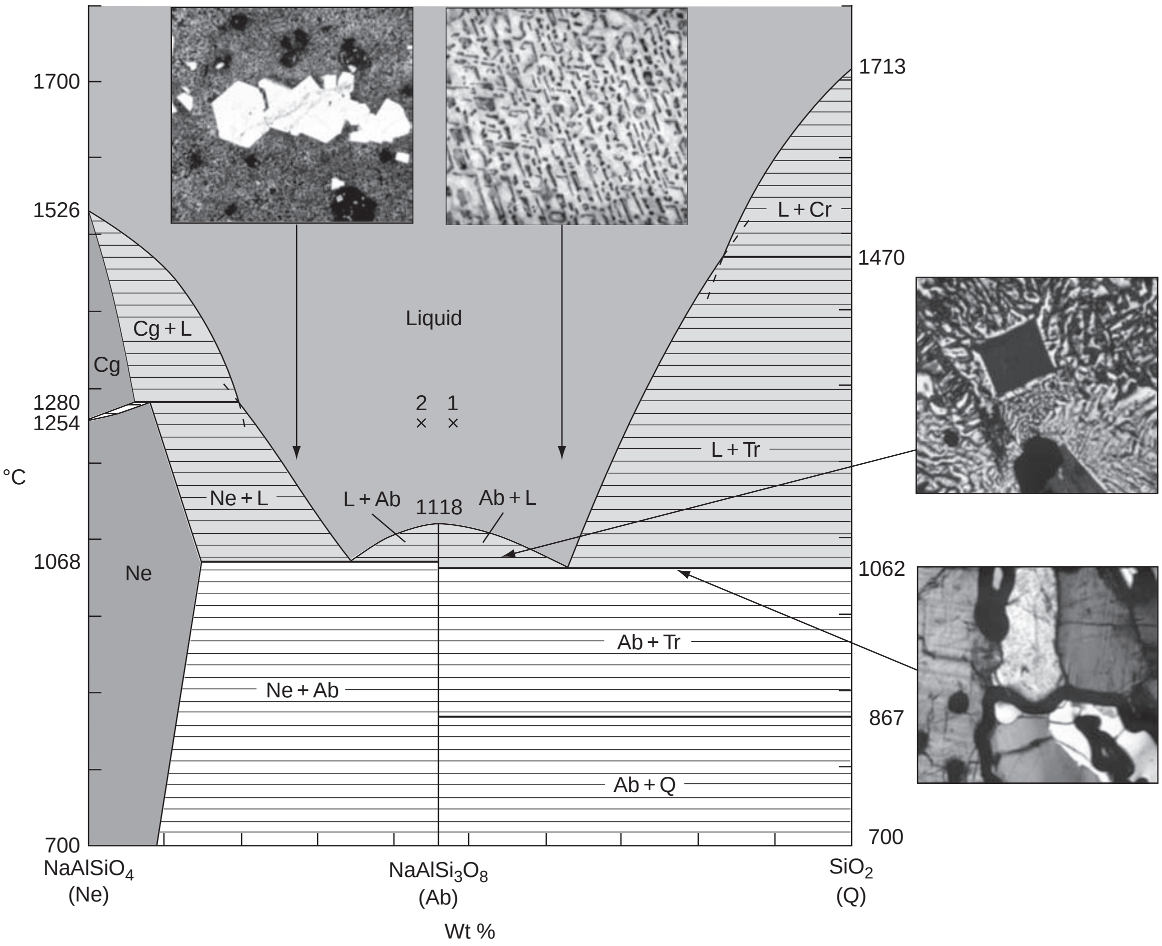


Fig. 10.5 System NaAlSiO$_4$ – SiO$_2$. Phases are cristobalite (Cr), quartz (Q), tridymite (Tr), albite (Ab), nepheline (Ne), and carnegieite (Cg). (After Greig and Barth, 1938; Tuttle and Bowen, 1958.) Photomicrographs in clockwise order from top left are of phenocrysts of nepheline, graphic granite, granophyric intergrowth of quartz surrounding plagioclase crystals (dark) in the final crystallizing residue of a basalt, and of experimentally produced partial melt between grains of quartz and feldspar – the melt is quenched to an isotropic glass.

In Chapter 2, the common occurrence of phenocrysts in igneous rocks was used as evidence that most magmas are not superheated (Bowen, 1928). We are now in a position to attach specific numbers to this statement. The presence of phenocrysts indicates that a magma is on the liquidus and not at some higher temperature. In the system diopside–anorthite, then, the presence of phenocrysts would indicate that the temperatures could not be more than 280 °C above the eutectic. We might, therefore, estimate that basaltic magmas have temperatures near 1300 °C. Such an estimate would be slightly high (actual temperatures are 1100 to 1200 °C) because natural pyroxenes contain some iron and plagioclase contains considerable sodium, both of which lower melting temperatures. The estimate is, nonetheless, of the correct order of magnitude.

Although discussion of more complex systems will allow us to account for other features of igneous rocks, the explanations provided by the simple binary system for the compositions of common igneous rocks, some of their main textures, and their temperatures of formation are of paramount importance. Later phase diagrams should be examined with these same points in mind.

10.5 PHASE RULE

In our discussion of crystallization in the diopside–anorthite system, reference was made to the fact that temperature remains constant at the eutectic while crystallization takes place. Although heat is continually being removed, the temperature does not change because heat is generated by the crystallization process. A more familiar example of this phenomenon is provided by water that has been cooled to the point that ice begins to form; as long as both water and ice are present and the pressure is 1 atm, the temperature remains at 0 °C. Another restriction encountered in the diopside–anorthite system was the fixed composition of the eutectic liquid. In fact, equilibrium between phases imposes numerous restrictions on variations in temperature, pressure, and composition. These relations can be quantified by the *phase rule*, which was discussed briefly in Section 8.5.

In Section 8.1, we saw that variables can be classified as extensive or intensive. At equilibrium, the intensive ones, T, P, and μ_i, must be constant throughout the system. In any phase, however, changes in these variables are related by the Gibbs Duhem equation (Eq. (9.7)). For example, in the two-component system diopside–anorthite, we can write for the liquid phase

$$S^{\mathrm{L}}\mathrm{d}T - V^{\mathrm{L}}\mathrm{d}P + n^{\mathrm{L}}_{\mathrm{Di}}\mathrm{d}\mu_{\mathrm{Di}} + n^{\mathrm{L}}_{\mathrm{An}}\mathrm{d}\mu_{\mathrm{An}} = 0 \tag{10.8}$$

If no other phase is present, we have only this one equation and four unknowns. As soon as any three variables are specified, the fourth is determined by the equation. We say, therefore, that there are three degrees of freedom, or that the variance of the system is 3. As long as only one phase is present, whether it is liquid or some other phase, there is only one Gibbs–Duhem equation and four unknowns. Thus, a single phase always gives a variance of three in a two-component system, and thus pressure, temperature, and composition can all be varied independently.

If two phases are present in a two-component system, for example liquid and diopside, Gibbs–Duhem equations can be written for both phases:

$$S^{\mathrm{L}}\mathrm{d}T - V^{\mathrm{L}}\mathrm{d}P + n^{\mathrm{L}}_{\mathrm{Di}}\mathrm{d}\mu_{\mathrm{Di}} + n^{\mathrm{L}}_{\mathrm{An}}\mathrm{d}\mu_{\mathrm{An}} = 0 \tag{10.8}$$

$$S^{\mathrm{DiS}}_{\mathrm{Di}}\mathrm{d}T - V^{\mathrm{DiS}}_{\mathrm{Di}}\mathrm{d}P + n^{\mathrm{DiS}}_{\mathrm{Di}}\mathrm{d}\mu_{\mathrm{Di}} + n^{\mathrm{DiS}}_{\mathrm{An}}\mathrm{d}\mu_{\mathrm{An}} = 0 \tag{10.9}$$

where the superscript S signifies solid–crystalline diopside in this case. Note that the last term in Eq. (10.9) is zero, because the diopside contains no anorthite; that is, there is no solid solution between these two minerals. We now have two equations and four unknowns. The variance of the system is therefore 2. This means that only two of the three possible variables are independently variable. For example, if the pressure is chosen to be 1 atm and the composition of the melt to be 60 wt% anorthite (l_2 in Fig. 10.4), the temperature is fixed at 1400 °C, the liquidus for these particular conditions. A special but commonly encountered condition is that of constant pressure (isobaric). For example, the phase diagram in Figure 10.4 represents the phase relations in the system diopside–anorthite at a pressure of 1 atm. If two phases are present in a two-component system, the variance of 2 is reduced to 1 by keeping the pressure constant. Such a variance is described as isobarically univariant (one degree of freedom) and appears as a line in an isobaric phase diagram. The liquidus of anorthite in Figure 10.4 is one such line. Along this line, temperature and composition are not independently variable, for as soon as one is selected, the other is determined.

If three phases are present in a two-component system, as happens at the eutectic in the system diopside–anorthite, three Gibbs–Duhem equations can be written:

$$S^{\mathrm{L}}\mathrm{d}T - V^{\mathrm{L}}\mathrm{d}P + n^{\mathrm{L}}_{\mathrm{Di}}\mathrm{d}\mu_{\mathrm{Di}} + n^{\mathrm{L}}_{\mathrm{An}}\mathrm{d}\mu_{\mathrm{An}} = 0 \tag{10.8}$$

$$S^{\mathrm{DiS}}_{\mathrm{Di}}\mathrm{d}T - V^{\mathrm{DiS}}_{\mathrm{Di}}\mathrm{d}P + n^{\mathrm{DiS}}_{\mathrm{Di}}\mathrm{d}\mu_{\mathrm{Di}} + n^{\mathrm{DiS}}_{\mathrm{An}}\mathrm{d}\mu_{\mathrm{An}} = 0 \tag{10.9}$$

$$S^{\mathrm{AnS}}_{\mathrm{An}}\mathrm{d}T - V^{\mathrm{AnS}}_{\mathrm{An}}\mathrm{d}P + n^{\mathrm{AnS}}_{\mathrm{Di}}\mathrm{d}\mu_{\mathrm{Di}} + n^{\mathrm{AnS}}_{\mathrm{An}}\mathrm{d}\mu_{\mathrm{An}} = 0 \tag{10.10}$$

and yet there are still only four unknowns. The variance is consequently reduced to 1 (univariant), and at constant pressure this becomes isobarically invariant. In Figure 10.4, therefore, the eutectic is represented by a point (invariant). If pressure is thought of as a variable plotted perpendicular to the page in Figure 10.4, the eutectic is then a univariant line that pierces the 1-atm phase diagram at the isobaric invariant eutectic. There is only one temperature and composition that the eutectic can possibly have at this pressure.

Clearly, if four phases are present in a two-component system, as would happen if the eutectic assemblage in the system diopside–anorthite were joined by a gas composed of these components, the four Gibbs–Duhem equations would uniquely define all the variables. Such a phase assemblage can therefore exist at only one set of conditions of temperature, of pressure, and of compositions of liquid and gas. Such an assemblage would be truly invariant.

The relations above can be generalized for a system composed of c components and ϕ phases. The Gibbs–Duhem equation for each phase will contain terms for each of the possible intensive variables, T, P, and μ_i. There will be $c+2$ of these terms (one for each component + T + P). There will be ϕ equations (one for each phase), and thus the number of independent variables will be $c+2-\phi$. The number of degrees of freedom, or variance, which is commonly designated by the letter f, is given by

$$f = c + 2 - \phi \tag{10.11}$$

This is known as the *Gibbs phase rule*.

It will be noted that zero is the minimum possible value for the variance, and in that case $c+2=\phi$. In other words, there can never be more than $c+2$ phases, if equilibrium exists. Rocks can therefore be expected to contain relatively small numbers of minerals. When they do not, a lack of equilibrium can be suspected, and textural evidence for incomplete reactions can be sought. This point is discussed further in later sections.

10.6 BINARY SYSTEMS WITH BINARY COMPOUNDS

In many binary systems, the end components can react together to form intermediate compounds. For example, in the system nepheline–quartz (Fig. 10.5), albite forms, and at high pressure jadeite, are also present. Both of these minerals can be described in terms of the two components $NaAlSiO_4$ (nepheline) and SiO_2 (quartz), and are therefore referred to as *binary compounds* in this system.

Binary compounds are divided into two types based on the way in which they melt. If they form a liquid of their own composition, they are said to melt *congruently*. This is the type of melting already encountered with diopside and anorthite. Another large group of petrologically important minerals, however, break down on melting to form a solid and a liquid, both of which have different compositions from the mineral from which they form. These are said to melt *incongruently*, because of the compositional mismatch between the mineral and the liquid. We discuss first the simpler case of congruent melting.

At atmospheric pressure in the system nepheline–quartz, albite is the only stable binary compound. It melts congruently at 1118 °C and effectively divides the binary system in two. Indeed, we could draw two separate binary diagrams, nepheline–albite and albite–quartz, and their treatment would be exactly the same as that for the system diopside–anorthite. Certain important petrological points can be made, however, if we keep them in a single diagram.

First, the presence of a congruently melting binary compound results in two separate eutectics, in this case between albite and tridymite and albite and nepheline. These eutectics correspond closely in composition to the two common rock types: granite and nepheline syenite, respectively.

The melting point of albite produces a *thermal maximum* on the liquidus in the central part of this diagram. This maximum has a profound effect on liquid fractionation paths in the system. This can be illustrated by considering the crystallization of two liquids, 1 and 2, which have very similar compositions in terms of the binary components, but which fall on either side of the composition of albite. Cooling of liquid 1 results in the crystallization first of albite, which enriches the remaining liquid in silica. Eventually, albite and tridymite crystallize together at the “granite” eutectic. Liquid 2 also crystallizes albite first, but because of the slight difference in composition, its residual liquid becomes enriched in nepheline and eventually crystallizes at the “nepheline syenite” eutectic. Thus, from two very similar composition melts, very different final liquids are produced.

Consideration of the crystallization paths in the vicinity of this thermal maximum should convince the reader that there is no way of changing a liquid from one side of the maximum to the other by fractional crystallization. Melts that start with compositions on the silica-rich side must eventually crystallize tridymite, and those on the other side must eventually crystallize nepheline. This division, then, justifies the classifying of igneous rocks into *oversaturated* (containing quartz) and *undersaturated* (containing feldspathoids), as discussed in Chapter 6.

Some igneous bodies do, however, contain over- and undersaturated rocks. In light of the phase diagram, we can see what a serious problem this poses and why it has attracted the attention of so many petrologists. As will be seen later, many ingenious schemes have been devised to cross the thermal barrier.

Both quartz and nepheline have high-temperature polymorphs, the stability fields of which can be shown in the phase diagram. The silica polymorphs show no solid solution with albite, and thus the boundaries between their stability fields can be represented by horizontal lines. The liquidus, on intersecting the boundary between the fields of cristobalite and tridymite, should, theoretically, show an inflection as a consequence of the different latent heats of fusion of these two phases (recall that the slope of the liquidus is a function of the latent heat of fusion). The differences are, however, very small, but application of Schreinemakers rules (Section 8.4) indicates that the liquidus of the high-temperature polymorph must be steeper than that of the low-temperature one, as shown in exaggerated form in Figure 10.5. The polymorphism of nepheline is complicated by solid solution with albite. Its discussion, therefore, is deferred to Section 10.10.

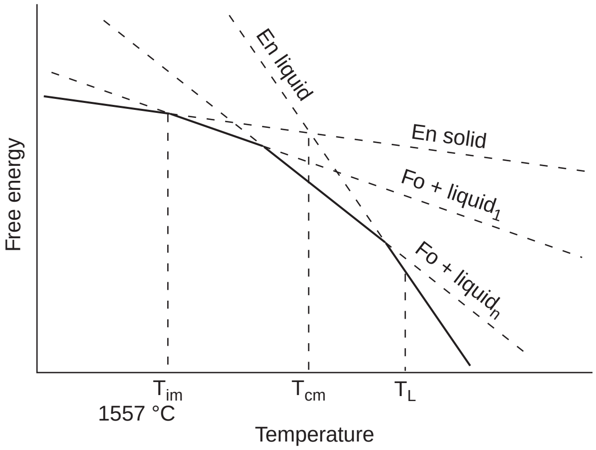


Fig. 10.6 Free energy versus temperature plot at constant pressure of solid and liquid enstatite and two compositionally equivalent mixtures of forsterite and liquid. Because these olivine–liquid mixtures have a lower free energy than that of enstatite at its congruent melting point, enstatite melts incongruently to forsterite + a more siliceous liquid.

Some common textures in felsic igneous rocks can be explained by the phase relations in this diagram (Fig. 10.5). We have already mentioned the eutectic intergrowth between quartz and alkali feldspar to produce the graphic granite texture. In many basaltic rocks, the final liquid residue has a composition near the granite eutectic. Consequently the early crystallizing plagioclase, which is anorthite rich, commonly zones out to albitic compositions on the rim where it is intergrown with quartz to form a texture resembling the graphic intergrowth but slightly less regular (inset in Fig. 10.5); this is referred to as a *granophyric* texture. Magmas rich in the nepheline component intersect the nepheline liquidus on cooling and consequently contain phenocrysts of nepheline. Heating of quartz–feldspar rocks results in the formation of liquid at the contacts between quartz and feldspar grains. This liquid has the composition of the granite eutectic regardless of the abundance of quartz and feldspar in the initial rock.

Incongruent melting is not as familiar to most of us as congruent melting; yet, many of the rock-forming minerals exhibit this phenomenon. The process is easily understood if examined in terms of free energy. This will be done for the incongruent melting of enstatite to form olivine and liquid. As was done with diopside, free energies of solid and liquid enstatite can be plotted as functions of temperature (Fig. 10.6). Where the two curves intersect, solid and liquid enstatite must be in equilibrium ($\Delta G = O$). This is the congruent melting point of enstatite (T_{cm} in Fig. 10.6). The free-energy plot of a compositionally equivalent mixture of olivine and liquid, however, has a steeper slope than that of solid enstatite, and over the temperature interval $T_{im} - T_L$ has lower values of free energy than either solid or liquid enstatite. Over this interval, solid or liquid enstatite can lower its free energy by changing into a mixture of forsterite and liquid. The congruent melting point of enstatite is therefore metastable. At low temperatures, enstatite has the lowest free energy and is thus the stable phase for this composition. At

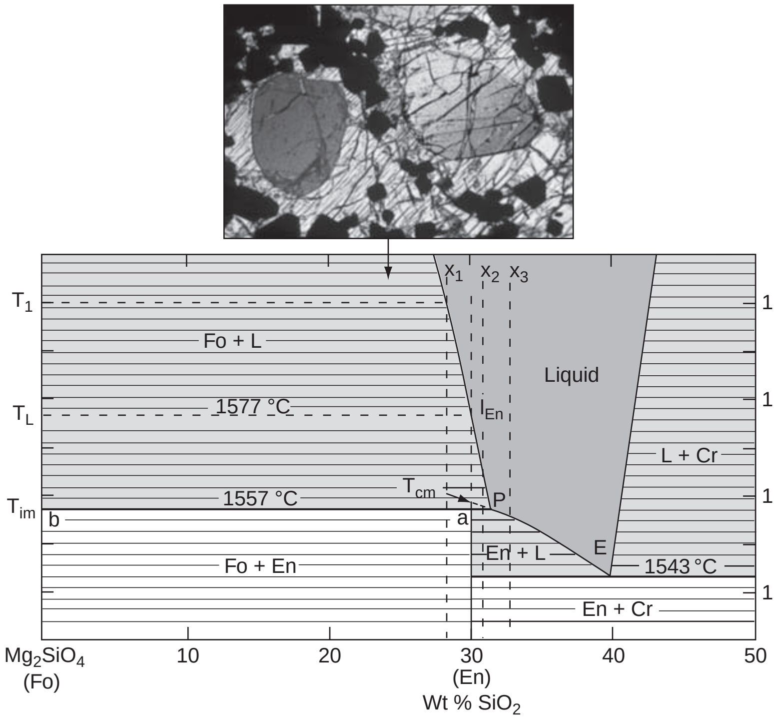


Fig. 10.7 Part of the system Mg_2SiO_4–SiO_2 at 1 atm pressure. (After Bowen and Andersen, 1914.) Photomicrograph of orthopyroxene reaction rims around olivine. Grain boundaries of olivine prior to reaction are indicated by presence of chromite grains (opaque), which accumulated with olivine from the magma; Stillwater Complex, Montana.

temperature T_{im}, a mixture of forsterite and liquid becomes the stable assemblage, and this, then, marks the incongruent melting point of enstatite (T_{im}). With rising temperature, the composition and proportion of liquid coexisting with forsterite continuously changes, so a family of G versus T curves could have been plotted, but only two have been shown for clarity. With rising temperature, these curves become steeper and the amount of forsterite decreases, becoming zero at temperature T_L. The liquid then has the composition of enstatite. Only above this temperature does the free energy of liquid enstatite become low enough for melt of this composition to be stable.

Let us now examine the binary phase diagram Mg_2SiO_4–SiO_2, which contains the binary incongruent-melting compound enstatite. Only the silica-poor side of the system is shown in Figure 10.7 in order to focus on the incongruent melting. A eutectic (E) between enstatite and cristobalite has a temperature of 1543 °C and a composition of 38% silica. From this point, the enstatite liquidus rises to what would be the congruent melting point of pure enstatite (T_{cm}), but before it reaches there, it is intersected (P) by the forsterite liquidus at 1557 °C. Above this temperature, which is the incongruent melting point of enstatite (T_{im}), the enstatite liquidus is metastable.

Heating of enstatite to 1557 °C results in the formation of liquid of composition P and crystals of forsterite. The presence of these three phases (En, Fo, Liquid) in a binary system at 1 atm pressure results in an isobaric invariant assemblage ($f = c + 2 - \phi = 1$). The composition of the liquid (P) and the temperature must remain constant until enstatite has completely melted, at which point the proportion of liquid to forsterite is given by the ratio of line lengths ab/aP. Continued heating causes the forsterite to dissolve in the liquid, which changes composition along the forsterite liquidus. With rising temperature, the proportion of forsterite crystals decreases and becomes zero at temperature T_L where the liquid has the composition of enstatite (l_{En}). The disappearance of forsterite results in a gain in degrees of freedom, and the liquid is then free to leave the liquidus and rise along the isopleth with further heating.

Liquid of composition x_1 in Figure 10.7 will, on cooling, intersect the forsterite liquidus at 1600 °C. Continued cooling under equilibrium conditions brings about crystallization of forsterite and enrichment of the melt in silica. When the liquid reaches point P (1557 °C), it begins to react with the crystals of forsterite to form enstatite. Because the system now has three phases, it is isobarically invariant, and the liquid remains at P until the enstatite-forming reaction has gone to completion. Reaction points in phase diagrams, such as P, are known as *peritectics*. The reaction at P consumes both liquid and forsterite, and depending on the bulk composition, one or the other of these phases can be consumed first, thus terminating the reaction. In the unique case of the composition being precisely that of enstatite, both liquid and forsterite are consumed simultaneously to form enstatite. The system is then totally solid. For composition x_1, however, the isopleth falls on the silica-poor side of enstatite, and cooling below the peritectic must involve only forsterite and enstatite. At the peritectic, therefore, all liquid must be consumed in the reaction, and only solids forsterite and enstatite remain. These minerals would show clear textural evidence of reaction, with enstatite rimming and eating into the olivine grains, as shown in the inset of Figure 10.7.

Liquid of composition x_2 would initially follow a similar sequence of crystallization steps to that of liquid x , but the

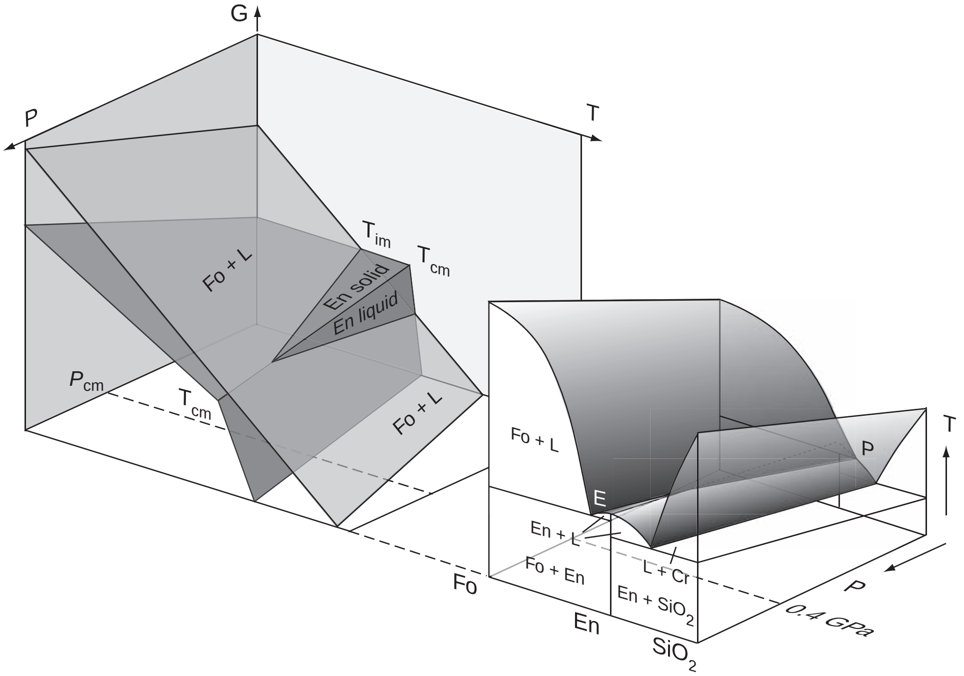


Fig. 10.8 Schematic plot of effect of pressure on free-energy surface of solid and liquid enstatite, and of chemically equivalent mixture of forsterite and liquid. Above a pressure of approximately 0.5 GPa, enstatite melts congruently, and the lower-pressure peritectic (P) changes to a eutectic (E). See text for discussion.

position of its isopleth indicates that at the peritectic, forsterite is the first phase to be totally consumed by the reaction. The system, then, would consist only of liquid and enstatite, and thus cooling could resume, with the melt descending the enstatite liquidus to the eutectic with cristobalite. Liquid of composition x_3, which lies on the silica-rich side of the peritectic, does not intersect the forsterite liquidus at all on cooling. Instead, enstatite forms initially as a primary phase rather than as a reaction rim on forsterite.

Unlike congruently melting compounds, incongruent ones do not change the number of eutectics present in a system. Under equilibrium conditions, crystallization must terminate at a peritectic or a eutectic, and from these, different assemblages form (forsterite + enstatite, and enstatite + cristobalite, respectively). With fractional crystallization, however, liquids, regardless of their initial compositions, can eventually reach the eutectic. The peritectic itself actually increases the opportunities for disequilibrium. Reaction rims, produced at the peritectic, commonly act as barriers to slowly diffusing reactants. It is not unusual, therefore, especially where cooling is rapid, to find rocks containing both quartz and magnesium-rich olivine, the olivine being mantled and separated from the quartz by orthorhombic pyroxene.

Liquids of peritectic composition, like those of eutectics, correspond closely to some common igneous rocks. For example, many continental flood basalts, when plotted in terms of the components of Figure 10.7, have compositions near the peritectic. Several explanations can be found for this. Melting of any mixture of enstatite and cristobalite must first produce a liquid of eutectic composition. Mixtures of enstatite and forsterite, on the other hand, do not melt until heated to the peritectic temperature, and then, regardless of the proportions of these minerals, the first-formed melt must be of peritectic composition. Thus, from a source region consisting of forsterite and enstatite, peritectic liquid is bound to be formed. It is interesting to note that this liquid, when crystallized, will produce some tridymite (or quartz), a phase not present in the source. Should liquids be formed with compositions above the peritectic, as might happen in a region of high heat flow, cooling to the peritectic will bring about the reaction that causes the system to become isobarically invariant. Because liquid of peritectic composition continues to exist for whatever time is necessary for the reaction to go to completion (assuming equilibrium), the chances of such liquid being tapped are greater than for liquid that is continually changing its composition along the liquidus.

Although enstatite melts incongruently at low pressures, above about 0.4 GPa it melts congruently (Hudon *et al.*, 2005). The reason for the change can be appreciated by examining the effect of pressure on the free energy surfaces that were shown in Figure 10.6 for solid and liquid enstatite, and a chemically equivalent mixture of forsterite and liquid. At atmospheric pressure, the congruent melting point of enstatite is metastable because the equivalent mixture of forsterite and liquid has a lower free energy. With increasing pressure the free-energy surfaces all increase (Eq. (7.32)), but not at the same rates (Fig. 10.8). The free energy of forsterite and liquid, in fact, rises so rapidly that above 0.5 GPa, it lies above the intersection of the solid and liquid enstatite surfaces. Above this pressure, therefore, enstatite melts

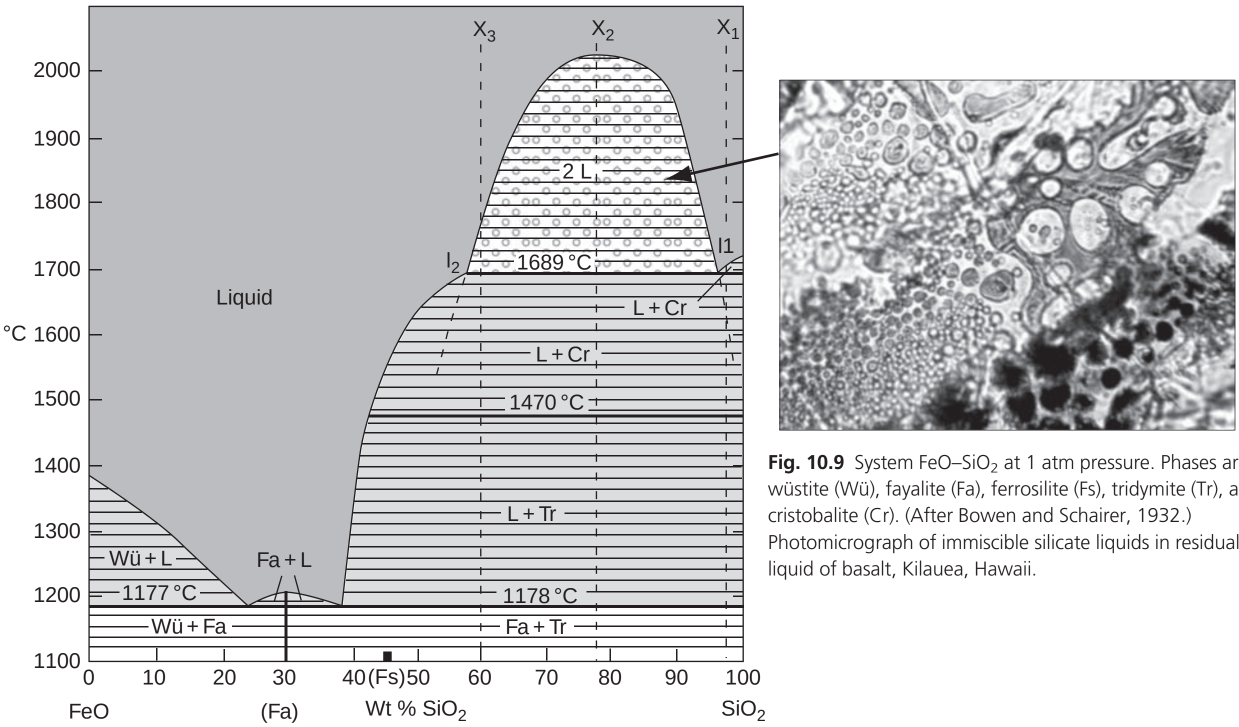


Fig. 10.9 System FeO–SiO_2 at 1 atm pressure. Phases are wüstite (Wü), fayalite (Fa), ferrosilite (Fs), tridymite (Tr), and cristobalite (Cr). (After Bowen and Schairer, 1932.) Photomicrograph of immiscible silicate liquids in residual liquid of basalt, Kilauea, Hawaii.

congruently because coexisting solid and liquid enstatite have lower free energy than the forsterite and liquid.

This change in the melting behavior of enstatite significantly affects the phase relations in the system Mg_2SiO_4–SiO_2. The univariant line marking the boundary between the liquidus surfaces of enstatite and forsterite (Fig. 10.8) changes from a peritectic at low pressure to a eutectic at high pressure. A single eutectic at low pressure, therefore, gives rise to two at high pressure, and perhaps more important, a thermal divide splits the system in two at high pressures. Fractional crystallization cannot move liquids from the silica-poor to the silica-rich side at high pressure. Also, melting of forsterite–enstatite mixtures no longer produces an oversaturated liquid as it does at 1 atm, and with increasing pressure the melts become progressively enriched in forsterite as the forsterite–enstatite eutectic moves toward forsterite. The olivine content of tholeiitic basalts may therefore be related to the depth at which the magma is generated (see Section 23.6 and Table 23.1).

10.7 BINARY SYSTEMS WITH LIQUID IMMISCIBILITY

There remains only one other phase relation involving liquids in binary systems that can be encountered without introducing solid solution, and that is liquid immiscibility. In developing the cryoscopic equation, the melt was assumed to behave ideally. Deviations from ideality, however, are common, and they can be large. In some systems, the lowering of free energy that should accompany mixing is less than expected, and in still others intermediate composition liquids can have free energies that are actually higher than liquids that are richer in one or the other component. The result is a saddle-shaped free-energy curve (Fig. 9.7(A)), which permits the coexistence of two compositionally different liquids, but both have identical chemical potentials of their components.

Many binary systems with silica exhibit liquid immiscibility. For example, systems involving SiO_2 with MgO, FeO, or CaO exhibit immiscibility, but those with alkalis or alumina do not. We shall look into the reasons for this difference following a discussion of the compositions of immiscible liquids. We shall use the system FeO–SiO_2 as an example (Fig. 10.9).

At a pressure of 10^5 Pa, the system FeO–SiO_2 has one binary compound, fayalite (Fe_2SiO_4), which melts congruently and produces two eutectics, one with wüstite (FeO) and the other with tridymite. The pyroxene ferrosilite ($FeSiO_3$) is stable only at high pressures, but its composition is included in Figure 10.9 for reference. Within the area marked 2L, single liquids unmix into two. This region of immiscibility extends to low temperatures, but only that part above the cristobalite liquidus is stable. However, silica liquids are readily supercooled (they are good glass formers) and it is possible to cool them down and intersect the metastable part of the two-liquid field. Glasses formed in this way consist of very small droplets of one glass in the other. On heating and annealing, a ceramic is formed that is of considerable commercial value. Corning, for example, uses such procedures to prepare PyroCeram®, which is used in such diverse products as the nose cones of rockets, and

Corningware® for cooking. The high-temperature stable immiscibility field shown in Figure 10.9 does reappear stably at much lower temperatures in more complex systems (see Fig. 10.24), and in the residual liquid of many basalts, such as the one from Kilauea shown in Figure 10.9.

Let us consider the equilibrium crystallization of three different liquids, each of which encounters the immiscibility field in a different way. Liquid of composition x_1 first encounters the liquidus of cristobalite, and crystallization of this phase enriches the melt in FeO, changing its composition to l_1. At this point, it encounters the two-liquid field, and any further enrichment in iron is impossible. The continued crystallization of cristobalite does generate FeO in the melt, but instead of this entering liquid l_1, a second liquid, l_2, forms. We can think of this as a reaction in which l_1 is converted to cristobalite plus l_2. The presence of these three phases in a two-component system produces an isobaric invariant assemblage; the temperature and compositions of liquids l_1 and l_2 cannot change until a phase has been lost. This does not mean, however, that the proportions of the two liquids cannot change; indeed, they must if cristobalite is crystallizing. When liquid l_2 first appears, it forms small iron-rich droplets (brown in thin section; see inset in Fig 10.9) in the silica-rich liquid (clear in thin section), but as crystallization continues the amount of iron-rich liquid increases. Eventually, all liquid l_1 is converted to cristobalite plus liquid l_2; then cooling resumes as the remaining single liquid descends the cristobalite liquidus to the eutectic with fayalite.

Liquid of composition x_2, on cooling, first encounters the immiscibility field and unmixes into approximately equal amounts of two liquids, one of which becomes progressively richer in silica, and the other richer in iron. Eventually the two liquids reach compositions l_1 and l_2, at which temperature cristobalite begins to crystallize. Again, this forms an isobarically invariant assemblage. Eventually, liquid l_1 is consumed and the iron-rich liquid, l_2, cools down the cristobalite liquidus to the eutectic.

Liquid of composition x_3 is relatively iron-rich to begin with, and so, on cooling, it first encounters the two-liquid field on the iron-rich side. This would bring about the nucleation of droplets of silica-rich liquid in the iron-rich host. Further cooling would bring the two liquids to compositions l_1 and l_2, and then the remainder of the crystallization is the same as in previous cases.

The heat removed from the isobaric invariant assemblage cristobalite + l_1 + l_2 is the latent heat of crystallization of cristobalite. The nutrients for the growth of this mineral (SiO_2) are concentrated in l_1, so it is likely that cristobalite will form in this liquid. But recall that equilibrium requires that the chemical potential of a component be the same in all phases. The chemical potential of silica in the iron-rich liquid, l_2, must therefore be the same as in liquid l_1 and the same as in cristobalite. Therefore, cristobalite can form from both liquids, but most will crystallize from the silica-rich one. We will have occasion to return to this point in discussing the evidence for liquid immiscibility in natural rock.The immiscibility field in the system FeO–SiO_2 extends over approximately the same compositional range as does the immiscibility field in the other systems exhibiting immiscibility mentioned previously. One side of the miscibility gap has a composition of almost pure SiO_2, whereas the other side has a composition corresponding closely to that of pyroxene. Herein lies the explanation for the immiscibility.

Silica-rich melts have structures similar to those of framework silicate minerals except that long-range order is lacking (Hess, 1980). They consist of networks of linked SiO_4 tetrahedra in which most oxygen form bridging Si–O–Si bonds. Such melts are said to be highly polymerized. Oxides of divalent cations with high ionization potentials also combine with oxygen in melts to form bonds of the type M–O–M, where M represents the cation. These cations, when mixed with silica-rich melt, are unable to bond with the bridging oxygens linking SiO_4 tetrahedra unless they break up the silica network. The following reaction then describes the depolymerization of the silica network brought about by addition of these cations to silica-rich melt:

$$\begin{array}{llll}\text{liquid silica} + & \text{liquid metal oxide} & = & \text{silicate melt}\\ \quad \mathrm{Si-O-Si} + & \quad \mathrm{M-O-M} & = & 2(\mathrm{Si-O-M})\end{array}$$

In this reaction, the bridging oxygen is changed to a nonbridging oxygen. Examples of nonbridging oxygens can be found in such silicate structures as the pyroxenes, where they bond with the cations between silica chains. In the melt, therefore, the presence of cations with high ionization potentials requires there be silicate ionic groups that can provide nonbridging oxygens. Of course, pyroxene-like chains are only one of the possible groups that can do this; rings and isolated tetrahedra can also provide these oxygens. When immiscible liquids form, a melt splits into a silica-rich fraction composed of a network of linked SiO_4 tetrahedra, and a relatively silica-poor fraction of less polymerized melt in which nonbridging oxygens bond with cations of high ionic potential. The latter melt tends to have a pyroxene composition, as seen by comparing the position of the silica-poor side of the immiscibility field in Figure 10.9 with the composition of ferrosilite.

Systems that do not exhibit immiscibility with silica involve cations with low ionization potentials that are able to accommodate themselves in the continuous linked network of silica-rich melts. Alumina, when present, is able to enter this network, substituting for silicon as long as cations such as K^+, Na^+, or Ca^{2+} are available for charge balance.

The immiscibility in Figure 10.9 occurs at temperatures that are too high to occur in the Earth at low pressure. We will encounter, however, an extension of this binary miscibility gap at much lower temperatures in three-component systems (Fig. 10.24). It is worth considering, therefore, what petrological principles can be read from this simple diagram, leaving aside for the moment the question of temperature.

Two liquids coexisting in the Earth for any length of time could possibly separate from each other if there is sufficient density contrast and viscosities are low enough. For example, the silica-rich liquid, l_1, could separate from the iron-rich one, l_2, by floating. Immiscibility therefore offers a possible means of fractionating magmas. All three liquids (x)

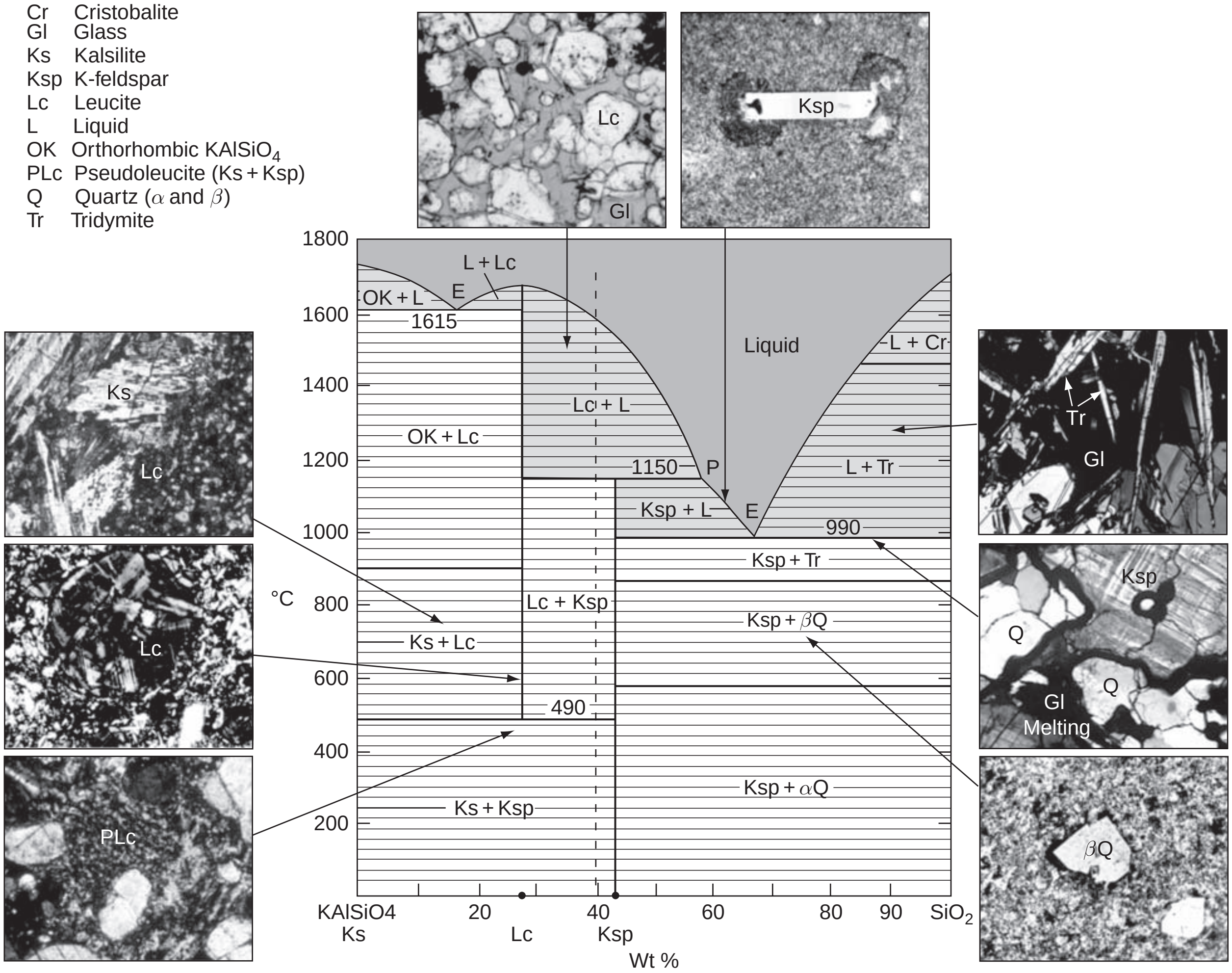


Fig. 10.10 System $KAlSiO_4$–SiO_2. Phases are kalsilite (Ks), orthorhombic $KAlSiO_4$ (OK), leucite (Lc), potassium feldspar (Ksp), low- (αQ) and high- (βQ) temperature quartz, tridymite (Tr), cristobalite (Cr), and liquid (L). P, peritectic; E, eutectic. (After Schairer and Bowen, 1955; Scarfe *et al*., 1966.) In a clockwise direction from lower left, photomicrographs show: pseudoleucite (PLc) an intergrowth of kalsilite and K-feldspar formed by inversion from leucite; leucite showing sector twinning formed from the isometric to tetragonal inversion; coexisting kalsilite and leucite; leucite phenocrysts in glass; sanidine phenocryst in rhyolite; fan-shaped twinned tridymite crystals in glass; melt generated at the contact between quartz and microcline grains in an experimental run; and dipyramidal high-temperature (β) quartz phenocryst in rhyolite.

whose crystallization histories were considered pass through the immiscibility field and, thus, could have been fractionated by this process. Yet each of these liquids eventually crystallizes the same minerals, and nothing in the assemblage itself indicates that immiscibility had been involved earlier. There, of course, might be certain features pertaining to the distribution of cristobalite that could be interpreted as resulting from immiscibility. But short of rapid quenching to form immiscible glassy globules, as happens in many volcanic rocks (Fig. 10.9), little evidence is likely to be preserved of this potentially important means of fractionating magma.

10.8 COMPLEX BINARY SYSTEMS WITH NO SOLID SOLUTION

Most previous sections in this chapter have dealt with binary systems that have each illustrated only one type of phase relation. Many binary systems combine some or all of these relations in a single diagram; consequently, they appear complex. But the combining of these phase relations does not change their behavior, so complex diagrams can be read with no additional problems.

The petrologically important system $KAlSiO_4$–SiO_2 (Fig. 10.10) exhibits a variety of binary phase relations. First, in addition to the polymorphs of the minerals forming the end components, kalsilite (Ks) and orthorhombic $KAlSiO_4$ (OK), and quartz (Q), tridymite (Tr), and cristobalite (Cr), the binary compounds leucite (Lc) and potassium feldspar (Ksp) are present. Leucite melts congruently at 1686 °C, but K-feldspar (sanidine) melts incongruently to leucite and liquid at 1150 °C, and thus only one additional eutectic is introduced into the system by these binary compounds. The eutectic between orthorhombic $KAlSiO_4$ and leucite at 1615 °C is at too high a temperature and too potassic a composition to be of much importance in nature. But the one between K-feldspar and tridymite at 990 °C is the “granite” eutectic in this potassium-bearing system. We have already met its sodic equivalent in the system albite–quartz (Fig. 10.5).

Let us consider the crystallization of the liquid marked by the isopleth in Figure 10.10 (dashed line). Under equilibrium conditions, leucite starts crystallizing from this liquid at 1600 °C, which enriches the melt in silica. Cooling continues until, at 1150 °C, the liquid reaches the peritectic, *P*, and reacts with the leucite to form K-feldspar. Because the isopleth immediately below the peritectic is in the field of leucite plus feldspar, all of the liquid must be consumed at the peritectic. Once this has happened, leucite and feldspar cool until, at 490 °C, leucite becomes unstable and breaks down to form an intergrowth of kalsilite and K-feldspar, which remains stable to low temperatures.

We have not previously encountered minerals that become unstable on cooling, but such behavior is common. Leucite, although unstable below 490 °C, does occur in some volcanic rocks, but this is due to metastability brought on by rapid cooling. With slow cooling, however, leucite does invert to the kalsilite–K-feldspar intergrowth. In doing so it commonly maintains the morphology of the original leucite crystal and is thus referred to as *pseudoleucite*. Prior to this inversion, leucite changes from an isometric structure to a tetragonal one, which produces the twinning commonly seen in leucite crystals (Fig. 10.10). Although pure potassic pseudoleucites have been found, most in igneous rocks contain considerable sodium in the form of nepheline. These likely involve some reaction or exchange of alkalis with the melt during their formation.

Because our chosen composition eventually crystallizes a feldspathoid rather than quartz, it would be classified as undersaturated (in silica); indeed, any composition on the silica-poor side of K-feldspar would be so classified. During part of its crystallization history, particularly at the peritectic, oversaturated liquids are generated. These liquids, if separated from the leucite crystals, would crystallize tridymite at the “granite” eutectic. Thus from an undersaturated melt an oversaturated one can be generated by fractional crystallization. This behavior is very different from that in the system $NaAlSiO_4$–SiO_2, where congruently melting sodic feldspar produces a thermal barrier, separating the under- and oversaturated rocks. The removal of leucite crystals from undersaturated magmas to produce oversaturated ones has been invoked to explain the association of these disparate rocks. Although the explanation is reasonable in terms of the phase diagram, we shall see later that leucite is restricted, by pressure, to near-surface environments. Moreover, field evidence indicating the separation of leucite crystals from magma has been found in only a few rare instances.

10.9 BINARY SYSTEMS WITH COMPLETE SOLID SOLUTION

So far we have considered systems in which there is no solid solution; that is, each solid phase has a specific composition, and only the liquid has variable composition. But many rock-forming minerals, such as the olivines, pyroxenes, amphiboles, and feldspars, belong to solid solution series. By changing the concentration of a component in a solid the chemical potential of that solid component is changed (Eq. (9.34)). Thus the cryoscopic equation derived in Section 10.2, where the chemical potential of the solid was taken to be fixed at the value of the pure mineral, is not applicable to systems with solid solution. Instead, the liquidus in these systems must be calculated in a manner that takes into account the variable chemical potentials in both the solid and liquid.

Consider equilibrium between a liquid and a solid, both of which form binary solutions. The olivines, ranging in composition from forsterite (Mg_2SiO_4) to fayalite (Fe_2SiO_4), provide a good example of such a system. At any pressure and temperature, olivine liquid will have specific chemical potentials of forsterite ($\mu^{\mathrm{L}}_{\mathrm{Fo}}$) and fayalite ($\mu^{\mathrm{L}}_{\mathrm{Fa}}$) depending on its composition. Similarly, crystalline olivine will have specific chemical potentials of forsterite ($\mu^{\mathrm{S}}_{\mathrm{Fo}}$) and fayalite ($\mu^{\mathrm{S}}_{\mathrm{Fa}}$) depending on its composition. If a crystal of olivine is in equilibrium with its melt, we can write

$$\mu^{\mathrm{L}}_{\mathrm{Fo}} = \mu^{\mathrm{S}}_{\mathrm{Fo}} \quad \text{and} \quad \mu^{\mathrm{L}}_{\mathrm{Fa}} = \mu^{\mathrm{S}}_{\mathrm{Fa}}$$

Using Eq. (9.34), we can write for the chemical potentials of forsterite in liquid and solid respectively

$$\mu^{\mathrm{L}}_{\mathrm{Fo}} = \mu^{*\mathrm{L}}_{\mathrm{Fo}} + RT\ln a^{\mathrm{L}}_{\mathrm{Fo}} \quad \text{and} \quad \mu^{\mathrm{S}}_{\mathrm{Fo}} = \mu^{*\mathrm{S}}_{\mathrm{Fo}} + RT\ln a^{\mathrm{S}}_{\mathrm{Fo}}$$

where $\mu^{*\mathrm{L}}_{\mathrm{Fo}}$ and $\mu^{*\mathrm{S}}_{\mathrm{Fo}}$ refer to the chemical potentials of pure forsterite liquid and solid respectively at the particular pressure and temperature considered. Equating these two equations, we obtain

$$\mu^{*\mathrm{L}}_{\mathrm{Fo}} + RT\ln a^{\mathrm{L}}_{\mathrm{Fo}} = \mu^{*\mathrm{S}}_{\mathrm{Fo}} + RT\ln a^{\mathrm{S}}_{\mathrm{Fo}}$$

which gives, on rearranging,

$$\mu^{*\mathrm{L}}_{\mathrm{Fo}} - \mu^{*\mathrm{S}}_{\mathrm{Fo}} = RT\ln\frac{a^{\mathrm{S}}_{\mathrm{Fo}}}{a^{\mathrm{L}}_{\mathrm{Fo}}} \tag{10.12}$$

But $\mu^{*\mathrm{L}}_{\mathrm{Fo}} - \mu^{*\mathrm{S}}_{\mathrm{Fo}}$ is the ΔG of fusion of pure forsterite at the specified pressure and temperature. If we assume that the ΔH and ΔS of fusion are constant over the temperature interval of interest, the same approximation for ΔG can be used as in Eq. (10.6), and then Eq. (10.12) can be expressed as

$$\ln\frac{a^{\mathrm{L}}_{\mathrm{Fo}}}{a^{\mathrm{S}}_{\mathrm{Fo}}} = \frac{\Delta H^{\mathrm{Fo}}_{\mathrm{m}}}{R}\left(\frac{1}{T^{\mathrm{Fo}}_{\mathrm{m}}} - \frac{1}{T}\right) \tag{10.13}$$

where $\Delta H^{\mathrm{Fo}}_{\mathrm{m}}$ is the heat of fusion of forsterite and $T^{\mathrm{Fo}}_{\mathrm{m}}$ is the melting point of pure forsterite. Equation (10.13) is a perfectly general equation that assumes nothing about the behavior of either solution. In the limiting case where the solid phase shows no solid solution, its activity would be 1, and Eq. (10.13) reduces to the previously derived cryoscopic equation (Eq. (10.6)). Equation (10.13), then, is the general form of the *cryoscopic equation*

To construct a phase diagram using this equation, it is necessary to know how the activities of the components are related to their mole fractions in both the solid and liquid solutions (Wood and Fraser, 1976). It will be recalled from Eq. (9.20) that a factor n, the number of equivalent sites on which mixing takes place in each formula unit of a solid solution, had to be introduced into the expression for the configurational entropy resulting from mixing. This same factor can be used to define activity where there is ideal mixing on n equivalent sites. The activity of an ideally mixed component in a solid solution in which the formula has n equivalent sites on which mixing takes place is given by

$$a_i = (X_i)^n \tag{10.14}$$

If the mineral has two distinct sites (M1 and M2, for example) on which mixing can take place the activity of an ideal solution is given by

$$a_i = X_i^{\mathrm{M1}} \times X_i^{\mathrm{M2}} \tag{10.15}$$

where X refers to the mole fraction of the cation (or anion) of the ith component occupying the particular M site. Olivine has two different octahedral sites, but they are so similar that magnesium and iron express little if any preference for one site over the other. As a result, the mole fraction of Mg^{2+} in one site is the same as in the other, and the activity of forsterite in crystalline olivine can be expressed as

$$a_{\mathrm{Fo}}^{\mathrm{S}} = X_{\mathrm{Mg}}^{\mathrm{M1}} \times X_{\mathrm{Mg}}^{\mathrm{M2}} = \left(X_{\mathrm{Fo}}^{\mathrm{S}}\right)^2 \tag{10.16}$$

Note that we would have obtained this same expression if we had assumed that olivine contained only one equivalent site (Eq. (10.15)).

An expression for the activity of forsterite in the melt is more difficult to formulate than that for the solid because less is known about the structure of melts. The simplest model of a silicate melt such as that of olivine is that it consists of cations and silicate anions that maintain, because of charge balance, the same general short-range order that they have in the crystalline state. Mixing is most unlikely to occur between cation and anion sites because of charge balance considerations. The entropy of mixing in such a melt would therefore be the sum of the entropies of mixing on the separate cation and anion sites. We can write the melting of olivine as

$$(\mathrm{Mg},\mathrm{Fe})_2\mathrm{SiO}_4 = 2(\mathrm{Mg},\mathrm{Fe})^{2+} + \mathrm{SiO}_4^{4+}$$

Because only one type of anion group is present in this system, there can be no contribution to the entropy of mixing from the anions. The entropy of mixing must therefore be due only to the mixing of the cations. The activity of forsterite in the melt can therefore be represented by

$$a_{\mathrm{Fo}}^{\mathrm{L}} = \left(X_{\mathrm{Fo}}^{\mathrm{L}}\right)^2 \tag{10.17}$$

We can now substitute the expressions for the activity of forsterite in the solid (Eq. (10.16)) and in the liquid (Eq. (10.17)) in Eq. (10.13) to obtain

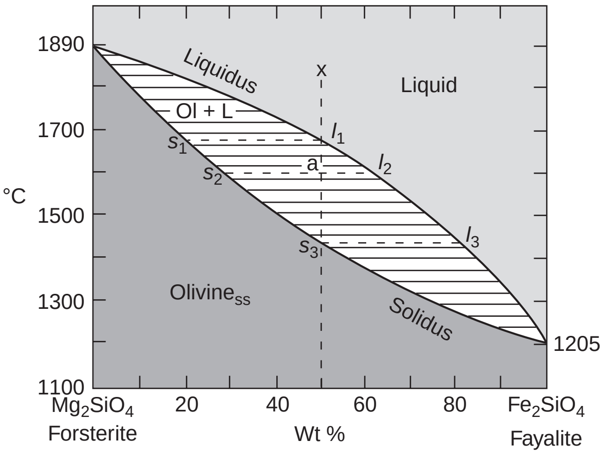


Fig. 10.11 System $\mathrm{Mg_2SiO_4}$–$\mathrm{Fe_2SiO_4}$ at 1 atm pressure. (After Bowen and Schairer, 1935.) See text for explanation.

$$2\ln\frac{X_{\mathrm{Fo}}^{\mathrm{L}}}{X_{\mathrm{Fo}}^{\mathrm{S}}} = \frac{\Delta H_{\mathrm{m}}^{\mathrm{Fo}}}{R}\left(\frac{1}{T_{\mathrm{m}}^{\mathrm{Fo}}} - \frac{1}{T}\right) \tag{10.18}$$

If a value for T in Eq. (10.18) is assumed, a ratio of mole fraction of forsterite in liquid to that in solid can be calculated, but we cannot determine the composition of either uniquely; there are one too many unknowns. A similar equation to Eq. (10.12) can, however, be written for the fayalite component from which it follows that

$$2\ln\frac{X_{\mathrm{Fa}}^{\mathrm{L}}}{X_{\mathrm{Fa}}^{\mathrm{S}}} = \frac{\Delta H_{\mathrm{m}}^{\mathrm{Fa}}}{R}\left(\frac{1}{T_{\mathrm{m}}^{\mathrm{Fa}}} - \frac{1}{T}\right) \tag{10.19}$$

where $\Delta H_{\mathrm{m}}^{\mathrm{Fa}}$ is the heat of fusion of fayalite and $T_{\mathrm{m}}^{\mathrm{Fa}}$ is the melting point of pure fayalite. Because there are only two components, $X_{\mathrm{Fa}}^{\mathrm{S}} = 1 - X_{\mathrm{Fo}}^{\mathrm{S}}$ and $X_{\mathrm{Fa}}^{\mathrm{L}} = 1 - X_{\mathrm{Fo}}^{\mathrm{L}}$, we can therefore rewrite Eq. (10.19) as

$$2\ln\frac{1 - X_{\mathrm{Fo}}^{\mathrm{L}}}{1 - X_{\mathrm{Fo}}^{\mathrm{S}}} = \frac{\Delta H_{\mathrm{m}}^{\mathrm{Fa}}}{R}\left(\frac{1}{T_{\mathrm{m}}^{\mathrm{Fa}}} - \frac{1}{T}\right) \tag{10.20}$$

Solving Eq. (10.18) and (10.20) simultaneously, we determine uniquely the compositions of liquid and solid coexisting at any temperature. This gives the liquidus and solidus of olivine (see Problem 10.5).

The experimentally determined phase diagram for the system $\mathrm{Mg_2SiO_4}$–$\mathrm{Fe_2SiO_4}$ at atmospheric pressure is given in Figure 10.11. Let us consider the equilibrium crystallization of a liquid of composition x in this system. This melt, on cooling, reaches the liquidus at 1675 °C (l_1) and crystallizes olivine with a composition of 80% forsterite (s_1). Crystallization of this magnesium-rich olivine enriches the residual melt in fayalite. As the melt descends the liquidus, equilibrium requires that crystalline olivine simultaneously change composition along the solidus. Thus the solid continuously reacts with the liquid to produce more fayalitic olivine. For example, when the liquid reaches l_2, the solid has a composition s_2, and the proportion of liquid to solid is given by a/l a. With continued cooling, the melt descends the liquidus until it reaches l_3, at which point the solid has composition s_3, which is identical with the starting bulk composition (see the isopleth). Consequently, the amount of liquid becomes zero at this temperature, and with only one phase remaining (s_3), the system gains a degree of freedom and crystalline olivine is free to leave the solidus and cool.

Because this crystallization process involves only two phases, liquid and solid (solution), it is never isobarically invariant. Thus, the reaction that takes place with cooling is univariant and occurs continuously with falling temperature. This is in contrast to the reaction at a peritectic in a binary system, where two solid phases plus liquid produce an isobaric invariant assemblage; in that case there is a discontinuity in the liquidus (the peritectic). These two reaction processes are commonly referred to as *continuous* and *discontinuous*, respectively.

The continuous reaction required to maintain equilibrium during crystallization of a solid solution phase increases chances for fractionation to occur. Mantling of early-formed refractory cores by later crystallizing material through which diffusion can occur only slowly causes residual melts to become increasingly enriched in the low-melting fraction. Zoned crystals are formed with high-temperature cores and low-temperature rims. This type of zoning, which is referred to as *normal*, is common in many minerals, especially in volcanic rocks, where rapid cooling provides insufficient time for equilibration. The ferromagnesian minerals typically become more iron-rich toward their rims, and the feldspars become richer in alkalis.

Whether minerals exhibit zoning or not depends on diffusion rates, which in turn depend, to a large extent, on the degree of polymerization in the mineral. For example, in a rock containing simultaneously formed olivine, pyroxene, and plagioclase, the olivine might be homogeneous, the pyroxene exhibit zoning only toward the rims, and plagioclase be strongly zoned throughout. The olivine structure, consisting of isolated silica tetrahedra, requires the least disruption to bring about homogenization of the iron and magnesium in the octahedral sites. In pyroxene, on the other hand, silica chains allow for easy diffusion only parallel to their lengths. Finally, the homogenization of plagioclase requires not only the diffusion of sodium and calcium but also of silicon and aluminum. Because this involves disruption of the framework structure, zoning in plagioclase is especially common, even with slow cooling rates.

Zoning is often more common on the rims of crystals than in the cores. Although this can be due to changes in cooling rate, Figure 10.11 provides another explanation. The solidus of any solid solution phase becomes flatter toward the low-temperature end, whereas the liquidus becomes steeper. Consequently, the composition of the solid has to change more during the later stages of crystallization than during the early stages, and this increases the likelihood of zoning.

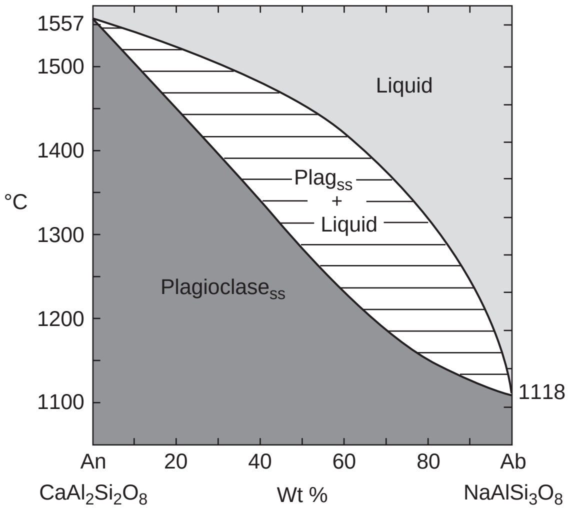


Fig. 10.12 System $CaAl_2Si_2O_8$ (anorthite) – $NaAlSi_3O_8$ (albite) at 1 atm pressure. (After Bowen, 1913; Weill *et al.*, 1980.) See text for explanation.

Melting in systems involving solid solution begins only when the temperature has risen to the solidus for the particular composition. This may be at considerably higher temperature than the minimum solidus temperature for the system as a whole. For example, 1118 °C is the lowest temperature at which a melt can exist in the plagioclase system at atmospheric pressure, and this only for a composition of pure albite (Fig. 10.12). Plagioclase of An_{50} composition, on the other hand, would have to be heated to 1280 °C before melting would begin. Once formed, though, this liquid could be fractionated down to the pure albite composition at 1118 °C. The melting and crystallization processes in systems exhibiting solid solution are, therefore, not necessarily the reverse of each other. This contrasts with the behavior in eutectic binary systems, where the initial melt formed on heating and the residual melt formed on cooling always have identical compositions.

10.10 POLYMORPHISM IN BINARY SOLID SOLUTIONS

Minerals belonging to solid solution series can undergo polymorphic transformations with changes in temperature or pressure, but these changes are compositionally dependent. A transformation is thermodynamically identical to that treated in Section 10.9 where one solution (solid) transformed (melted) into another solution (liquid). The fact that one solution is a solid and the other is a liquid in no way affects the general applicability of the results. Indeed, the cryoscopic equation describes equally well the compositional dependence of the temperature of a polymorphic transformation; we have simply to change the superscripts from liquid and solid to high- and low-temperature polymorphs, the ΔH_m to the heat of transformation, and T_m to the temperature of polymorphic transformation of the pure end member. The phase diagrams relating to polymorphism are therefore topologically identical to those described above for melting. For example, solid solution can either raise or lower the temperature of a polymorphic transformation. In an analogous

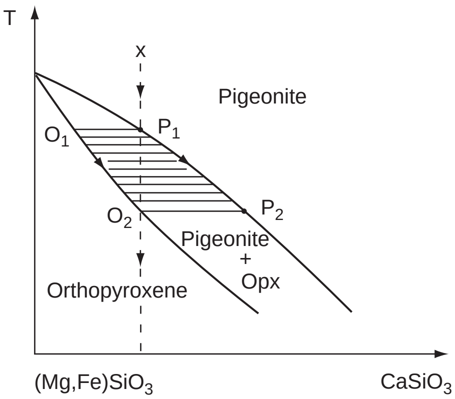


Fig. 10.13 Idealized phase relations in part of system $(Mg,Fe)SiO_3$–$CaSiO_3$ showing polymorphic transformation of pigeonite to orthopyroxene (Opx). See text for explanation.

manner the melting temperature of fayalite is raised by the addition of forsterite, and the melting temperature of forsterite is lowered by the addition of fayalite.

We have already encountered a polymorphic transformation whose temperature is raised by solid solution. Pure $NaAlSiO_4$, at low temperatures, forms the mineral nepheline, but above 1254 °C, at atmospheric pressure, it forms carnegieite (Fig. 10.5). Albite is soluble in both phases, but more so in nepheline. Consequently, the transformation temperature rises with increasing albite content. At 1280 °C, however, carnegieite and albite melt, preventing any further rise in the transformation temperature.

The change from orthopyroxene to pigeonite with increasing temperature exemplifies a transformation that is lowered by solid solution with $CaSiO_3$ (Fig. 10.13). Cooling or heating through such a polymorphic phase change involves similar types of steps to those encountered in crossing a liquidus and solidus. Cooling down the isopleth in Figure 10.13 from point x results in the appearance of orthopyroxene of composition O_1 from the pigeonite at P_1. Because the orthopyroxene has a different composition from pigeonite, its formation must involve diffusion of calcium through the solid. This is most easily achieved by nucleating numerous units of the low-temperature polymorph throughout the high-temperature phase. With continued cooling pigeonite changes its composition to P_2 and the orthopyroxene to O_2, whereupon, under equilibrium conditions, the transformation would be complete, because the low-temperature polymorph now has the starting composition. Theoretically, the polymorphic transformations in these solid solution phases should extend across the entire binary system, but in this particular example, other phase relations intervene (see Section 10.20). In other systems, however, polymorphic transformations can extend across the entire system.

Although polymorphic transformations in end members of solid solution series occur at definite temperatures, intermediate members transform over a temperature range and involve compositional readjustments. Thus, the transformation of olivine to a spinel structure at a depth of approximately 400 km in the mantle should not be expected to occur at one depth but should be spread out over a range of depths (temperatures).

10.11 BINARY SYSTEMS EXHIBITING PARTIAL SOLID SOLUTION

In Section 9.7, we dealt with the thermodynamics of nonideal solutions and why certain minerals belonging to solid solution series unmix into distinct compositional phases on cooling below the solvus. The alkali feldspars, clino- and orthopyroxenes, calcium-rich and calcium-poor amphiboles, ilmenite and hematite, and chalcopyrite and pyrrhotite are but a few of the common mineral pairs that form solid solutions at high temperatures but unmix on cooling. The question to be addressed here is what effect the presence of a solvus has on the liquidus and solidus relations.

The presence of a solvus indicates that mixing in the solids is far from ideal. Even at temperatures well above a solvus, nonideal mixing can make its presence known through deviations in the shape of the liquidus and solidus from those expected in ideal systems. If the deviations from ideality are great enough, the solvus can extend to temperatures that are high enough to intersect the solidus, and thus significantly change the phase relations.

We shall first consider a phase diagram in which a solvus intersects the solidus. At first, this diagram appears very different from that for an ideal solid solution, such as that of the olivine binary. It is, however, closely related, and it is helpful, in studying the more complex system, to keep track of the simple component parts. To help do this, a series of diagrams are presented in which a solvus moves closer to and eventually intersects the solidus (Fig. 10.14). Such stages might develop through increasing pressure. Typically, a solvus is raised to higher temperatures by increased pressure, but so is the solidus. Thus it is necessary, for purposes of illustration, to have the solvus rise more rapidly. Once the solvus intersects the solidus, its upper part becomes metastable, being replaced by the assemblage solids plus liquid. The presence of two solids with liquid at the point of intersection results in isobaric invariance. A peritectic, P, is formed on the liquidus.

The system MnO–FeO provides an example of such a system. The solid solution between these two is interrupted by a solvus, which intersects the solidus at 1430 °C. The solids are restricted to either manganese-rich or iron-rich compositions by the solvus. Manganese-rich ones are labeled MnO_{ss} in Figure 10.14(D), the subscript ss indicating that it is a solid solution. The iron-rich solid is labeled FeO_{ss}. Beneath the solvus only mixtures of MnO_{ss} and FeO_{ss} phases are stable, the compositions of which are given by points on the solvus at the temperature of interest.

Let us consider the equilibrium crystallization of a liquid of composition x. It reaches the liquidus of MnO_{ss} at l_1 with the formation of crystals of composition s_1. As the liquid cools, the solid continually reacts with it, becoming more iron-rich, until its composition reaches . Simultaneously, the liquid changes

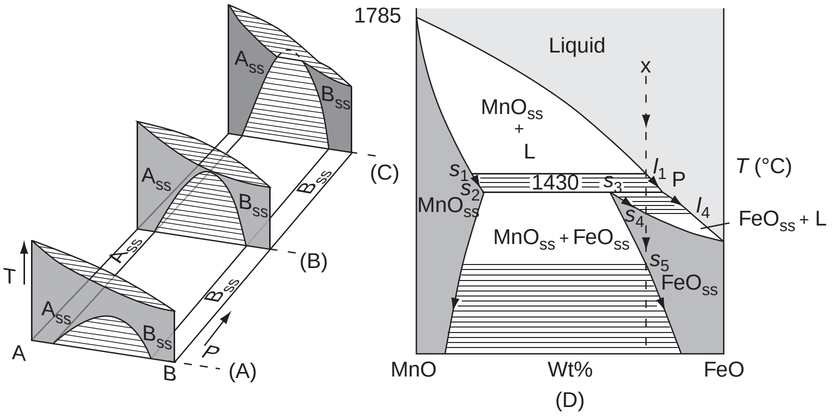


Fig. 10.14 (**A**-**C**) Solvus in the two-component system A–B rising with increasing pressure until it intersects the solidus. (**D**) Schematic 1-atm binary-phase diagram for the system MnO–FeO. Peritectic is indicated by letter P. See text for explanation.

composition to P, the peritectic, at which point the MnO_{ss} reacts with the liquid to form FeO_{ss} of composition s_3. The reaction continues under the isobaric invariant conditions until all MnO_{ss} is consumed. Had the isopleth passed to the manganese-rich side of s_3, all liquid would have been consumed at the peritectic instead. (Why?) Because only FeO_{ss} and liquid remain, continued cooling causes the solid to descend the FeO_{ss} solidus. When it reaches s_4, the isopleth composition, the last liquid of composition l_4 crystallizes. The gained degree of freedom allows the FeO_{ss} to leave the solidus and cool until it intersects the solvus at s_5. Continued cooling below this point causes exsolution of MnO_{ss} from the FeO_{ss}, with the compositions of both phases descending opposite sides of the solvus to low temperatures. Note that MnO_{ss} appears twice in the crystallization, first as a primary crystallizing phase, which is eliminated at the peritectic, and then later as an exsolved phase from the FeO_{ss}.

This system exhibits both continuous reactions (s_1–s_2 and s_3–s_4) and a discontinuous one (peritectic). Departures from equilibrium, therefore, can produce both zoned and mantled crystals, which on cooling intersect the solvus over a range of temperatures and compositions. This can produce complicated exsolution patterns. Such disequilibrium enriches the residual liquid in the low melting fraction (Problem 10.8).

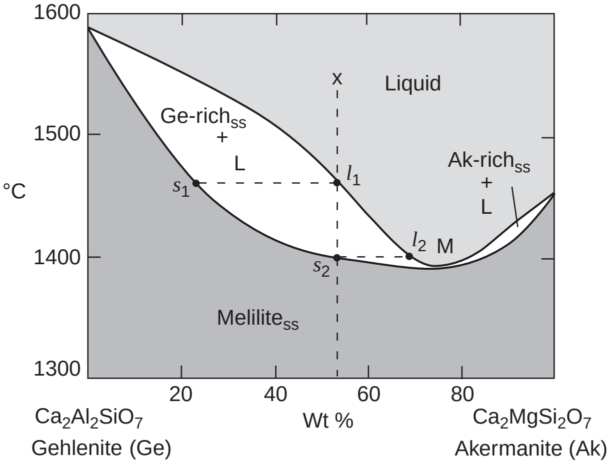


Fig. 10.15 System $Ca_2Al_2SiO_7$ (gehlenite)–$Ca_2MgSi_2O_7$ (akermanite). (After Osborn and Schairer, 1941). Minimum is marked by letter M. See text for explanation.

10.12 BINARY SYSTEMS WITH LIQUIDUS PASSING THROUGH A MINIMUM

All points on the liquidus in systems such as those considered in Sections 10.9 to 10.11 fall between the melting points of the pure end members. In systems that deviate markedly from ideality, however, the liquidus more commonly decreases to a *minimum* between the end members (M in Fig. 10.15). In such systems, the liquid is always closer in composition to the minimum than the coexisting solid solutions, regardless of which side of the minimum it is on, but at the minimum, the liquid and solid have identical compositions. A minimum resembles a eutectic in that the lowest possible temperature at which a liquid can exist occurs within the binary, but it differs from a eutectic in having only two phases present, liquid and one solid (solid solution), instead of liquid and two solids. Two phases are present along the entire length of the liquidus, and the minimum is just the lowest point on the continuously variable isobaric univariant liquidus; there is no discontinuity such as occurs at a eutectic. In still other rarer systems, the liquidus rises to a maximum within the binary system, but no petrologically relevant systems are known to exhibit this type of behavior.

The melilites provide an excellent example of a system exhibiting a liquidus minimum (Fig. 10.15). Gehlenite ($Ca_2Al_2SiO_7$) and akermanite ($Ca_2MgSi_2O_7$), which form a complete solid solution series at high temperature, melt at 1590 °C and 1454 °C respectively. The liquidus decreases from the end members to a minimum temperature of 1385 °C at a composition of 72% akermanite. Liquids, such as that of composition x, on cooling to the liquidus, crystallize gehlenite-rich melilite, whereas those on the other side of the minimum crystallize akermanite-rich melilite. Crystals of composition s_1 form first from liquid x, but under equilibrium conditions, they continuously react with the liquid, changing composition to s_2 while the liquid cools from l_1 to l_2. At this stage, all liquid crystallizes if equilibrium has been maintained, and the melilite leaves the solidus and continues cooling.

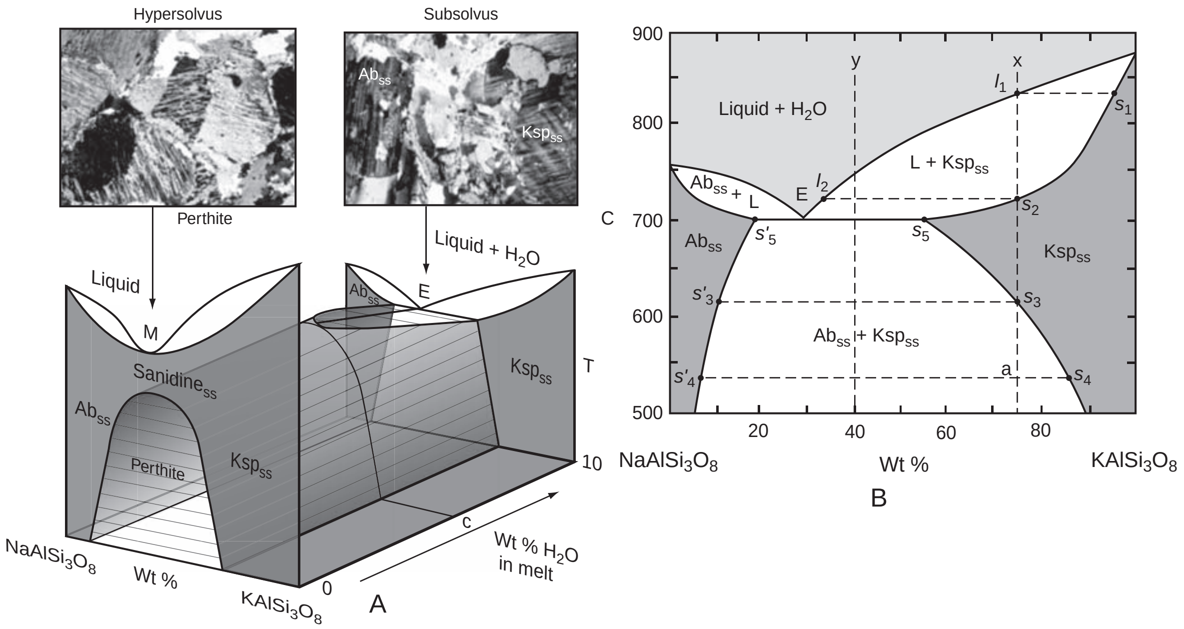


Fig. 10.16 (**A**) Schematic representation of effect of addition of H_2O to liquid in system $NaAlSi_3O_8$–$KAlSi_3O_8$ at 0.5 GPa. The minimum (M) in the anhydrous system changes to a eutectic (E) above a critical concentration (C) of H_2O in the melt. Photomicrographs illustrate typical hypersolvus and subsolvus textures. (**B**) H_2O-saturated phase diagram for alkali feldspars at 0.5 GPa. All phases coexist with an H_2O-rich vapor. (After Morse, 1970.) See text for explanation.

Fractional crystallization in systems such as this enriches residual melts in the minimum composition; thus, melilites containing 72% akermanite can be expected to be common. Melilites are rather rare igneous minerals, and occur only in very silica-poor rocks. They, however, do have compositions near the minimum, but they contain, in addition, moderate amounts of iron and a third melilite component, soda melilite ($CaNaAlSi_2O_7$), which cannot be shown in this binary diagram (see Fig. 10.37).

The alkali feldspars provide an example of a much more common solid solution series exhibiting a minimum in the liquidus. This system also has a solvus that, under certain conditions, can intersect the solidus, producing another type of binary liquidus phase relation.

At low pressures potassium feldspar melts incongruently to leucite and silica-rich liquid (Fig. 10.10), neither of which have compositions that can be expressed in terms of the two components $NaAlSi_3O_8$ and $KAlSi_3O_8$ (Fig. 10.16). Additional components are needed to describe all phases that can form in the potassium-rich part of the system, so this part is said to be nonbinary. At high pressures, however, potassium feldspar melts congruently, and the system is binary throughout.

At 0.5 GPa, the solidus is well above the solvus, so crystallization in this system is similar to that in the melilite system (front part of Fig. 10.16(A)). An alkali feldspar crystallizes which, initially, is richer than the liquid in either the albite or potassium feldspar component depending on which side of the minimum it falls. Eventually, under equilibrium conditions, a single feldspar crystallizes whose composition is that of the starting liquid. If the liquid is fractionated, compositions approaching the minimum, which is at approximately 70% Ab, will form. On reaching the solvus, the single feldspar unmixes into sodium- and potassium-rich fractions (*perthite*), which upon further cooling descend on opposite sides of the solvus.

If water is added to this system at 0.5 GPa, it dissolves in the liquid and lowers the chemical potential of alkali feldspar. Water is not soluble in the solid feldspar, nor does it cause hydrous minerals to form. Therefore, water does not affect the chemical potential of the solids, and the result, according to the cryoscopic equation, is a lowering of the alkali feldspar liquidus. The system is now really ternary, but because the added component, H_2O, does not form a crystalline phase, the representation as the simpler binary system is permissible, as long as we remember that the melt is not truly binary but contains water. The pressures in such systems are commonly recorded in the petrologic literature as P_{H_2O}, the subscript indicating the hydrous nature of the system.

As the water content of the alkali feldspar melt increases, so the liquidus temperature progressively decreases. At 0.5 GPa, however, the melt becomes saturated in water when it reaches 10 wt%, and no further lowering of the liquidus temperature can occur. Any additional water at this pressure forms a separate vapor phase. By increasing the pressure, however, more water can be dissolved in the melt and the liquidus temperature is lowered still further. Conversely, decreasing the pressure decreases the ability of the melt to contain water and raises the liquidus temperature. Silicate melts, like carbonated beverages, require pressure to keep most dissolved gases in solution. Thus, low liquidus temperatures due to dissolved water can occur only under high pressure.

Addition of water to alkali feldspar melt lowers liquidus temperatures, but it does not affect the alkali feldspar solvus, because water does not enter the crystalline feldspar. Consequently, there comes a stage, as water is added to the system (C in Fig. 10.16(A)), where the alkali feldspar minimum intersects the solvus. Then, because two solid phases coexist with the melt, a eutectic (or peritectic) relation must exist. The phase relations in the water-saturated alkali feldspar system at 0.5 GPa are depicted in Figure 10.16(B) – all phases coexist with a water-rich vapor.

Let us consider the equilibrium crystallization of a hydrous melt of composition x in Figure 10.16(B). Cooling of the melt to the liquidus at l_1 results in the appearance of crystals with composition s_1, which is rich in the K-feldspar component. Continued cooling causes the liquid to become more sodic as it descends the liquidus to l_2. Simultaneously, the solid continuously reacts with the liquid, changing its composition along the solidus to s_2. Because s_2 falls on the isopleth for composition x, all liquid must have crystallized by this stage.

It is important to bear in mind that the liquid in this system contains water. Thus as melt x cools and crystallizes anhydrous alkali feldspar, water must continuously be transferred from the melt to the vapor phase in order to maintain the concentration of water at the saturation limit (10 wt%) in the ever-decreasing amount of residual liquid. Crystallization of any anhydrous mineral from a water-saturated melt must cause nucleation and growth of bubbles of vapor. As will be seen in Section 11.6, this vapor phase plays a pivotal role in the generation of *pegmatites*, rocks with extremely coarse grain size and commonly economically important concentrations of rare elements.

Returning to the cooling of composition x, we see that following the disappearance of the liquid, the solid at s_2 is free to cool. On reaching the solvus at s_3, exsolution lamellae of albite solid solution (s'_3) form. Further cooling depletes the host crystal in sodium, changing its composition along the solvus, while the exsolution lamellae become enriched in sodium. At any particular temperature the compositions of the two alkali feldspars are given by points on either side of the solvus, such as s_4 and s'_4 at 530 °C. The proportion of exsolved phase to host crystal is, then, s_4a to s'_4a.

Although liquid x does not reach the eutectic under equilibrium conditions, a small amount of fractionation would make it do so. Also, if the composition were slightly closer to the eutectic, such as y in Figure 10.16(B), the residual liquid would reach the eutectic, regardless of whether there is fractionation. Indeed, many liquids in this system reach the eutectic, at which point two different feldspars crystallize, one a potassium feldspar solid solution of composition s_5 and the other an albite solid solution of composition s'_5. Both of these feldspars develop exsolution lamellae of the other on cooling.

Crystallization in the anhydrous alkali feldspar system gives rise to only a single primary feldspar, whereas in the water-saturated system many liquids crystallize two different alkali feldspars. In the first case, crystallization takes place above the solvus and is said to be *hypersolvus*. In the second, the solidus intersects the solvus, so crystallization is described as being *subsolvus*. We will encounter these terms again in discussing granites, but their significance is clear from the phase relations in Figure 10.16(A). Hypersolvus rocks, which are characterized by only one primary feldspar (later unmixing to perthite), form under relatively dry conditions and at relatively high temperatures. In contrast, subsolvus rocks, which are characterized by two different primary alkali feldspars, form under relatively wet conditions and at relatively low temperatures (see insets in Fig. 10.16(A)).

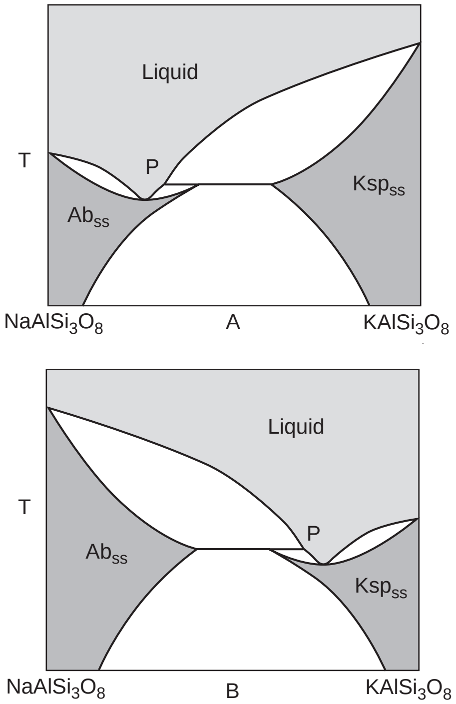


Fig. 10.17 Possible peritectic relations in the alkali feldspar system. See text for explanation.

In examining the intersection of the alkali feldspar solidus with the solvus in Figure 10.16(A), the astute reader may have wondered precisely how these two surfaces first make contact. For example, does the minimum coincide precisely with the top of the solvus, in which case a eutectic results; or does the minimum come down on the sodium- or potassium-rich side of the solvus, in which case a peritectic results (Fig. 10.17)? This is not a trivial question, for it implies significant differences in the phase relations.

Because the eutectic in Figure 10.16(B) lies to the sodium-rich side of the solvus, the minimum may also be on this side, in which case the peritectic relation would be as shown in Fig. 10.17(A). If the minimum is on the other side of the solvus, the peritectic relation would be like that shown in Figure 10.17(B). Reactions at the peritectic result in mantling of one feldspar by the other. Such textures are common in rocks, but are complicated by the solid solution of the third feldspar component, anorthite, which preferentially enters the

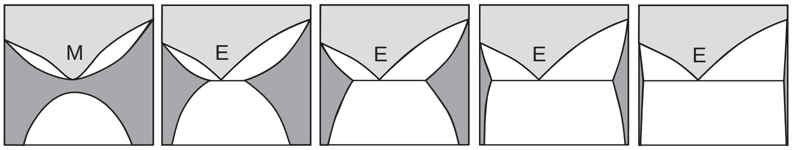


Fig. 10.18 A series of binary systems grading from complete high-temperature solid solution (left) to no solid solution (right).

albite solid solution and thus raises the liquidus surface of albite. The result is that the peritectic in Figure 10.17(B) represents the reaction most commonly encountered in rocks, that is, potassium feldspar rimming plagioclase. The position of the minimum is, however, affected by other factors in the melt, such as the activity of silica, and possibly under some conditions the peritectic of Figure 10.17(A) exists. This could then explain the *rapakivi* texture, where potassium feldspar is rimmed by plagioclase (Problem 10.10).

Regardless of which peritectic exists, the intersection of the solidus with the solvus increases with increasing water content. As a result, the peritectic moves closer to the solvus and on crossing it, becomes a eutectic, giving the phase relations in Figure 10.16(B). The ability of a peritectic to change to a eutectic (or the converse) with changing concentration of a third component is common in systems exhibiting partial solid solution. We will encounter it again in the pyroxene system where a eutectic between augite and pigeonite changes to a peritectic with increasing iron concentrations.

As the amount of solid solution in a binary system decreases, the intersection of the solidus with the solvus increases. The series of phase diagrams in Figure 10.18 illustrates this change. The first diagram is similar to the "dry" alkali feldspar system. This is followed by one similiar to the "wet" alkali feldspar system. In the successive diagrams, the amount of solid solution decreases until a diagram similar to the diopside–anorthite system, with essentially no solid solution, is formed. The system with which we began our discussion of binary phase diagrams is, therefore, an end member of this more general type of diagram.

This completes the discussion of possible phase relations involving the liquidus in binary systems. No new relations exist in systems containing more components, but their geometrical representation is more complex. Consequently, before proceeding to the discussion of ternary systems, readers should ensure they have a firm understanding of the binary-phase relations by doing Problem 10.1 to 10.10.

10.13 TERNARY SYSTEMS

Although binary systems illustrate all of the types of melting behavior found in multicomponent systems, they fall short of simulating accurately those in magmas. The addition of a third component greatly increases their verisimilitude and forms systems that closely portray the phase relations in common magmas. Graphical representations of systems with the additional component are, however, difficult because four dimensions are necessary to represent all possible variables. Nonetheless, by keeping some of these constant, useful and readily intelligible diagrams can be constructed. Graphical representations of more complex systems become increasingly difficult to read and are consequently less useful. For truly complex systems, graphical representation is impossible, and here is where a computer program such as MELTS is necessary to follow the crystallization of a magma.

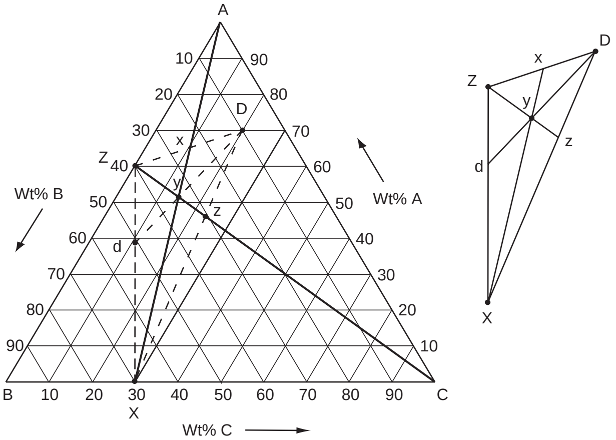


Fig. 10.19 Graphical representation of ternary compositions. See text for discussion.

Ternary systems can have four independent variables, temperature, pressure, and two of the chemical components; the third component is not independent, for it is specified as soon as the concentrations of the other two components are set. By considering isobaric conditions, as was done in most of the binary systems, the variables are reduced to three, and the system can be represented in three dimensions.

Compositions in ternary systems are normally plotted on triangular graphs, where the apexes represent the three components (Fig. 10.19). A composition plotted at an apex indicates 100% of that component, whereas a point on the opposite side of the triangle represents 0% of that component. Point X in Figure 10.19, for instance, contains no A and is a mixture of B and C in the proportions XC (70%) and XB (30%) respectively. If A is added to this mixture, the ratio of B to C will be unchanged, and compositions, such as y (50% A, 35% B, 15% C), will form along the line XA. Because y lies halfway along the line XA, it must, according to the lever rule, contain equal proportions of the compositions at both ends of the line, in this case A and the mixture of B and C. The fraction of A in composition y is therefore given by yX/XA. The fraction of C in y would similarly be yZ/ZC, and the amount of B could be determined with a similar line drawn through B and y. Just as the addition of A to composition

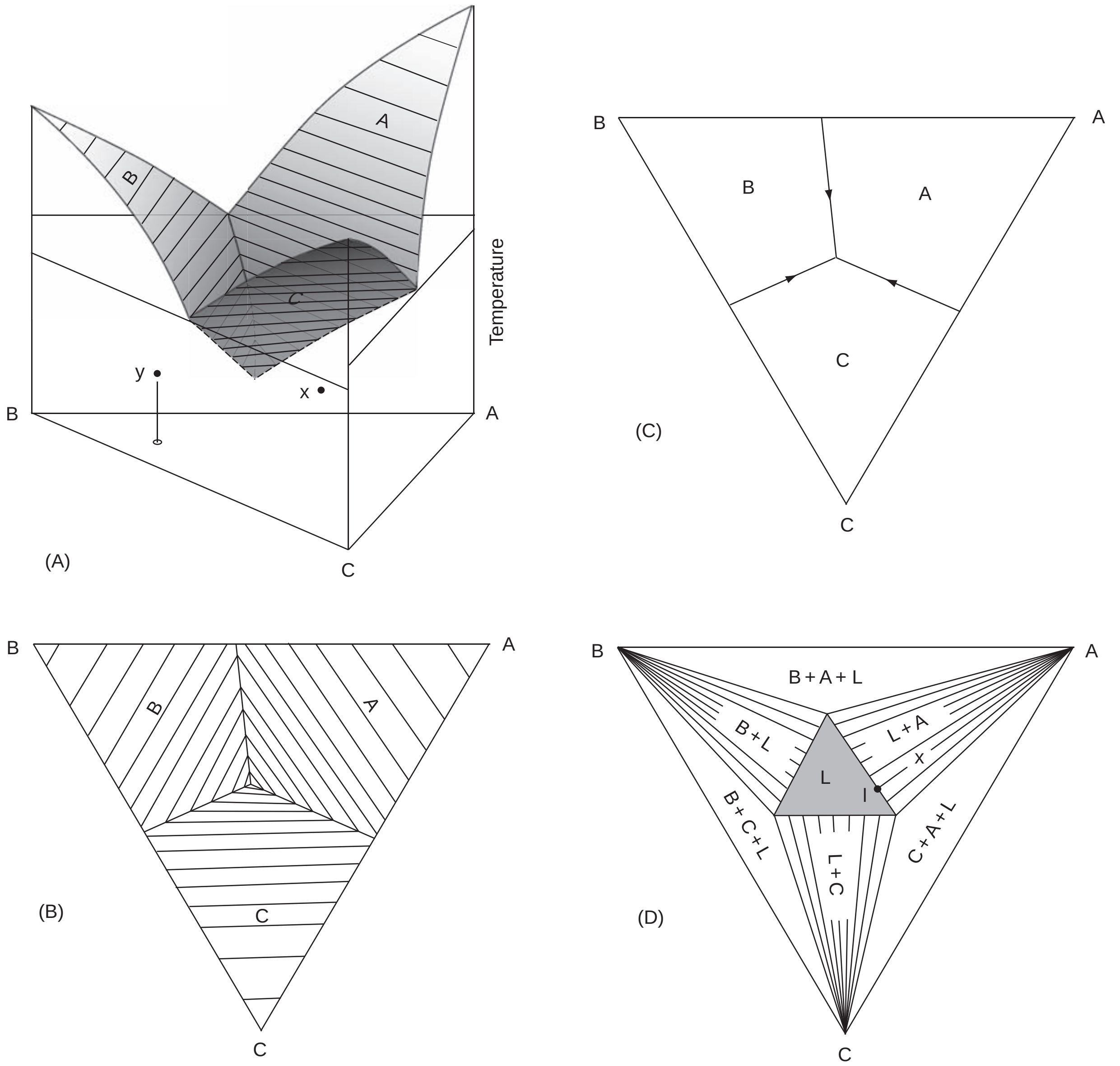


Fig. 10.20 Isobaric ternary system involving three pure phases, A, B, and C, which melt congruently. In (**A**), temperature is represented in perspective on the vertical axis. In (**B**), phase relations are projected onto triangular base, and temperature is indicated by isotherms. In (**C**), only directions of falling temperature on boundary curves are indicated. (**D**) Isothermal section through system.

X produced y, subtraction of A from y can produce X. Addition or subtraction of a component in these diagrams, thus, simply involves a linear change from the starting composition either toward or away from the component, respectively. The amount added or subtracted is determined with the lever rule (Problem 10.11).

Although compositions are usually plotted in equilateral triangles, this is by no means essential. Indeed, it is often necessary to describe a composition in terms of a different set of components that do not lie at the apexes of an equilateral triangle. For example, composition y might crystallize minerals having compositions X, Z, and D. The percentage of each mineral in y can be determined graphically in the same way as before, regardless of whether triangle XZD is equilateral. The fraction of X is xy/xX, of D is dy/dD, and of Z is zy/zZ.

The triangular compositional diagram can now have a third dimension added on which temperature can be represented. The resulting three-dimensional diagram portrays the phase relations at a particular pressure and is therefore described as isobaric. Figure 10.20(A) is a perspective view of the simplest type of ternary system, consisting of three congruently melting phases A, B, and C. Each face of this triangular prism is a simple binary-phase diagram, each with its own eutectic. From each binary eutectic, a line extends into the ternary system, marking compositions of ternary liquids that are in equilibrium with the binary eutectic minerals; this is known as a *cotectic*. The three cotectic lines meet at a ternary eutectic, which is the lowest possible temperature at which a melt can exist in the system at the specified pressure. The liquidus line in the binary diagram becomes a liquidus surface in the ternary system. Because the liquidus surface of each phase is bounded by an intersection with another liquidus surface, the line of intersection is known as a *boundary curve*. In this particular diagram, the boundary curves are cotectics, but other types of boundary curves will be encountered in other diagrams.

Although the perspective representation in Figure 10.20(A) is easily visualized, quantitative reading of such diagrams is difficult. For example, there is no way of knowing what composition is represented by point x; it could be a mixture of A and B, B and C, or any intermediate composition. Composition y, however, can be read because a vertical line shows not only the temperature of y (height) but also its composition projected onto the triangular base of the diagram. If every composition had to be indicated by a projection line, the diagram would soon become so congested as to be unusable. Instead, the three-dimensional diagram is reduced to two dimensions by projecting temperatures onto the triangular base of the prism (Fig. 10.20(B)). Compositions can now be read directly from the triangular diagram. Isotherms on the liquidus surface of each phase appear in the projection as lines that can be read in the same way as contours on a topographic map to indicate directions of falling or rising temperature. The boundary curves on the liquidus surface project as lines, and the ternary eutectic as a point. To avoid clutter, the isotherms are often omitted from the diagram, and directions of falling temperature are represented by arrows on the boundary curves (Fig. 10.20(C)).

The lowering of melting points in ternary systems is described by the same cryoscopic equation (Eq. (10.6)) used in binary systems. It will be recalled that this equation is independent of the substance causing the lowering of melting. Thus, in Figure 10.20, the lowering of the liquidus of A is dependent only on the mole fraction of A in the liquid (assuming ideal mixing in the liquid), regardless of whether dilution is effected by B, C, or a mixture of these. Isotherms will therefore be straight lines of constant mole fraction in ideal systems, and if the molecular weights of the two diluting components (B and C) are the same, the isotherms will be parallel to the BC side of the triangle.

In binary systems, the lowering of the liquidus of one of the phases is limited by the liquid becoming saturated in the other phase used to cause the dilution. This produces the binary eutectic. In ternary systems, however, the liquidus can be lowered still further by diluting the liquid with the third component. Eventually, of course, no further lowering of temperature is possible when the liquid becomes saturated in both the other phases, and this then produces the ternary eutectic. Clearly, addition of a fourth component would allow further lowering of the melting point until a quaternary eutectic was reached.

A useful way of presenting the phase relations in a ternary system is through a series of isothermal, isobaric sections. One such section is shown in Figure 10.20(D) for a temperature above the ternary eutectic but below any of the binary eutectics (the heavy isotherm in Fig. 10.10(B)). At this temperature, any composition lying within the central triangular area is liquid. The sides of this triangular area are formed by the intersection of the three liquidus surfaces with the plane of the isothermal section. A liquid having a composition on one of these lines (e.g. the one closest to A) would be saturated in crystals of A. Compositions lying within the triangular area bounded by the liquidus of A and A itself consist of crystals of A and a liquid. The liquid must have a composition on the liquidus. This composition can be determined by drawing a straight line from A through the bulk composition to the point of intersection with the liquidus. For example, a bulk composition of x in Figure 10.20(D) at the temperature and pressure of this diagram would consist of crystals of A and liquid of composition *l*, and by the lever rule the fraction of crystals would be x*l*/A*l*. Because the line A*l* ties together the compositions of coexisting phases, in this case crystals of A and liquid *l*, it is known as a *tie line*. Note that tie lines also radiate from B and C, tying the compositions of the respective liquids with their coexisting crystals.

In an isothermal section, cotectic lines become points where the lines pierce the isothermal plane. These points are the corners of the triangle bounding the field of liquid. A liquid having a composition at one of these points would be in equilibrium with two crystalline phases, for example A and B for a liquid on the AB cotectic. A bulk composition lying anywhere within the triangular area bounded by A, B, and the liquid on the AB cotectic consists of crystals of A, B, and cotectic liquid. Regardless of the bulk composition, the liquid has only one possible composition, that of the point of the cotectic. In terms of the phase rule, a ternary system consisting of three phases (crystals of A and B, and cotectic liquid, for example) is divariant, but the two variables are used in specifying the pressure and temperature of the isobaric, isothermal section. The resulting invariance means that the liquid cannot change its composition and remain in equilibrium with the two crystalline phases.

Isothermal, isobaric sections through ternary systems can be divided into fields consisting of one, two, or three phases: for example, liquid (L), liquid + crystal (L + A), and liquid + two crystalline phases (L + A + B). Within the liquid + crystal field, the liquid can vary its composition only along the liquidus surface, and consequently these fields are ruled with tie lines to indicate the permissible compositions of coexisting phases. Within the liquid + two crystalline phase field, the compositions of all phases are fixed and are given by the apexes of the triangular field, which is known as a *tie triangle*.

At successively lower-temperature isothermal sections, the field of liquid shrinks, whereas the triangular fields of liquid plus two crystalline phases expand. Eventually, the liquid field shrinks to a point at the ternary eutectic. At temperatures below this, isothermal sections would be of solids only and would simply be the triangle ABC.

Isothermal sections are a convenient means of presenting phase data without any possible confusion that might result from more complex projections. For example, it is simple to determine what would happen to a part of the Earth's mantle consisting of the solids A, B, and C in the proportions of

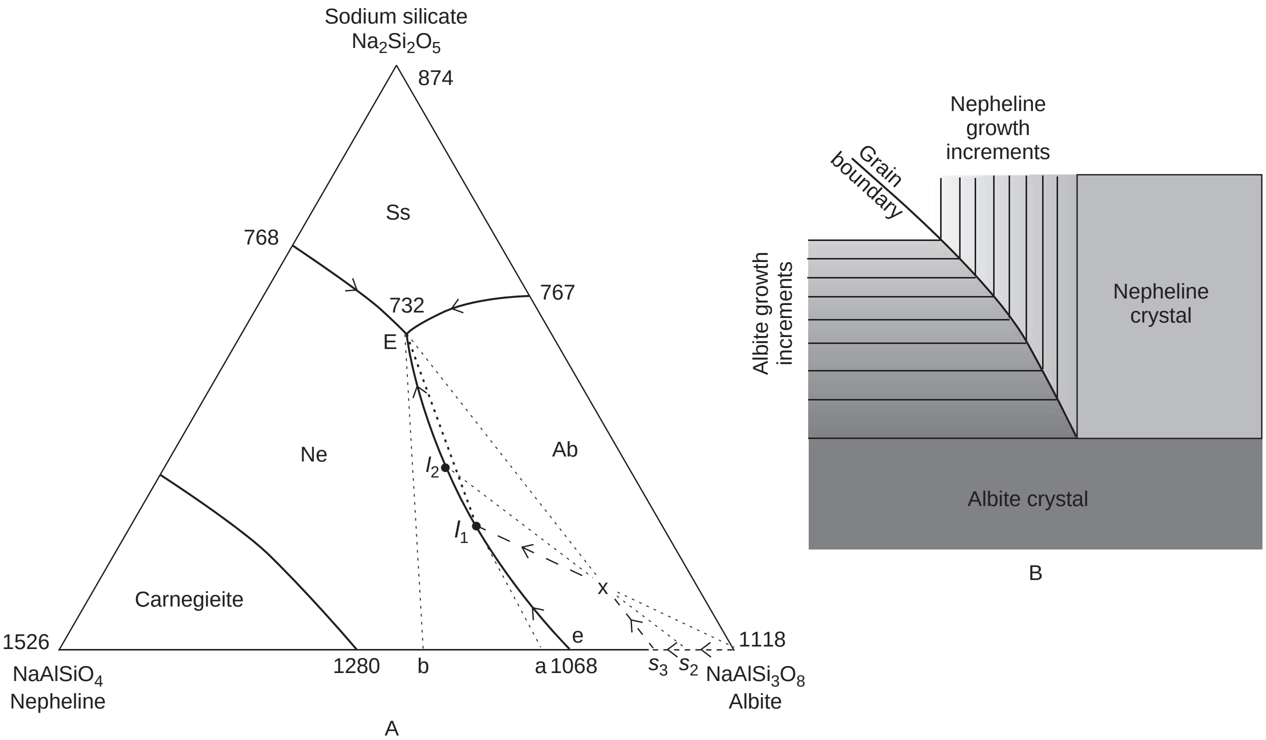


Fig. 10.21 (**A**) Ternary system albite–nepheline–sodium silicate (after Schairer and Bowen, 1956). Long-dashed line traces the liquid path from a starting composition of x to the cotectic, which it then follows to the ternary eutectic (E). Dotted lines are construction lines to determine the solids in equilibrium with the liquid. The short-dashed line traces the composition of the bulk solids from the Albite corner through s_2 and s_3 to x. Points between a and b mark the intersection of tangent lines drawn from the cotectic with the Nepheline–Albite join and mark the instantaneous composition of solids crystallizing at the point of tangency on the cotectic. The heavy dotted line between the eutectic and l_1 indicates possible compositions produce by mixing of these two liquids. Temperatures are in °C. (**B**) Curvature in grain boundary between albite and nepheline crystals resulting from changes in proportions of these phases forming along cotectic. See text for discussion.

composition x in Figure 10.20(D) if it were heated to the temperature of the isotherm in this diagram. All of the B and C and some of the A would melt to form a liquid of composition *l*, leaving only crystals of A. If we are also interested in the amount of liquid formed, that too can be read easily from the ratio of xA/A*l*.

10.14 SIMPLE TERNARY SYSTEMS WITH CONGRUENTLY MELTING PHASES WITH NO SOLID SOLUTION

The system albite–nepheline–sodium silicate is a ternary system with three congruently melting minerals. Nepheline exhibits some solid solution toward albite, but for the purpose of this discussion, all of the minerals can be considered to have the stoichiometric compositions indicated by the corners of the ternary diagram (Fig. 10.21(A)). Also, at temperatures above 1280 °C, nepheline forms the polymorph, carnegieite, but this does not affect the phase relations in the lower-temperature part of the system.

Consider the slow, equilibrium crystallization of a magma having the composition of x in Figure 10.21(A). This composition lies within the primary liquidus field of albite. Thus, on cooling, albite will be the first mineral to crystallize. When albite begins to crystallize, the magma must become depleted in albite. Subtraction of an albite component from the melt (to form the crystals) is represented graphically by a line that extends from x in a direction directly away from albite. With continued crystallization of albite, the composition of the melt moves away from x toward l_1. At l_1 the liquidus of nepheline is encountered; that is, the melt becomes saturated with nepheline, as well as albite.

It is important in tracing the sequence of crystallization in a ternary system to keep track of the composition and amount of the solids that form. During the first stage of crystallization of melt x, the composition of the solids formed must plot at the albite corner of the diagram, because this is the only mineral that crystallizes during this stage. The amount of solid formed is determined with the lever rule; for example, at the instant the melt reaches the cotectic (and before any nepheline has formed) the melt has the composition l_1, and the solids consist only of albite. The liquid l_1 and the solid (albite) must sum to the starting bulk composition, x. The ratio of solid to liquid is therefore given by the ratio of l_1x/xAb, which in this case is 50:50. Eruption and rapid quenching of the magma at this stage of crystallization would result in the formation of a vitrophyric rock containing 50 wt% of albite phenocrysts in a glassy groundmass having the composition *l*

If, instead of erupting, the magma continues to cool and crystallize slowly at depth, the residual melt will change its composition along the cotectic as albite and nepheline crystallize. Any mixture of solid albite and nepheline has a composition that plots on the line between these minerals. But only one composition on this line will effect the necessary change in liquid composition to keep the melt on the cotectic. For example, melt of composition l_1 must crystallize albite and nepheline in the proportion indicated by point a if it is to move along the cotectic toward l_2. Clearly, the direction of change along the cotectic at any point is given by the tangent to the cotectic at that point. The intersection of this tangent with the line joining the crystallizing minerals must represent the bulk composition of these minerals forming at that instant. Point a indicates that the amount of albite crystallizing from melt l_1 is given by aNe/NeAb and that of nepheline is given by aAb/NeAb.

The tangent to the cotectic allows us to determine the instantaneous rate of crystallization of the cotectic phases, but we must also keep track of the total amount of solid formed during the entire cooling period. Once the cotectic is reached, the early-formed albite has a cotectic mixture of albite and nepheline added to it. The bulk composition of these solids consequently moves away from the albite corner of the ternary diagram along the line towards nepheline. When the melt cools to composition l_2 on the cotectic, the total solid formed has a composition of s_2 on the line nepheline–albite. This is determined by drawing a straight line from l_2 through the bulk starting composition, x, to the line Ne–Ab. This graphical construction simply means that the phases on either end of this line (i.e. the liquid l_2 and the mixture of solids, s_2) must sum to the starting composition, x.

When the cotectic liquid reaches the ternary eutectic, the bulk composition of the solids is represented by s_3. At the eutectic, sodium silicate joins albite and nepheline as crystallizing phases and the system becomes isobarically invariant (four phases in a ternary system). The liquid, consequently, can no longer change its composition. Nonetheless, albite, nepheline, and sodium silicate must crystallize from this liquid, so these minerals must form in the same proportions as present in the ternary eutectic. To the solids at s_3 a eutectic mixture of albite, nepheline, and sodium silicate is added, moving the bulk composition of the solids from s_3 toward the starting composition x. When the solids reach x, the last eutectic liquid disappears and crystallization is complete (Problem 10.12).

In the discussion above the liquid has been taken to be in equilibrium with the solids at all stages during solidification. It is possible, however, that during crystallization the solids may separate from the liquid, for example by crystals sinking to the bottom of a magma chamber where they are no longer in contact with the liquid from which they crystallized. Such crystallization is said to occur *fractionally* and the liquid undergoes *fractionation*. Fractionation is an extremely important process, for it may account for much of the variation in the composition of igneous rocks.

Let us consider the fractional crystallization of magma with a composition of x in the ternary system of Figure 10.21(A). When this magma becomes 50% crystallized, the residual liquid has a composition of l_1 and the solids consist only of albite. If all of these crystals are removed, by sinking for example, only magma of composition l_1 would remain. When this new magma eventually solidifies, it would produce a rock having the composition of l_1, which contains considerably less albite than would have formed from the original magma, x. In this example, crystals were removed at only one stage; the removal, however, could be continuous, giving rise to complete fractional crystallization. In either case the result of fractional crystallization, whether partial or complete, is to produce rocks that are enriched in the low-melting eutectic fraction.

The sequence of appearance of phases indicated in the phase diagram would account for the obvious textures of a rock having a composition such as x. Albite, for example, would form phenocrysts, and sodium silicate would appear only in the groundmass. The phase diagram also provides information on the relative rates at which various minerals crystallize. When the melt has the composition l_1, albite and nepheline crystallize in the proportions indicated by point a; that is, approximately twice as much albite would be forming as nepheline at this stage. By the time the cotectic reaches the eutectic, however, the proportions of albite and nepheline crystallizing change to point b, which corresponds to approximately equal proportions of these minerals.

If magma crystallizes slowly, changes in the proportions of crystallizing phases along a cotectic will affect the way in which grains are intergrown. For example, when the liquid has the composition l_1, albite crystals would be expected to grow at twice the rate of nepheline crystals. If two such crystals were juxtaposed as shown in Figure 10.21(B), the thicker growth bands on the albite crystal would cause the boundary between these crystals to grow at a steep angle to the albite face. As the residual liquid approached the eutectic, the proportion of nepheline would increase until equal amounts of the two minerals were forming, at which point the angle of the grain boundary would be near 45°. Variation in the curvature of cotectics is therefore one of the factors that determines the shape of grain boundaries.

An important petrologic consequence of a curved cotectic is the possibility of producing monomineralic rocks from the mixing of different magmas on the cotectic. Many large igneous bodies are formed by repeated injections of magma, and many of the rocks are products of mixing of these successive pulses. Let us assume that an initial pulse of magma with composition x has cooled and fractionated to the eutectic composition when a second batch of magma is intruded with a less evolved composition, say magma of composition l_1. Mixing of eutectic liquid with that at l_1 produces liquids that have compositions lying in the primary field of albite. Thus magmas that were initially crystallizing several phases, precipitate only albite on mixing. In this way, layers of monomineralic rock (albite in this case) may be formed in large intrusive bodies.

10.15 TERNARY SYSTEMS WITH CONGRUENTLY MELTING BINARY PHASES

In Section 10.6 a congruently melting binary compound was shown to divide a binary system into two simple binaries, each with its own eutectic. Similarly, a congruently melting binary phase divides a ternary system into two simple ternary diagrams, each with its own ternary eutectic. For example, the system diopside–nepheline–silica contains the binary phase albite, and the line joining diopside and albite divides this system in two (Fig. 10.22). On the silica-rich side of this join, three cotectics meet at a ternary eutectic involving diopside, albite, and tridymite. On the silica-poor side of the join, three cotectics meet at another ternary eutectic, involving diopside, albite, and nepheline.

The directions of falling temperature on the boundary curves in Figure 10.22 are, for the most part, self-evident. Of particular interest, however, is the diopside–albite boundary curve, for it is common to both eutectics. To see how temperatures vary along this boundary, consider the crystallization of the two liquids labeled 1 and 2 in Figure 10.22. Both plot in the primary field of diopside; thus, they both start by crystallizing diopside, which is then followed by albite when the liquids reach the cotectic. From here on, the liquids follow the cotectic until they reach a eutectic. Because liquid 1 reaches the diopside–albite cotectic to the left of the diopside–albite join, crystallization of these two minerals must move the composition of the liquid away from the join toward the Di–Ab–Ne eutectic. The temperature on this part of the cotectic must therefore decrease toward this eutectic. Liquid 2, on the other hand, reaches the cotectic to the right of the diopside–albite join, and crystallization of these minerals causes the composition of the second liquid to move away from the join toward the Di–Ab–Tr eutectic. On this side of the join, the temperature of the cotectic must decrease toward the Di–Ab–Tr eutectic. If temperatures on the diopside–albite cotectic decrease toward both eutectics, there must be a temperature maximum between the eutectics, and this obviously occurs where the diopside–albite join intersects the cotectic.

The same conclusion can be reached by applying *Alkemade's theorem*, which states that the direction of falling temperature on any boundary curve is always in the direction away from the intersection of the boundary curve with the line joining the two phases that coexist on the boundary curve. This rule holds even when the boundary curve or the join has to be extended to obtain an intersection.

The temperature maximum on the cotectic in Figure 10.22 acts as a thermal barrier that prevents any crystallizing magma from crossing the dashed line joining diopside and albite. This join behaves in much the same way that a topographic divide prevents water from crossing from one drainage basin into another. Any magma with a composition to the left of the diopside–albite join must eventually crystallize at the Di–Ab–Ne eutectic, whereas any magma to the right of the join eventually crystallizes at the Di–Ab–Tr eutectic. The diopside albite join, therefore, divides rocks that plot in this ternary system into ones that contain nepheline and ones that contain tridymite (quartz). It will be recognized that this division is, in fact, the primary one used in the IUGS classification of igneous rock (Section 6.6). Rocks that are classified as alkaline or undersaturated (with respect to silica) plot to the left of the diopside–albite join in Figure 10.22, whereas those that are classified as subalkaline, oversaturated, or tholeiitic plot to the right of this join. This classification can now be seen to have a firm physical chemical justification.

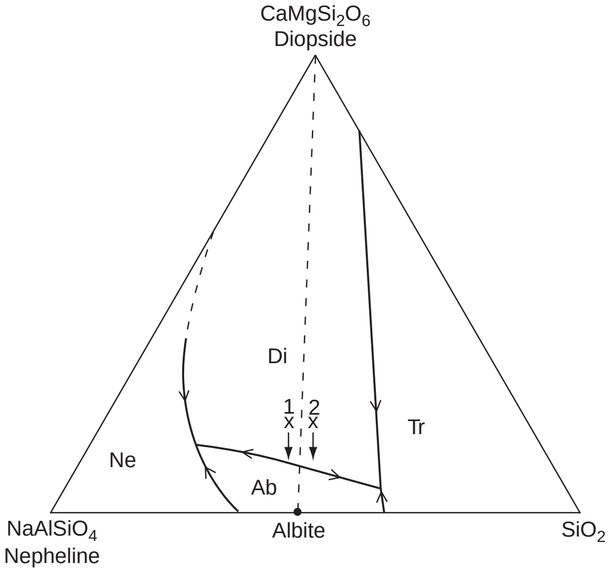


Fig. 10.22 System diopside–nepheline–silica. Small volume of nonternary olivine omitted from nepheline–diopside side of diagram where the boundary line is dashed. (Simplified from Schairer and Yoder, 1960.)

Ternary systems may contain more than one binary compound, but as long as each compound melts congruently, no complications are introduced. Each group of three minerals can be treated as a simple ternary diagram with a single eutectic. Congruently melting ternary compounds can also be treated in exactly the same way (Section 10.21).

10.16 TERNARY SYSTEMS WITH AN INCONGRUENT-MELTING BINARY PHASE

In Section 10.6 it was shown that if a binary compound melts incongruently in a binary system, an additional isobaric invariant point is introduced, the peritectic or reaction point, but there is still only one eutectic. Similarly, in a ternary system, the presence of an incongruently melting binary compound introduces a ternary peritectic, but there is still only one ternary eutectic.

In the ternary system forsterite–anorthite–silica (Fig. 10.23), the binary compound enstatite melts incongruently (review Fig. 10.7). A boundary curve separating the liquidus fields of forsterite and enstatite extends into the ternary system from the binary peritectic. Along this line, olivine continuously reacts with liquid to produce enstatite. The direction of falling temperature on such boundary curves is

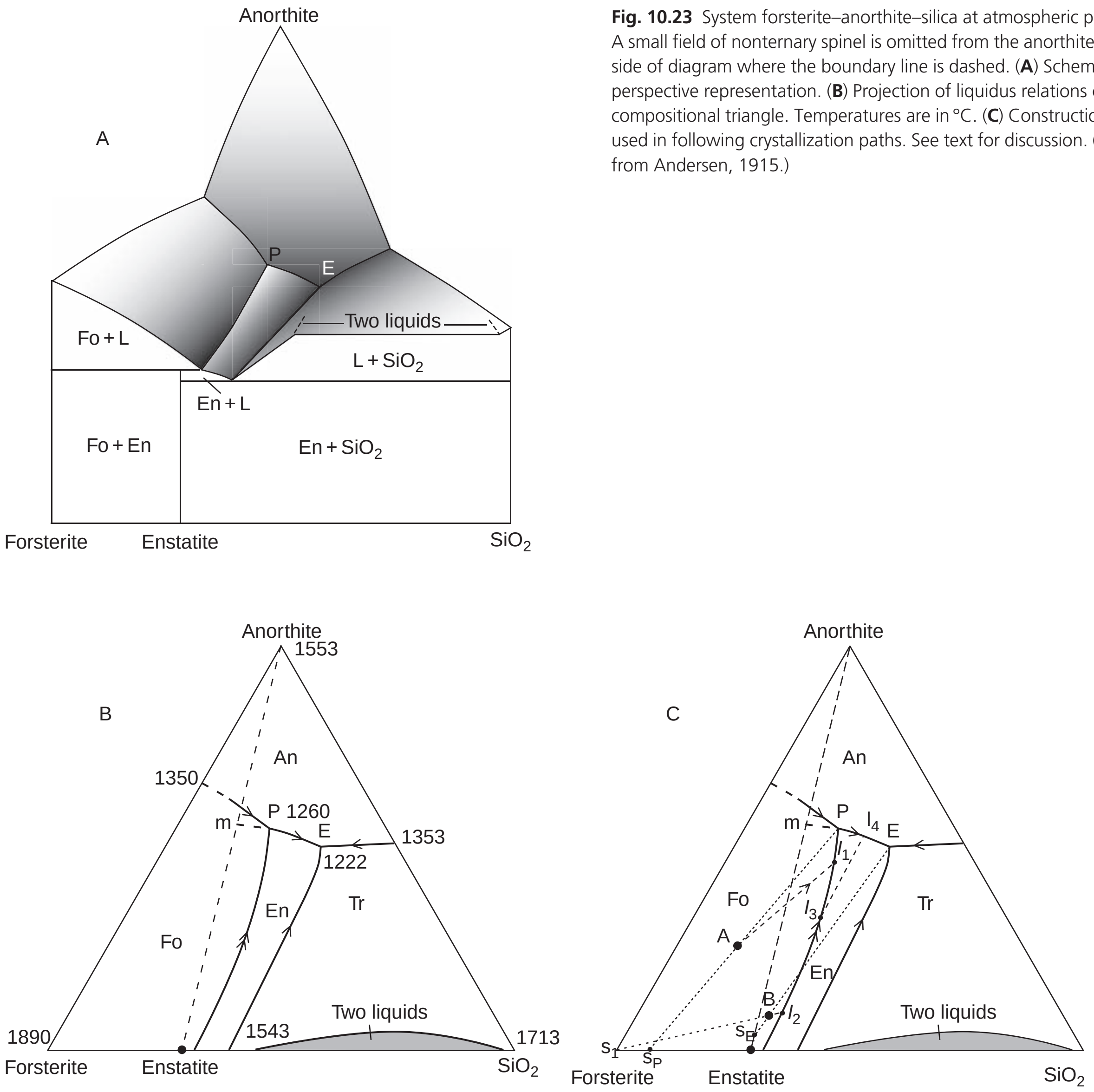


Fig. 10.23 System forsterite–anorthite–silica at atmospheric pressure. A small field of nonternary spinel is omitted from the anorthite–forsterite side of diagram where the boundary line is dashed. (**A**) Schematic perspective representation. (**B**) Projection of liquidus relations onto compositional triangle. Temperatures are in °C. (**C**) Construction lines used in following crystallization paths. See text for discussion. (Simplified from Andersen, 1915.)

commonly marked with a double arrow to indicate a reaction relation and to contrast it with a cotectic boundary curve, which is marked with a single arrow. The olivine–enstatite boundary is not a valley on the liquidus surface (Fig. 10.23 (A)) but is a line of inflection. Above the inflection is the steep liquidus of forsterite, and below it is the gentler-sloping liquidus of enstatite, which continues downward until it intersects the silica liquidus; this produces the cotectic valley that leads down to the ternary eutectic (E). Where the forsterite–enstatite boundary line intersects the liquidus surface of anorthite, coexistence of the four phases liquid, forsterite, enstatite, and anorthite produces an isobaric invariant assemblage – the ternary peritectic (P). At this peritectic, all olivine must be consumed by reaction before the liquid is free to continue cooling down the enstatite–anorthite cotectic to the ternary eutectic.

The directions of falling temperature on the boundary curves in Figure 10.23(B) can be determined using Alkemade's theorem. Consider first the forsterite–enstatite boundary curve; this line does not intersect the forsterite–enstatite join, so the join must be extended toward silica in order to find the point of intersection and temperature maximum on the boundary curve. This point is the binary peritectic. The temperature on the boundary curve must therefore decrease into the ternary system. The fact that this boundary curve does not intersect the forsterite–enstatite join between the minerals indicates that the boundary curve is a reaction line. Had the point of intersection fallen between the two, it would have indicated that a positive amount of each mineral crystallizes along the boundary, and thus the boundary would be a cotectic. The proportion of olivine and enstatite crystallizing from this liquid would depend on the position of the point of intersection relative to each mineral. The closer it was to enstatite, the smaller would be the proportion of olivine, the amount becoming zero when the intersection occurs at enstatite. When the point of intersection is to the right of enstatite, as it is in Figure 10.23, a negative amount of forsterite is indicated; that is, olivine is being consumed by reaction.

The direction of falling temperature on the enstatite–anorthite boundary curve is determined by tracing the metastable extension of this curve to the point of intersection with the enstatite–anorthite join (m in Fig. 10.23(B)); this point is the temperature maximum on this boundary. The temperature must therefore decrease from m toward the ternary eutectic.

In considering crystallization in ternary systems with a peritectic, it is essential to keep track of the composition of both the liquid and the solids. Consider the crystallization of a liquid with composition A in Figure 10.23(C). First, A plots in the subtriangle An–Fo–En; under equilibrium conditions, therefore, it must eventually crystallize to these three minerals. But only at the ternary peritectic, P, do these minerals coexist with a melt. Liquid A must therefore eventually crystallize at this point. Initially, however, olivine crystallizes from liquid A. This depletes the melt in forsterite and changes the composition of the liquid in a straight line away from forsterite until it reaches the reaction curve at l_1. Because forsterite is the only mineral to crystallize during this initial stage, the composition of the solids must be represented by s_1. As the liquid follows the reaction curve from l_1 to the ternary peritectic, *P*, forsterite reacts with the liquid to produce enstatite. The bulk composition of the solids, therefore, moves from s_1 towards enstatite, reaching point s_P when the liquid reaches the peritectic. The composition of the solids at any stage of crystallization is determined by drawing a line from the liquid through the initial bulk composition (A) to the line joining the crystallizing phases. The additional phase and decreased variance at the peritectic means that the liquid can no longer change its composition. It must therefore crystallize forsterite (negative amount), enstatite, and anorthite in the proportions given by point P (any other point would change the liquid's composition). Addition of a mixture of solids of composition P to the already formed solids moves the bulk composition of the solids from s_P toward P. But in doing so, the solids approach and eventually reach composition A, at which point the peritectic liquid is exhausted and a mixture of crystals of forsterite, enstatite, and anorthite is free to cool.

The crystals of forsterite that first form in liquid A are euhedral (Fig. 10.23(C)), but when the liquid reaches the reaction curve they become rounded and rimmed by enstatite. Such reaction textures are common in tholeiitic rocks. For instance, early crystallizing olivine in the Stillwater Complex of Montana sank and accumulated along with chromite near the base of the intrusion. The liquid trapped between these crystals had the peritectic composition and therefore reacted with the olivine to produce orthopyroxene. The outline of the original olivine grains is preserved by the position of the chromite grains (see Fig. 10.7).

Let us now consider the crystallization of a liquid with composition B in Figure 10.23(C). This liquid also first crystallizes olivine, but because the liquid plots in the subtriangle An–En–SiO_2, it must eventually crystallize at the ternary eutectic. Early crystallization of olivine moves the composition of the liquid from B to l_2 on the reaction curve, whereupon enstatite begins to form. As the liquid follows the reaction curve, the solids change their bulk composition from olivine toward enstatite. When the liquid reaches l_3, all forsterite must be consumed by reaction. This can be shown by drawing a straight line from the liquid through the bulk composition, B, to the forsterite–enstatite join; this line passes through enstatite, and thus no forsterite remains. The loss of this phase increases the variance by one, and the liquid is free to leave the isobaric univariant reaction curve. Because enstatite is the only phase crystallizing, the liquid moves directly away from enstatite across the enstatite liquidus to l_4, where it intersects the anorthite liquidus. Note that whereas enstatite originally forms only as a reaction product around forsterite, on leaving the reaction curve, it crystallizes directly from the melt and would consequently form euhedral crystals. At l_4, anorthite begins to crystallize, so that the solids are now represented by points on the line En–An. For example, when the liquid first reaches the ternary eutectic, the solids have a composition represented by point s_E. To this composition, a eutectic mixture of En–An–Tr is added, which eventually moves the composition of the solids to the initial starting composition, B, at which point the eutectic liquid is exhausted.

Under equilibrium conditions, liquid of composition B would give rise to a rock composed of phenocrysts of enstatite, smaller phenocrysts of anorthite (because it forms later) in a eutectic groundmass of enstatite, anorthite, and tridymite. If the starting liquid had been a little less rich in anorthite, the fractionating liquid would have encountered the tridymite liquidus before that of anorthite, in which case there would have been phenocrysts of tridymite, and anorthite would have been restricted to the eutectic groundmass. Slight variations in the initial composition of the liquid can result in very different sequences of crystallization (Problem 10.13).

When crystallization involves reactions, disequilibrium commonly results. For example, once olivine is completely rimmed with orthopyroxene, further reaction between liquid and olivine involves solid diffusion through the orthopyroxene. This is an extremely slow process and except in very large, slowly cooled intrusions, is not likely to go to completion. Thus, it would not be surprising in a rock of composition B in Figure 10.23(C) to find orthopyroxene phenocrysts with cores of olivine. The consequence of disequilibrium is to increase the quantity of residual liquid. Indeed, it is possible for liquids plotting in the subtriangle An–Fo–En, such as liquid A, to fractionate all the way to the ternary eutectic. This happens simply because early crystallizing olivine fails to react with the liquid to produce enstatite. Thus, if a liquid at

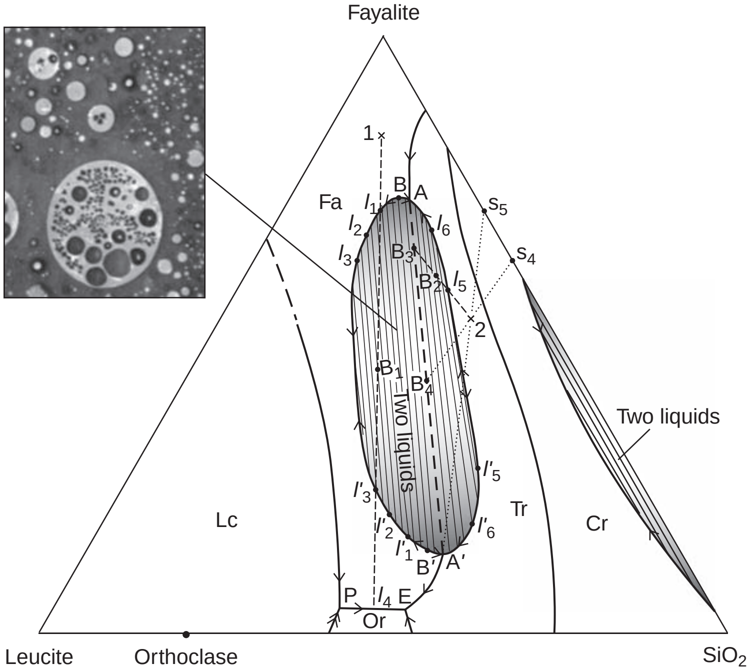


Fig. 10.24 System fayalite–leucite–silica at atmospheric pressure. Tie lines join coexisting liquids across two-liquid fields. See text for discussion. (After Roedder, 1979.) Photomicrograph shows a polished section in reflected light of an experimental run that was rapidly quenched in the middle of the two-liquid field. Various size droplets of the iron-rich liquid (light colored) are enclosed in the silica-rich host (dark).

the peritectic is unable to react with olivine because of kinetic factors, the liquid is free to continue down the cotectic to the ternary eutectic.

Another source of disequilibrium is to have early crystallizing olivine sink and separate from the liquid with which it should react. In this way, a magma of composition A in Figure 10.23(C) could fractionate all the way to the ternary eutectic. This process is extremely important in producing a diversity of igneous rocks in large intrusions where there is sufficient time for separation of early crystallizing minerals from the magma.

10.17 TERNARY SYSTEMS WITH LIQUID IMMISCIBILITY

In Figure 10.23 a small region of the ternary system Fo–An–SiO_2 is labeled "two liquids." Compositions within this field are metastable and should unmix into silica-rich and silica-poor liquids. As indicated in Section 10.7, many silicate systems exhibit this immiscibility, but not those involving feldspar. Thus we see in Figure 10.23 that as soon as a small amount of anorthite is added to the system, the two-liquid field disappears. Similar immiscibility fields occur on the silica liquidus at temperatures near 1690 °C in many silicate systems at atmospheric pressure. This immiscibility occurs at too high a temperature to be of significance in natural occurrences. In some systems, however, this same immiscibility reappears at less silica-rich compositions and at temperatures well within the range found in crustal rocks. We will examine the best known of these, the system fayalite–leucite–silica (Fig. 10.24).

The system Fa–Lc–SiO_2 contains the common high-temperature two-liquid field along the Fa SiO side of the ternary. Addition of only a few percent leucite to these liquids is sufficient to depress the two-liquid field below the silica liquidus. The two-liquid field, however, does extend metastably below this liquidus and reappears in the central lower-temperature region of the ternary diagram. The region of immiscibility can best be visualized as an anticline plunging downward from the Fa–SiO_2 side of the ternary toward the Fa–Lc side. It plunges beneath the liquidus surface of cristobalite on leaving the side of the ternary system but then crops out as an elliptical region in the low-temperature region along the fayalite–tridymite cotectic. The compositions of coexisting liquids across these two-liquid fields are given by the tie lines. Note that the fayalite–tridymite cotectic, on reaching the two-liquid field, reappears on the other side of the two-liquid field at the same temperature; that is, both liquids A and A′ coexist with fayalite and tridymite. On the silica side of this cotectic, immiscible liquids coexist with tridymite, whereas on the fayalite side they coexist with fayalite. The irregular sigmoid trace of the fayalite–tridymite cotectic in this system is a result of the same nonideal mixing in the liquids that produces the immiscibility. Note also that the binary-phase orthoclase melts incongruently, as is evident from the fact that its liquidus field does not extend to the composition of orthoclase; instead, the liquidus of leucite extends out over orthoclase.

Let us first consider the crystallization of liquid 1 in Figure 10.24. Because this liquid lies in the primary field of fayalite, crystallization of this mineral will cause the composition of the liquid to move in a straight line away from the fayalite apex of the ternary system until the liquid reaches the liquidus of orthoclase at l_4. In doing this, however, the composition of the liquid would have to traverse the low-temperature two-liquid field, which is impossible if equilibrium is maintained. Therefore, although the straight line away from the fayalite apex gives the variation in the bulk composition of the liquid, the actual compositions of the liquids are given by the margins of the two-liquid field. Thus, as liquid 1 moves away from the fayalite apex it intersects the two-liquid field at l_1, whereupon droplets of immiscible liquid nucleate in it with composition l'_1. Continued crystallization of fayalite moves the bulk composition of the liquids on the straight line away from fayalite, while the immiscible liquids change along the margin of the two-liquid field. When the bulk composition of the two liquids reaches B_1 for example, the fayalite-rich liquid has changed from l_1 to l_2, and the fayalite-poor one from l'_1 to l'_2. The proportions of these two liquids at this stage are given by the lengths of the lines $B_1l'_2$ and B_1l_2 respectively. The photomicrograph in Figure 10.24 shows a polished section in reflected light of an experimental run that was held in the two-liquid field. As the bulk composition of the liquids moves away from the fayalite apex, the proportion of the fayalite-poor liquid increases. Eventually, the bulk liquid composition reaches l'_3, at which point the last drop of fayalite-rich liquid disappears at l_3. The liquid, now single-phased, proceeds from l'_3 to l_4, whereupon orthoclase begins crystallizing, and the cotectic liquid descends to the ternary eutectic, where fayalite and orthoclase crystallize with tridymite.

It is interesting to note that although during a considerable fraction of the crystallization history of liquid 1 two liquids are present, no textural evidence of this would remain in the final equilibrium crystallization products, which would consist simply of phenocrysts of fayalite in a groundmass of orthoclase, tridymite, and fayalite. Only if immiscible liquids were rapidly quenched would the evidence of immiscibility be preserved. The experimental run products shown in the photomicrograph in Figure 10.24 were quenched by plunging the high-temperature experimental charge into cold water. Immiscible liquids can also be preserved in residual glasses in some basaltic lavas where cooling is too rapid to permit complete crystallization (see Fig. 10.9). Although textural evidence of immiscibility is unlikely to be preserved in coarsely crystalline rocks formed from a liquid that has traversed a two-liquid field, significant fractionation of the two liquids could occur within the two-liquid field. The fayalite-rich and fayalite-poor liquids have very different densities; therefore, they may separate by floating or sinking in each other. This effect could be augmented by the ability of immiscible globules to coalesce and form larger globules, which could then separate more rapidly.

Let us consider the crystallization of one other liquid, 2, which lies in the primary field of tridymite. Crystallization of tridymite causes the composition of the liquid to move away from the silica apex of the ternary system until it reaches the boundary of the two-liquid field at l_5. At this point an immiscible fayalite-poor liquid nucleates with a composition of l'_5. Continued crystallization of tridymite causes the bulk composition of the two liquids to keep moving directly away from the silica apex. Thus when the bulk composition of the two liquids is B for example, the immiscible liquids have compositions l_6 and l'_6. Eventually when the bulk composition of the liquids reaches B_3, the immiscible liquids have compositions A and A′. But these compositions lie on the fayalite–tridymite cotectic, so the tridymite and two liquids are joined by fayalite. These four phases create an isobaric invariant assemblage. Only the proportions of the two liquids, and not their compositions, change as tridymite and fayalite crystallize. For example, when the bulk composition of the liquids has changed to B_4, there will be equal amounts of the two liquids, and the solids will have a bulk composition given by the point s_4 (this point is obtained by drawing a straight line from the bulk composition of the liquids through the initial bulk composition to the line joining silica and fayalite). Eventually, the bulk composition of the liquids reaches A′, at which point the last drop of liquid A disappears and the solids have the composition s_5. Continued crystallization of tridymite and fayalite proceeds as the remaining liquid moves down the cotectic to the ternary eutectic.

In this second example, crystallization causes the liquids to intersect the tridymite–fayalite cotectic in the two-liquid field. Although liquids A and A′ are in equilibrium with both tridymite and fayalite under these conditions, crystallization involves largely the formation of fayalite from liquid A. This is evident from the fact that the tangent to the cotectic within the two-liquid field passes almost through fayalite. We can think of liquid A as being converted into a mixture of crystals of fayalite and liquid A′. As a consequence, crystals of fayalite grow almost entirely in the fayalite liquid – the source of their nutrients. These crystals are in equilibrium with the fayalite-poor liquid, but they do not grow from it, at least not during the two-liquid stage.

10.18 TERNARY SYSTEMS WITH ONE BINARY SOLID SOLUTION WITHOUT A MINIMUM

Many ternary systems contain phases that belong to solid solution series. Systems containing MgO and FeO, for example, will have ferromagnesian minerals that typically exhibit some range of solid solution, and systems containing alkalis and CaO may contain feldspar solid solutions. In this and the following two sections, we will consider the phase relations in systems containing solid solutions. We consider first a system where one of the binaries that make up the ternary system exhibits complete solid solution. In following sections we consider solid solutions with a minimum, partial solid solution, and finally, more than one solid solution.

One of the most important ternary systems in petrology is that of diopside–albite–anorthite (Fig. 10.25). One of the binaries that bounds this system is that of the plagioclase feldspars, and the third component, diopside, is one end member of the clinopyroxene solid solution series. This system therefore contains two of the most important minerals in basalts, and liquids in this system can be considered representative of simplified basaltic magmas.

An easy way to visualize the phase relations in the ternary system Di Ab An is to start with the binary plagioclase

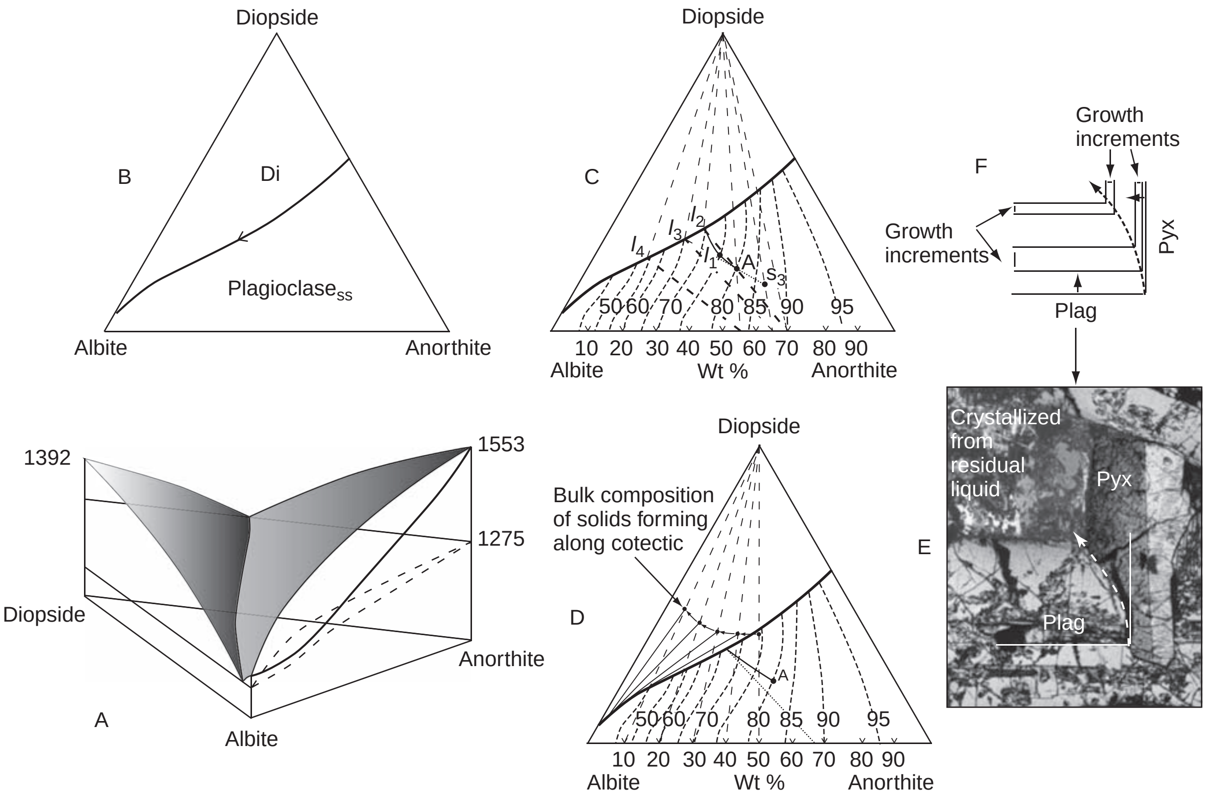


Fig. 10.25 One-atmosphere ternary system diopside–albite–anorthite shown in perspective (**A**) and projected onto the compositional triangle (**B**). In (**C**) the liquidus surface of plagioclase is ruled with dashed lines marking liquid compositions that coexist with plagioclase of the designated composition. The liquid line of descent from A, through l_1, l_2, l_3, to l_4 results from equilibrium crystallization of liquid A. In (**D**) the complete fractional crystallization path of liquid A is shown. Photomicrograph shows curving boundary between plagioclase and augite crystals resulting from curvature of cotectic. See text for discussion. (After Bowen, 1915a; Barron, 1972; Weill *et al.*, 1980.)

system (see Section 10.9 and Fig. 10.12). Addition of diopside to any composition on the plagioclase liquidus lowers the liquidus temperature according to the cryoscopic equation. The plagioclase liquidus–solidus loop in a system containing, for example, 5 wt% diopside looks almost the same as that in the pure plagioclase binary, except that it moves to lower temperatures, especially at the anorthite end. This lowering of temperature of the plagioclase liquidus continues until liquids become saturated in diopside. This occurs along the cotectic marking the boundary between the fields of plagioclase and diopside. The composition of plagioclases coexisting with liquids on this cotectic and the projected composition of the cotectic liquids are shown by the dashed lines on the binary plagioclase system (Fig. 10.25(A)).

Because the plagioclase liquidus has no minimum, the temperature on the cotectic line in the system Di–Ab–An decreases continuously from the anorthite–diopside eutectic to the diopside–albite eutectic. Consequently, this system has no ternary minimum or eutectic; liquids on the cotectic simply fractionate toward the diopside–albite eutectic. The resulting phase diagram is simple, with just two liquidus surfaces, one for diopside and the other for plagioclase (Fig. 10.25(B)).

Let us consider the equilibrium crystallization of a liquid having the composition A in Figure 10.25(C). This liquid first crystallizes plagioclase on cooling. Unfortunately, the composition of the plagioclase that forms initially cannot be read directly from a simple diagram such as that in Figure 10.25(B). Although we might suspect that the first-formed plagioclase would be relatively anorthite-rich, its precise composition cannot be determined unless the diagram indicates the compositions of plagioclase that coexist with liquids on the plagioclase liquidus in the ternary system. The simplest way to do this is with lines on the plagioclase liquidus joining liquid compositions that coexist with a particular plagioclase composition (Fig. 10.25(C)). Such lines indicate that liquid A first crystallizes plagioclase of An_{80} composition.

Crystallization of An_{80} from liquid A would cause the liquid to change its composition in a straight line away from An_{80}. But as the composition changes, the coexisting plagioclase must become more albitic, and this, in turn, causes the path followed by the liquid to curve towards the plagioclase–diopside cotectic. This path can be deduced quite simply from Figure 10.25(C). For example, when the plagioclase has a composition of An_{75}, the liquid must lie on the line marking the composition of liquids that can coexist with An_{75}. At the same time, the line joining the liquid and An_{75} must pass through the initial bulk composition, A. The liquid at this stage must therefore have a composition of l_1. Similarly, we can deduce that when the liquid reaches the cotectic at l_2, the coexisting plagioclase will have a composition of An_{69}.

Once the cooling liquid reaches the cotectic, coprecipitation of plagioclase and diopside causes the liquid to descend the cotectic. The three coexisting phases, liquid, plagioclase, and diopside, can be joined by lines to form a *phase triangle*. The first three-phase triangle joins l_2 with diopside and An_{69}. As the liquid descends the cotectic, the phase triangle pivots from its diopside apex as the coexisting plagioclase becomes more sodic. When the liquid reaches l_3, for example, the phase triangle joins plagioclase of An_{65} composition with diopside. The bulk composition of the solids at this stage must lie on the diopside–plagioclase side of this phase triangle. At the same time, a line drawn from the bulk composition of the solids to the liquid at l_3 must pass through the initial bulk composition, A. The solids coexisting with l_3 will therefore have a composition of s_3.

As the liquid continues to descend the cotectic, the three-phase triangle moves to the left, and the diopside–plagioclase side of this triangle moves progressively closer to the initial bulk composition. When the diopside–plagioclase join passes through point A, the last drop of liquid must crystallize. The composition of this final liquid is determined by noting that liquid of composition A must eventually crystallize to a mixture of diopside and plagioclase of composition An_{55} (this is determined by drawing a straight line from diopside through A to the plagioclase binary). The only liquid that can coexist with An_{55} and diopside is l_4. This, then, is the final liquid to crystallize from the starting composition A under equilibrium conditions (Problem 10.15).

Equilibrium crystallization requires that plagioclase remain in contact with the liquid and that crystals continuously adjust their composition to be in equilibrium with the liquid. Solid diffusion rates in plagioclase are, however, very low, even at magmatic temperatures, as is evidenced by the common occurrence of zoned plagioclase crystals in igneous rocks. Equilibrium crystallization is therefore likely to occur in nature only under conditions of very slow cooling. Disequilibrium, or fractional crystallization, is more likely to occur.

Let us again consider the crystallization of liquid A, but this time under conditions of complete fractionation; that is, crystals do not react with the liquid. This occurs, for example, if crystals separate from the liquid as they form, or if crystals become mantled with layers through which diffusion is unable to transfer the reactants in the time available.

Liquid A again first crystallizes plagioclase of An_{80} composition, which causes the liquid to change composition away from An_{80} (Fig. 10.25(D)). Consequently, the plagioclase becomes more sodic, but because it does not react with the liquid, the path followed by the fractionating liquid is less curved than under equilibrium conditions. In addition, the line joining the liquid and coexisting plagioclase does not pass through the initial bulk composition, but lies to the left of it. This is because a fraction of the initial bulk composition is lost to the early crystallizing plagioclase, which either is removed from the system or is hidden in the cores of zoned crystals. The fractionating liquid reaches the cotectic with diopside at a more sodic composition than it does under equilibrium conditions. The liquid then follows the cotectic, and because the crystals are prevented from reacting, the liquid fractionates all the way to the diopside–albite eutectic. The result is that fractional crystallization leads to a much wider range of liquid compositions than does equilibrium crystallization.

Equilibrium and complete fractional crystallization are two extreme ways of crystallizing a melt. In nature, crystallization is likely to follow some intermediate path. Volcanic and hypabyssal rocks tend to form under conditions closer to those of fractional crystallization, whereas plutonic ones form under conditions approaching those of equilibrium crystallization.

Many basaltic rocks form from liquids similar to those on the plagioclase–diopside cotectic. The common *ophitic texture*, where plagioclase laths are embedded in large crystals of clinopyroxene, is believed to result from the cotectic crystallization of these minerals. The abundance of plagioclase and clinopyroxene in this cotectic mixture, of course, depends on where the tangent to the cotectic line intersects the plagioclase–diopside (clinopyroxene) join. Along much of the cotectic, the tangent intersects the join at about 40 wt% plagioclase, which is similar to the amount found in typical ophitic intergrowths. Toward the low-temperature end of the cotectic, however, the tangent swings progressively toward pyroxene (Fig. 10.25(D)); that is, more pyroxene and less plagioclase crystallizes. Evidence of this changing proportion can be seen in the curvature of pyroxene–augite grain boundaries as they grow into patches of residual liquid in basaltic rocks. The photomicrograph in Figure 10.25, for example, shows a zoned crystal of plagioclase attached approximately at right angles to a zoned crystal of augite. The boundary between these two crystals is essentially a graph of the growth rate of one crystal against that of the other. Because the proportion of pyroxene to plagioclase crystallizing from the fractionating liquid increased, the grain boundary is concave toward the plagioclase. Thus, preserved in this grain boundary is a record of the changing fractionation trend of the clinopyroxene–plagioclase cotectic in this natural system.

10.19 TERNARY SYSTEMS WITH ONE BINARY SOLID SOLUTION WITH A MINIMUM

A ternary system in which a binary solid solution exhibits a liquidus minimum has either a ternary minimum or ternary eutectic. The system SiO_2–albite–orthoclase provides an important example of such a system, for it indicates the phase relations in simplified granites. Figure 10.26(A) is a perspective view of this system at low pressure. To simplify the illustration, orthoclase has been shown as melting congruently, where in fact it melts incongruently to leucite and a silica-rich liquid (Fig. 10.10). At pressures above 0.3 GPa, however, orthoclase does melt congruently, and even at lower pressures, the simplification is of no concern as long as orthoclase-rich compositions are avoided.

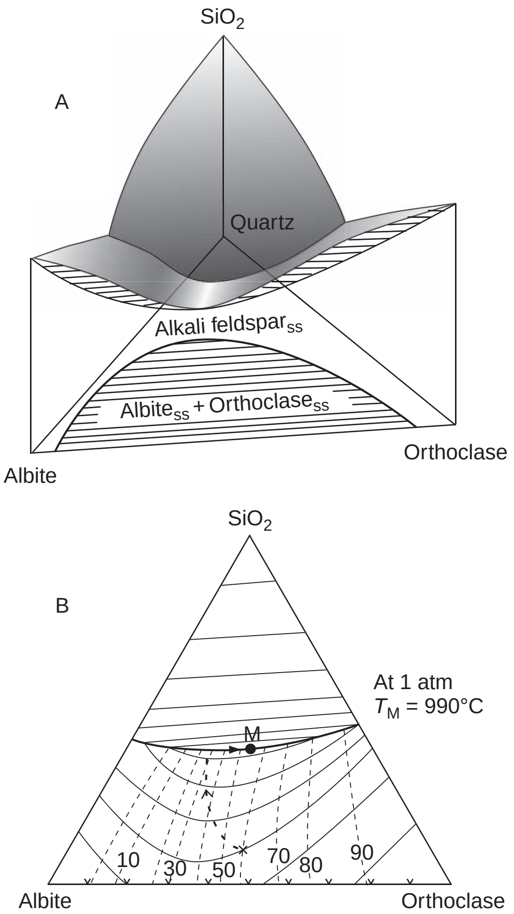


Fig. 10.26 (**A**) Perspective representation of ternary system SiO_2–albite–orthoclase at low pressure. Incongruent melting of orthoclase to leucite plus liquid is omitted. The system has a minimum at M. (**B**) Projection of system onto the compositional triangle. Schematic isotherms on liquidus shown with solid lines. Dashed lines join liquid compositions that coexist with a particular composition of alkali feldspar. Heavy dashed line traces fractional crystallization path of liquid *X*. (After Tuttle and Bowen, 1958; Barron, 1972.)

The liquidus minimum on the feldspar boundary extends into the ternary system as a U-shaped valley on the feldspar liquidus. A *ternary minimum* (M) is created where this valley intersects the liquidus of the silica phase. Tridymite is the stable silica polymorph at the minimum under atmospheric pressure, but with a slight increase in pressure (0.02 GPa), quartz becomes stable. Under most conditions, then, quartz and a single alkali feldspar crystallize together at the ternary minimum. As in the system diopside–albite–anorthite, the precise composition of the feldspar crystallizing cannot be read from the simple phase diagram, unless the feldspar liquidus is contoured with lines joining liquids that coexist with a specific feldspar composition (Fig. 10.26(B)). With such lines, the paths followed by liquids during crystallization are easily determined (Problem 10.16). In general, these paths trend toward and then down the U-shaped valley to the ternary minimum. Under equilibrium conditions, of course, the line joining a liquid and its coexisting feldspar must, at all times, pass through the initial bulk composition, as was explained in the system diopside–albite–anorthite. Under fractional crystallization, however, a liquid will descend more rapidly to the bottom of the U-shaped valley, and then turn down the valley to the minimum (Fig. 10.26(B)).

Although the liquidus minimum on the feldspar binary is well above the alkali feldspar solvus, ternary liquids descending the U-shaped valley crystallize feldspars that get progressively closer to the critical temperature on the solvus; they do not, however, reach it, for the ternary minimum at atmospheric pressure is 990 °C, whereas the critical temperature on the solvus is 648 °C (see Problem 9.4). Consequently, at low pressures a single feldspar crystallizes from low-temperature liquids in the ternary system. In nature, such liquids give rise to granites that contain a single alkali feldspar. Because these crystallize at temperatures above the alkali feldspar solvus, they are known as *hypersolvus* granites (Fig. 10.16).

With increasing pressure the critical temperature on the alkali feldspar solvus rises (see Problem 9.4), but so do liquidus temperatures. Pressure by itself, then, does not significantly change the phase relations in the SiO_2–albite–orthoclase system. If water is present, however, increased pressure causes the water to dissolve in the melt, which, in turn, lowers the liquidus temperatures because of the decreased chemical potentials of the other components. The water, however, does not change the minerals that crystallize from the melt, and so we can still plot the phase relations in a ternary diagram even though the system, which now consists of SiO_2–albite–orthoclase–H_2O, is quaternary. Figure 10.27 shows the water-saturated liquidus-phase relations in this quaternary system at two different pressures.

At low pressure, no significant amount of water dissolves in the liquids in the system SiO_2–albite–orthoclase–H_2O, so the phase relations are identical to those shown in Figure 10.26 for the "dry" system. With increased pressure, however, the critical temperature on the alkali feldspar solvus increases, whereas the liquidus temperatures decrease, as long as water is present in excess (water-saturated). When the pressure reaches 0.4 GPa, the water-saturated minimum drops below the critical temperature on the alkali feldspar solvus. Two feldspars, consequently, must crystallize from this liquid, which changes the minimum into a ternary eutectic. The compositions of these two feldspars are given by the coexisting alkali feldspars on the solvus at the eutectic temperature.

Extending up from the ternary eutectic in Figure 10.27(A) are three cotectics, quartz–albite$_{ss}$, quartz–orthoclase$_{ss}$, and albite$_{ss}$–orthoclase$_{ss}$. As the alkali feldspar cotectic extends away from the ternary eutectic, its temperature rises until it reaches the critical temperature for that pressure. At this point, a single alkali feldspar crystallizes from the melt rather than two, and the boundary between albite$_{ss}$ and orthoclase$_{ss}$ ceases to exist; this is a critical point on the boundary curve. Below the critical point, the boundary curve marks the bottom of a V-shaped valley on the liquidus. With rising temperature, the angle of the V decreases, and at the critical point the valley becomes U-shaped and continues as such to the minimum on the alkali feldspar side of the system. As the pressure increases, the critical end point on the alkali feldspar cotectic

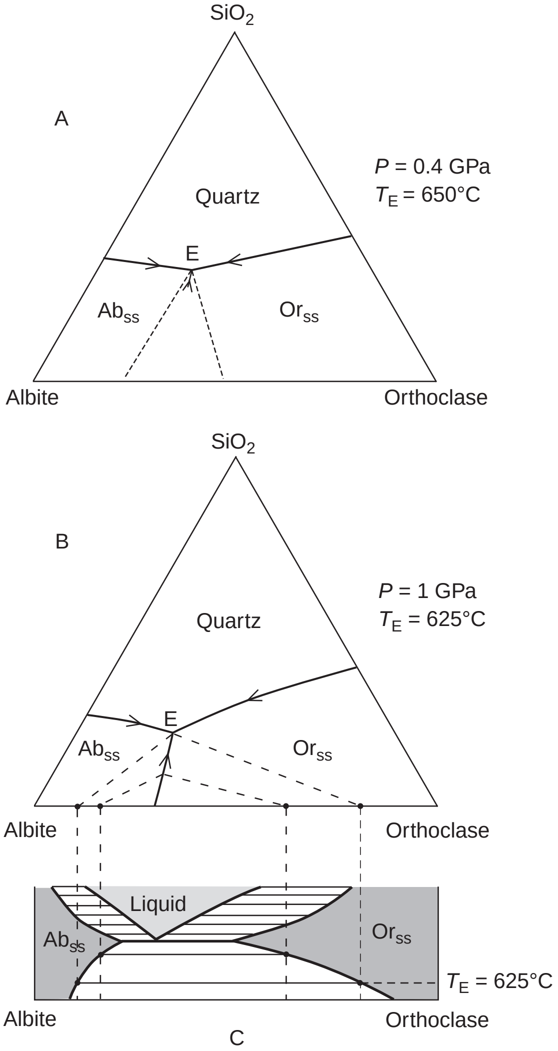


Fig. 10.27 Ternary projection of water-saturated liquidus relations in the quaternary system SiO_2–Ab–Or–H_2O at 0.4 GPa (**A**) and 1.0 GPa (**B**). At these pressures, the system has a eutectic (E) rather than a minimum, because two alkali feldspars coexist with quartz. Compositions of feldspars coexisting with particular liquids are joined with dashed lines. Positions of these feldspars on the alkali feldspar solvus are shown in (**C**). (After Tuttle and Bowen, 1958; Luth *et al.*, 1964.)

gets progressively closer to the alkali feldspar join, and eventually reaches it. At 1 GPa, for example, an alkali feldspar cotectic extends all the way from the ternary eutectic to the feldspar binary (Fig. 10.27(B)).

The compositions of feldspars crystallizing from liquids on the alkali feldspar cotectic in the water-saturated 1-GPa system can be read directly from the compositions on the solvus at the temperature of the liquid (Fig. 10.27(C)). Two such pairs of compositions are shown for a liquid partway down the cotectic and another at the ternary eutectic in Figure 10.27(B) and (C). Natural granites that crystallize under water-saturated conditions also crystallize two alkali feldspars, an albite solid solution and an orthoclase solid solution (Fig. 10.27(B)). Because this crystallization takes place below the solvus, these rocks are known as *subsolvus* granites (Fig. 10.16). Such rocks clearly indicate the presence of water in the magma, which, in turn, indicates that the magma must have been under considerable pressure, and hence at considerable depth, at the time of formation.

All degrees of water saturation are possible in granitic magmas. At high pressure, if the concentration of water in the melt is decreased, liquidus temperatures rise, and the phase relations resemble those at lower pressures, where a critical point exists on the alkali feldspar cotectic. At still lower concentrations, the ternary eutectic is replaced by a minimum and a single feldspar crystallizes. We can conclude, therefore, that hypersolvus granites are dry, whether formed at low or high pressures, and subsolvus granites are wet, which requires at least moderate pressures to keep the water in solution in the magma.

10.20 TERNARY SYSTEMS WITH MORE THAN ONE SOLID SOLUTION SERIES

In this section, we consider two important phase diagrams, one involving the feldspars and the other involving the pyroxenes. The feldspar diagram involves two solid solution series, the alkali feldspars and the plagioclase feldspars. The pyroxene diagram also contains two prominent solid solution series, one involving the common augites, and the other involving the orthopyroxenes. The pyroxene diagram is made more complicated by the fact that the two solid solution series exhibit considerable solid solution between each other. In addition, a third pyroxene series – pigeonites – can form under certain conditions.

The feldspar ternary system is very much like the diopside–albite–anorthite system in that it is divided in two by a cotectic, which extends down into the ternary system from the anorthite–orthoclase eutectic. Unlike the diopside–plagioclase cotectic, however, the orthoclase$_{ss}$–plagioclase cotectic does not extend all the way to the alkali feldspar side of the ternary system. Instead, it is terminated at a critical endpoint (c in Fig. 10.28), which is produced by the complete solid solution that occurs between feldspars at liquidus temperatures in the alkali-rich part of this system.

Feldspars near the alkali feldspar side of the ternary feldspar system form a continuous solid solution series at liquidus temperatures. Indeed, at atmospheric pressure the critical temperature on the alkali feldspar solvus (K in Fig. 10.28 (B)) is 200 °C below the liquidus minimum. Addition of anorthite, however, causes the solvus to rise and expand until, on the anorthite–orthoclase side of the system, the solid solution is very limited. The solvus in the ternary system can be visualized as an anticline plunging from the anorthite–orthoclase side down to the albite–orthoclase side of the ternary system (Fig. 10.28(B)). The crest of this anticline is the trace of the critical temperature on the ternary feldspar solvus (*K*–*K*′–K′ in Fig. 10.28(B)). At *K*′, the critical temperature on the solvus reaches the solidus in the ternary system and produces the critical endpoint on the plagioclase–orthoclase boundary (c).

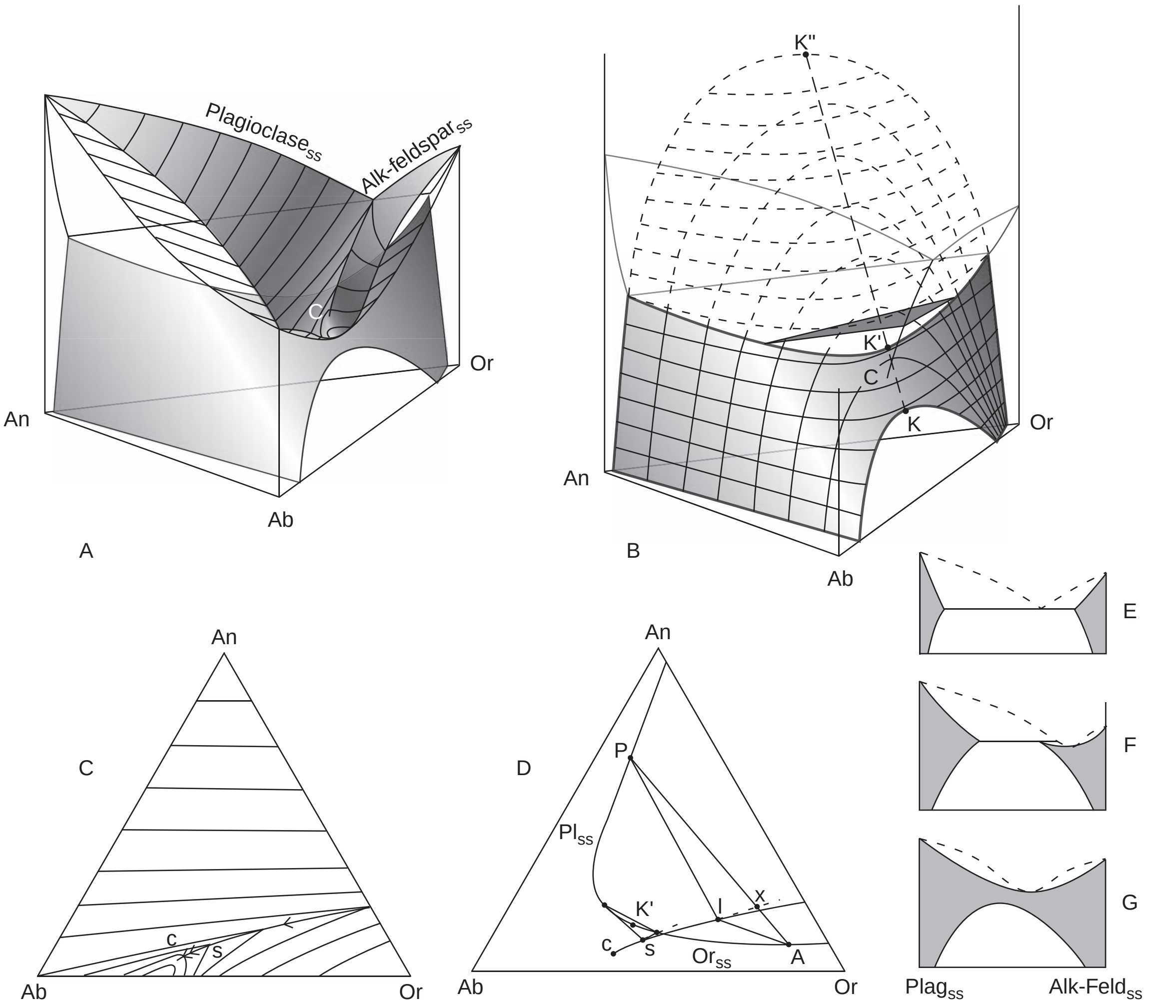


Fig. 10.28 Schematic representation of phase relations in system anorthite–albite–orthoclase at low pressure. (**A**) Liquidus surface and boundary curve with critical endpoint (c) between fields of plagioclase and alkali feldspar. Curved horizontal line at top of shaded surface is intersection of the ternary feldspar solvus with the ternary solidus. (**B**) Ternary feldspar solvus. Line K-K′-K″ joins critical points on solvus at different alkali feldspar contents. Unshaded part of solvus is metastable with respect to liquid. (**C**) Isotherms and boundary curve on feldspar liquidus surface. Above point s the boundary is cotectic; below s it is a reaction curve. (**D**) Three different liquid–feldspar phase assemblages; see text for explanation. (**E–G**) Pseudobinary sections through ternary feldspar system above point s (**E**), between point s and c (**F**), and between c and the Ab–Or side of diagram (**G**).

On the higher-temperature part of the boundary curve in the feldspar system, plagioclase and potassium-rich alkali feldspar coprecipitate along a cotectic. The composition of the alkali feldspar and plagioclase coexisting with a liquid such as *l* in Figure 10.28(D) would be A and P, respectively. The proportions in which these feldspars crystallize would be given by Px/xA, respectively, x being the point where the tangent to the cotectic at *l* intersects the join AP. The phase relations along the upper part of this boundary curve can be thought of in terms of a simple pseudobinary diagram with plagioclase$_{ss}$ and alkali feldspar$_{ss}$ crystallizing on either side of the solvus from a cotectic liquid (Fig. 10.28(E)). The representation is not truly binary because if the solid phases plot in the plane of the diagram, the liquid must plot at more albitic compositions. For this reason the liquidus in Figure 10.28(E) is dashed.

As liquids descend the cotectic, the point of intersection of the tangent line with the plagioclase alkali feldspar join shifts toward the alkali feldspar composition until, at point s, it actually passes through the alkali feldspar. Beyond this point the boundary between the fields of plagioclase and alkali feldspar is a reaction curve, with plagioclase reacting with the liquid to produce alkali feldspar, as represented in the pseudobinary diagram in Figure 10.28(F). This reaction relation explains the common rimming of plagioclase by alkali feldspar. Below point s, the lowest temperature in the pseudobinary diagram is no longer on the boundary curve, but is at the minimum, which occurs in the alkali feldspar field. This can also be seen from the shape of the isotherms in Figure 10.28(A) and (C).

As liquids descend the boundary curve still further, compositions of coexisting feldspars approach one another until at point *K′*, the critical temperature on the ternary feldspar solvus, they become identical, that is, only one feldspar crystallizes. This point is the critical endpoint on the plagioclase–alkali feldspar boundary (c in Fig. 10.28). Between this point and the alkali feldspar side of the ternary system the phase relations are as shown in the pseudobinary diagram of Figure 10.28(G).

In the low-temperature part of the ternary feldspar system, then, feldspar can crystallize in three different ways. Along most of the boundary curve plagioclase and potassium-rich alkali feldspar crystallize together. Between point s and the critical endpoint, however, alkali feldspar forms as a reaction rim around plagioclase. Finally, below the critical endpoint, only a single alkali feldspar forms. All three modes of crystallization are commonly encountered in igneous rocks.

Some igneous rocks exhibit feldspar phase relations that differ from those shown in Figure 10.28. This is because most magmas are more complex systems than the simple ternary one. For example, if water is present and the pressure is sufficient to dissolve the water in the melt, liquidus temperatures can be lowered enough to make the cotectic boundary extend all the way through the system from the anorthite–orthoclase eutectic to an albite$_{ss}$–orthoclase$_{ss}$ eutectic. In some rocks, potassium-rich alkali feldspar is rimmed by plagioclase – the *rapakivi* texture (Fig. 4.79(E)) – suggesting a reaction relation that is the reverse of the one shown in Figure 10.28(F). A discussion of all the possible phase relations in the feldspars is given by Abbott (1978).

Pyroxene phase relations can be discussed most simply in terms of the ternary system CaSiO$_3$–MgSiO$_3$–FeSiO$_3$ (Fig. 10.29(A)). Most of the common pyroxenes have compositions that plot in the lower half of this system; that is, they plot below the line joining the compositions of diopside (CaMgSi$_2$O$_6$) and hedenbergite (CaFeSi$_2$O$_6$). The lower half of this ternary plot is commonly referred to as the *pyroxene quadrilateral*.

Common augite, which is monoclinic (clinopyroxene = Cpx), has compositions that plot between diopside and hedenbergite. Augite that coexists with calcium-poor pyroxene typically contains significantly less than 50 mol% CaSiO$_3$, especially at intermediate Mg/Fe ratios. Pigeonite is also monoclinic but generally contains only about 11 mol% CaSiO$_3$. At high pressure, pigeonite can range in composition from the magnesium to the iron end member. In most rocks formed under crustal conditions, however, pigeonite is restricted to intermediate compositions of magnesium and iron. Orthorhombic pyroxenes (Opx), which typically contain between 4 and 5 mol% CaSiO$_3$, can have compositions ranging from the magnesium end member (enstatite) to iron-rich compositions. The iron end member, ferrosilite, is not stable under crustal pressures (<1.15 GPa); instead, it forms fayalite plus quartz.

Before considering the ternary pyroxene phase relations, it is important to understand the enstatite–diopside binary (Fig. 10.29(B)). Addition of FeSiO$_3$ to this system lowers temperatures, but it does not significantly change relations between the phases. The enstatite–diopside binary can therefore be used as a model for pseudobinary diagrams drawn at any constant Mg/(Mg + Fe) ratio in the ternary system.

At low temperatures, enstatite is orthorhombic and shows limited solubility toward diopside. At high temperatures, enstatite inverts to a monoclinic form, which is able to dissolve considerably more diopside. This high-temperature form is in fact pigeonite. Other polymorphs of enstatite are known, but they are omitted here for simplicity and because they have little effect on the ternary pyroxene phase relations. In Figure 10.29(B) the clinopyroxene on the diopside side of the binary is labeled augite$_{ss}$ in order to simplify correlation with the ternary system. It shows limited solubility toward MgSiO$_3$. At low temperature, it can coexist with orthoenstatite, and at high temperature, with clinoenstatite (pigeonite). This binary diagram, therefore, has two different solvi, a low-temperature one between orthopyroxene and augite and a high-temperature one between pigeonite and augite. The high-temperature one intersects the solidus to create a eutectic between pigeonite$_{ss}$ and augite$_{ss}$.

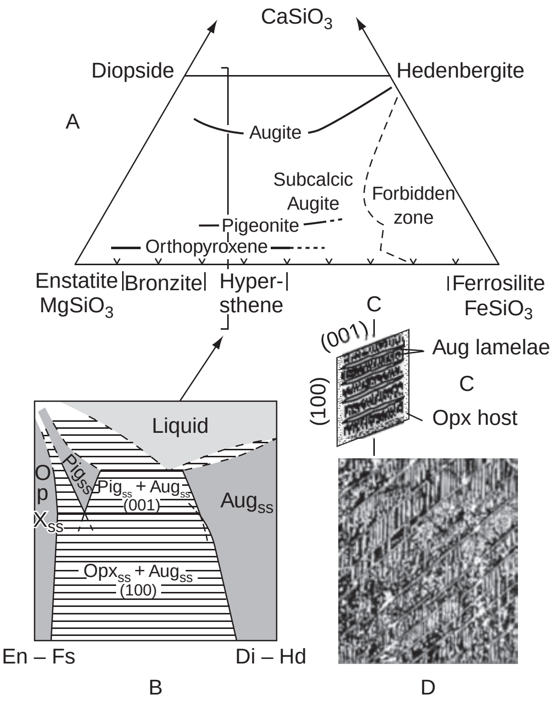


Fig. 10.29 (**A**) Compositional range of common pyroxenes in terms of CaSiO$_3$–MgSiO$_3$–FeSiO$_3$. Pyroxenes within "forbidden zone" are unstable with respect to calcium-rich clinopyroxene + fayalitic olivine + quartz. This zone shrinks with increasing pressure, disappearing above 1.15 GPa. (**B**) Simplified schematic pseudobinary representation of phase relations near the En–Di side of the pyroxene quadrilateral. A high-temperature polymorph of enstatite and a small field of incongruent melting involving olivine have been omitted. Exsolution lamellae involving pigeonite and augite approximately parallel (001), whereas those involving opx and augite approximately parallel (100). (**C**) Orientation of exsolution lamellae in pigeonite that slowly cools and inverts into opx to form inverted pigeonite. (**D**) Photomicrograph of inverted pigeonite in norite, Lake St. John anorthosite massif, Quebec.

Coexisting calcium-rich and calcium-poor pyroxenes, on cooling slowly, develop exsolution lamellae. Pigeonite and augite exsolve from one another on planes that very nearly parallel (001), whereas orthopyroxene and augite exsolve from one another on planes that very nearly parallel (100). Pigeonite that cools slowly first develops fine (001) lamellae of augite. On cooling to the clino- to ortho-inversion temperature (Fig. 10.29(B)), the pigeonite must exsolve a significant amount of augite in order to change to the stable orthopyroxene composition. It does this by exsolving augite onto the thin lamellae that form during the initial cooling through the pigeonite–augite solvus. This thickens these lamellae, which are eventually preserved in the orthopyroxene host that forms by inversion from the pigeonite (Fig. 10.29(C) and (D)). More augite exsolves from the orthopyroxene as it continues to cool, but this takes place as thin lamellae nearly parallel to (100). Orthopyroxene that contains the thick exsolution lamellae of augite at an oblique angle to the *c* axis is referred to as *inverted pigeonite*. It is particularly common in slowly cooled subalkaline gabbroic intrusions. Pigeonite itself is restricted to subalkaline lavas and shallow intrusions, where it cools too rapidly to invert to the orthorhombic form.

The phase relations on the enstatite–diopside binary decrease by about 6 °C for each 1 mol% increase in Fe/(Fe +Mg) on extending into the ternary system (Fig. 10.30). For example, the inversion of pigeonite to orthopyroxene + augite, which occurs at just over 1300 °C in the iron-free system, occurs at only 1100 °C when Mg/(Mg+Fe) is 0.7. Similarly, the critical temperature on the pigeonite–augite solvus decreases from 1460 °C to 1300 °C over this same interval. Many subalkaline basaltic magmas have magnesium numbers (Mg/(Mg + Fe) above 0.7 and temperatures near 1200 °C. Consequently, they start crystallizing below the stability field of pigeonite with the formation of augite + orthopyroxene (Fig. 10.30). With cooling and fractional crystallization, these magmas become more iron rich, following paths such as the one shown by the dotted line in Figure 10.30. The minimum stability temperature of pigeonite, however, decreases more rapidly with iron enrichment than does the temperature of most fractionating magmas. Consequently, pigeonite becomes stable during the later stages of crystallization of these magmas, with the result that the assemblage augite + orthopyroxene is replaced by augite + pigeonite. The precise composition at which this change occurs depends on when the fractionating magma crosses the inversion curve. In many magmas it takes place when Mg/(Mg + Fe) is approximately 0.7. The pigeonite formed during the later stages of crystallization, on slow cooling, changes to inverted pigeonite.

The only significant difference between the phase relations in the binary and ternary systems is that coexisting phases in the ternary system may have different values of Mg/(Mg + Fe). These values can be shown clearly in ternary isothermal sections, such as the one illustrated in Figure 10.31 for 1200 °C. Here pigeonite is seen to be more iron-rich than coexisting augite and orthopyroxene. Similarly, the cotectic liquid that coexists with augite and pigeonite is considerably more iron-rich than the crystalline phases.

Because magmas contain more components than do liquids in the simple pyroxene quadrilateral, they have lower melting points. The actual liquidus temperatures and liquidus phase relations in the pyroxene quadrilateral are, therefore, not of great importance to the interpretation of pyroxenes in natural rocks. But many of the additional components in magmas do not exhibit significant solid solution in the pyroxenes, and therefore their presence does not substantially change the subsolidus phase relations in the pyroxene quadrilateral. Consequently, the subsolidus relations can be used to interpret pyroxenes that crystallize from complex

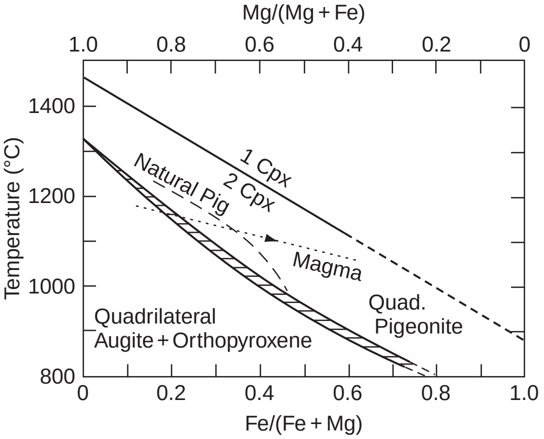


Fig. 10.30 Variation of critical temperature on augite–pigeonite solvus and minimum stability of pigeonite as a function of mole proportions of Mg/(Mg+Fe) in the Di–En–Hd–Fs system at atmospheric pressure. Curves are dashed at high iron contents to show that pyroxenes within "forbidden zone" are unstable at low pressure. Long-dashed line indicates minimum stability of natural pigeonites containing nonquadrilateral components. Dotted line indicates typical fractionation bath of basaltic magma. (After Lindsley, 1983.)

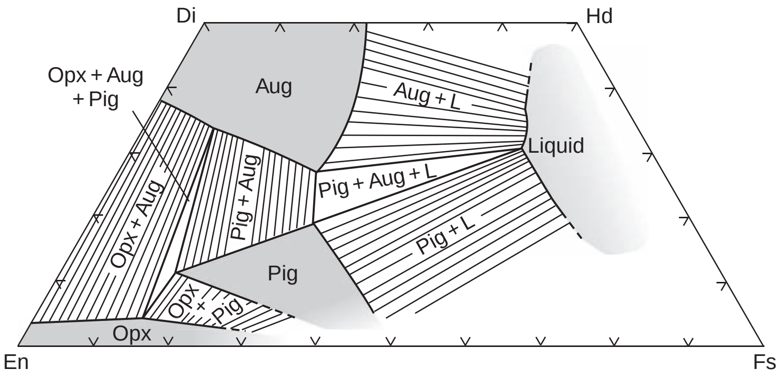
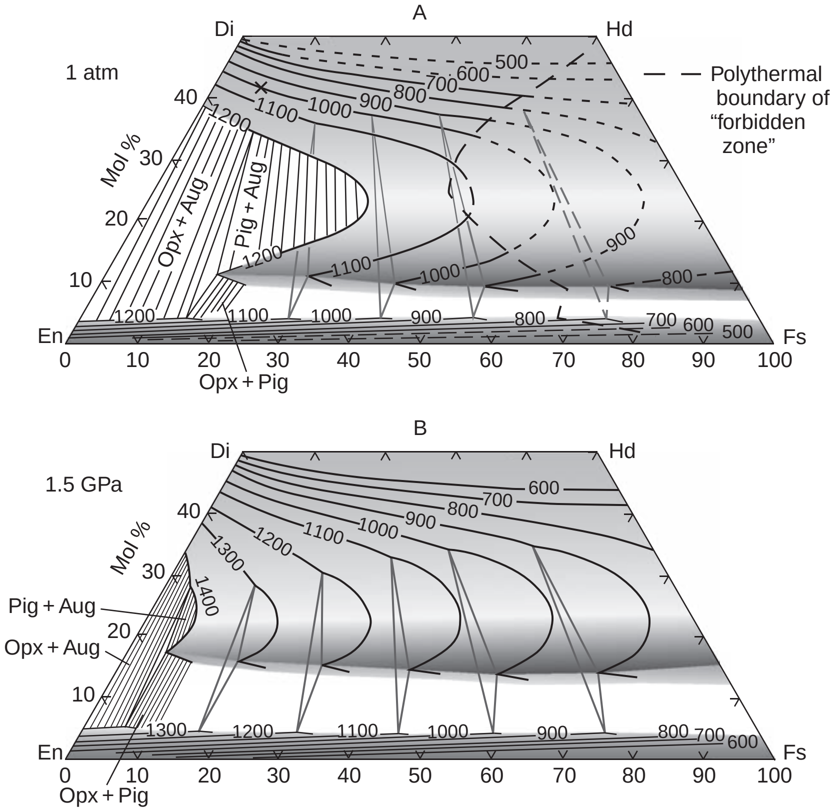


Fig. 10.31 Isothermal section through system Di–En–Hd–Fs at 1200 °C and atmospheric pressure. (After Lindsley, 1983.)

Fig. 10.32 Polythermal relations for augite + opx, augite + pigeonite, and augite + opx + pigeonite (triangle) in the Di–En–Hd–Fs system at atmospheric pressure (**A**) and 1.5 GPa (**B**). The augite + pigeonite and augite + orthopyroxene solvi are shaded and contoured with isotherms with temperatures in °C. Tie lines are drawn only on 1200 °C isothermal section in (A) and on the 1400 °C isothermal section in (B). (Drawn from Lindsley, 1983; published with permission of *American Mineralogist*.)



magmas at temperatures well below the liquidus in the pyroxene quadrilateral. Moreover, compositions of the pyroxenes provide useful geothermometers in subalkaline rocks (Lindsley, 1983).

A complete three-dimensional representation of the subsolidus phase relations in the pyroxene quadrilateral would be far too complicated to be of great use. For purposes of geothermometry, however, we need include only the augite–orthopyroxene and augite–pigeonite solvi and the reaction of pigeonite to form orthopyroxene + augite. These phase relations can be illustrated in a series of *isothermal sections* which when superimposed on one another form a *polythermal diagram* (Fig. 10.32).

Let us consider the 1200 °C isothermal section at atmospheric pressure (Fig. 10.32(A)). Coexisting pyroxenes with Mg/(Mg+Fe) > 0.85 consist of augite and orthopyroxene, with compositions that plot on the solvus. Compositions slightly more iron-rich than this consist of augite + orthopyroxene + pigeonite (three-phase triangle in Fig. 10.32). This assemblage represents the reaction of the high-temperature pigeonite to the low-temperature assemblage of augite + orthopyroxene. Note that in the isothermal section, the pigeonite is more iron-rich than the coexisting pyroxenes (see also the T versus X representation in Fig. 10.30). Compositions more iron-rich than the three-phase triangle consist of augite and pigeonite, but with increasing iron content the compositions of coexisting augite and pigeonite approach one another until only a single clinopyroxene (*subcalcic augite*) forms when Mg/(Mg+Fe) < 0.6. At still more iron-rich compositions, pyroxenes plotting to the right of the dashed line in Figure 10.32 are not stable with respect to calcium-rich augite + quartz + fayalitic olivine. This part of the pyroxene quadrilateral is referred to as the *forbidden zone*. It becomes smaller with increasing pressure and disappears above 1.15 GPa.

At lower temperatures, the phase relations are similar, except that augite has a higher calcium content and orthopyroxene and pigeonite have lower calcium contents. An analysis of a single pyroxene from any coexisting pair provides an estimate of the temperature at which this pair last equilibrated. Each pair provides two independent estimates of this temperature, and if three pyroxenes are present, three estimates can be made. Orthopyroxene compositions, however, are not very sensitive to temperature changes, as is evident from the close spacing of the isotherms on the orthopyroxene side of the augite–orthopyroxene solvus. Finally, if only a single type of pyroxene is present, its composition provides only a minimum temperature of crystallization; that is, the temperature of the solvus below which a second pyroxene would form. For example, a single pyroxene with composition X in Figure 10.32(A) must have formed above 1000 °C; otherwise, orthopyroxene would also have been present (Problem 10.19).

Increasing pressure does not change the pyroxene phase relations much, except for decreasing, and eventually eliminating, the forbidden zone. The augite–orthopyroxene field widens at a given temperature, but the augite–pigeonite field is unaffected. Also, the three-pyroxene triangle, at a given temperature, shifts to more iron-rich compositions with increasing pressure. This has the effect of increasing the calcium content of pigeonites that coexist with augite + orthopyroxene. For example, at low pressure such a pigeonite typically contains about 11 mol% $CaSiO_3$, whereas at 1.5 GPa it contains 16 mol%.

The phase relations in Figure 10.32 are for the pure Di–En–Hd–Fs system. Natural pyroxenes, however, may contain other components, such as Al, Ti, Fe^{3+}, Cr, and Na, which can modify the phase relations. For example, the minimum stability of most natural pigeonites having Mg/(Mg+Fe) > 0.6 is approximately 50 °C higher than that shown in Figure 10.30 for pure Ca–Mg–Fe pigeonite. This results from the greater solubility of the nonquadrilateral components in augite and orthopyroxene relative to that in pigeonite (Problem 10.20). For compositions where Mg/(Mg+Fe) < 0.5 the lower stability limits of natural and pure pigeonites are identical. How each of the nonquadrilateral components affects the phase relations in the simple system is only partly understood. It is clear, however, that the temperatures in Figure 10.32 are applicable only to pyroxenes in which the quadrilateral components total more than 98%. When nonquadrilateral components exceed 2%, Figure 10.32 can still be used to determine temperatures, if compositions are correctly projected into the pyroxene quadrilateral. The scheme for doing this is given by Lindsley (1983).

10.21 TERNARY SYSTEMS WITH BINARY AND TERNARY COMPOUNDS

Many ternary phase diagrams contain ternary compounds; that is, compounds that plot within the ternary diagram. Although these may appear complex at first, they can be thought of as consisting of a number of simple ternary diagrams, the corners of which are the various compounds in the system. To illustrate such a diagram we choose the system SiO_2–Ca_2SiO_4–Mg_2SiO_4 (Hudon *et al.*, 2005), which is important because it contains most of the ferromagnesian minerals that occur in anhydrous basaltic compositions (Fig. 10.33).

This system contains 14 different minerals, two of which, diopside and akermanite, plot within the ternary system. Six minerals are polymorphs: cristobalite and tridymite of SiO_2; wollastonite and pseudowollastonite of $CaSiO_3$; and protopyroxene, orthopyroxene and pigeonite of $(Mg,Ca)SiO_3$. Compositions in this diagram are plotted in mole proportions. It will be noted that the binary compounds pyroxene and merwinite have liquidus fields that do not extend over the mineral composition. These minerals therefore melt incongruently and have boundary curves that involve reactions that end at ternary peritectics. The direction of falling temperature on the boundary curves is determined by applying Alkemade's theorem. The joins between minerals that coexist along boundary curves have been shown to indicate the location of temperature maxima. Once these have been determined, the directions of falling temperature can be plotted and peritectics and eutectics located. It is left to the reader to determine the location of these isobaric invariant points (Problem 10.21).

10.22 QUATERNARY SYSTEMS

Quaternary systems do not introduce new principles with which to analyze rocks, but they do provide an almost complete representation of the phase relations in most common rocks. With the additional compositional variable, however, graphical representation becomes more complex. Only an introductory treatment of these systems will therefore be given here.

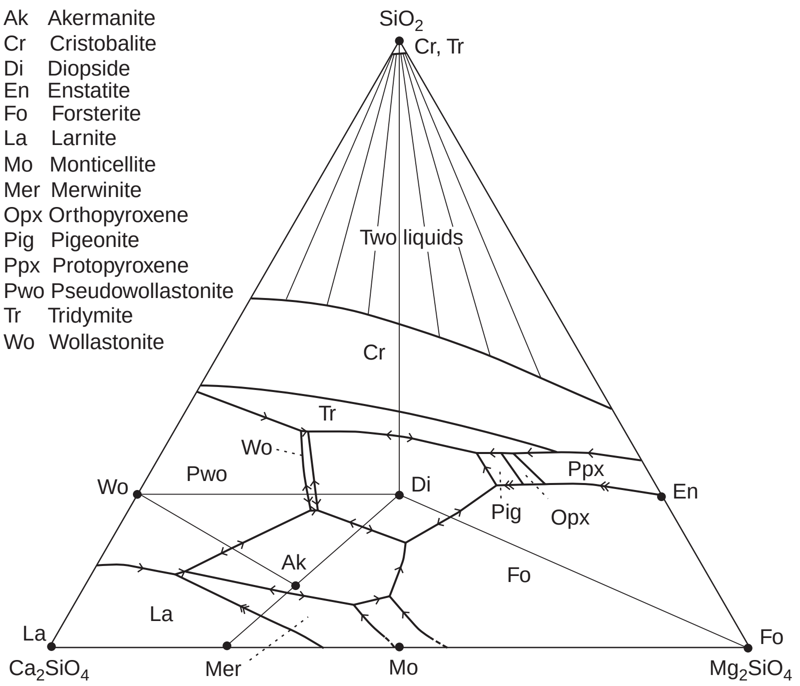


Fig. 10.33 System SiO_2–Ca_2SiO_4–Mg_2SiO_4 (from Hudon *et al.*, 2005). Compositions are plotted in mole proportions. Joins between coexisting phases have been drawn with thin lines to indicate the composition of temperature maxima on boundary curves.

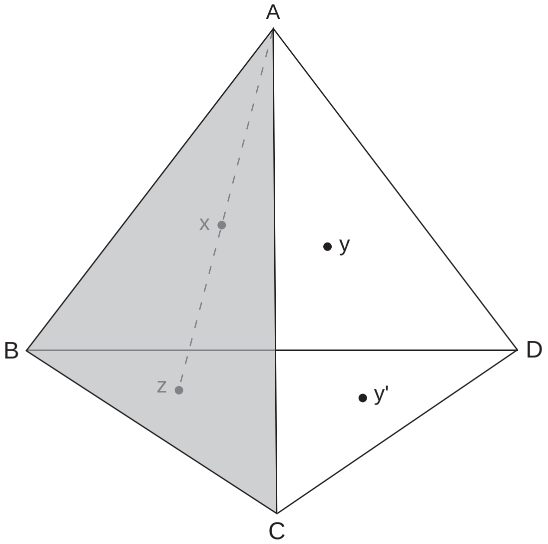


Fig. 10.34 Perspective tetrahedral representation of quaternary system. See text for discussion.

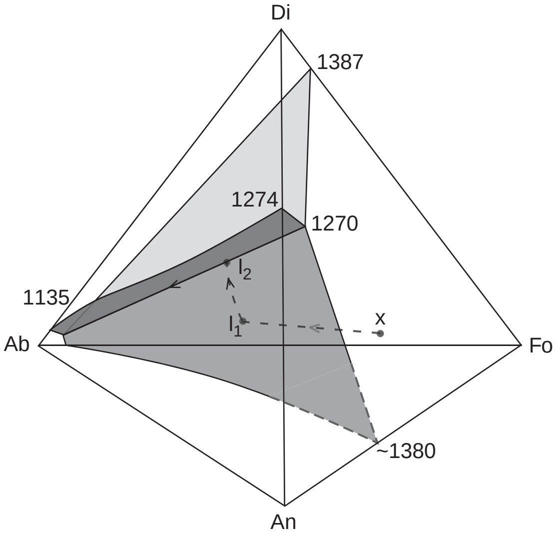


Fig. 10.35 Simplified perspective representation of system diopside–forsterite–anorthite–albite. A small volume of the nonquaternary phase spinel is omitted from the Fo–An boundary surface (dashed region). Temperatures are indicated in °C. See text for discussion. (Simplified from Yoder and Tilley, 1962.)

The four components (three independent compositional variables) require three dimensions for graphical representation. A tetrahedral plot is used for this purpose, where each apex represents one of the components (Fig. 10.34). The weight fraction of a component in any composition plotting within the tetrahedron is determined by extending a line from the component through the composition to the opposite face of the tetrahedron; the weight fraction of that component is then given by dividing the length of the line from the tetrahedral face to the composition by the total length of the line (xz/Az in Fig. 10.34). The amounts of the other components are determined with similar construction lines drawn from the other components.

Although compositions can be plotted easily in this way, compositions cannot be read from the diagram if only the point is given. Composition y in Figure 10.34, for example, could be positioned on the front face, ACD, within the tetrahedron, or on the back face, ABD. Other information, such as a construction line or projected point (y′ in Fig. 10.34), must be included to define the composition uniquely. Herein lies a limitation to graphical representations in quaternary systems.

Because three dimensions are used for the compositional variables, temperature and pressure cannot also be plotted as variables. Diagrams must therefore represent phase relations at a given pressure (isobaric) and temperature (isothermal). Although such diagrams are used in the study of metamorphic rocks (Chapter 16), it is common when dealing with igneous diagrams to plot only those phase relations involving liquids. By limiting the diagram in this way, temperature can be included as a variable, the temperatures of important points being labeled on the diagram.

Each face of the tetrahedron in a quaternary liquidus diagram is formed by an isobaric ternary diagram. The front right face in Figure 10.35, for example, is the system diopside–forsterite–anorthite. Each of the primary liquidus phase *fields* in such a ternary system extends into the quaternary system as a primary phase *volume*, which is bounded by surfaces that are extensions of the ternary boundary lines into the quaternary system. Three boundary surfaces intersect along an isobaric univariant line, and four intersect at an isobaric invariant point (eutectic or peritectic). In Figure 10.35 the ternary eutectic involving Di–Fo–An extends into the quaternary system as a cotectic along which diopside, forsterite, and plagioclase crystallize together. No quaternary invariant point is reached along this cotectic because the fourth component, albite, forms a solid solution with anorthite. The cotectic thus falls continuously in temperature from 1270 °C on the Di–Fo–An face to 1135 °C, where it passes out of the tetrahedron through the Di–Fo–Ab face. Where this line passes through the face is known as a *piercing point*.

Cooling of a magma with composition x in Figure 10.35 first results in crystallization of forsterite, which drives the magma composition directly away from forsterite until it reaches the boundary surface with plagioclase at l_1, whereupon a calcic plagioclase begins to crystallize. Crystallization of olivine and plagioclase causes the magma to change its composition along the boundary surface from l_1 to l_2. This path is curved slightly because the plagioclase continuously changes its composition as the liquid changes its composition. The degree of curvature depends on how closely crystallization approaches equilibrium. At l_2 the liquid reaches the primary field of diopside, and from there on crystallization of forsterite, plagioclase, and diopside forces the liquid to descend the cotectic. The distance that the liquid travels down the cotectic depends on the degree of fractionation.

In the system Di–Fo–An–Ab the perspective representation in Figure 10.35 is relatively clear, but in more complex systems – those with binary, ternary, or quaternary compounds, for example – phase relations can be far too complex to be portrayed clearly in this way. As a result, the phase relations are commonly presented in the form of a flowchart (Fig. 10.36). Such charts lack any compositional information, but this is not a serious loss because compositions of points in a perspective quaternary diagram cannot be read without additional information in any case. The flowchart does, however, provide a simple means of following the sequence of phases that appear during the crystallization of a magma.

Let us examine the flowchart for the system Di–Fo–An–Ab (Fig. 10.36). In the bounding ternary system Di–Fo–An there are three binary eutectics (Fig. 10.35). In the flowchart, these binary eutectics are shown as points where the liquidus lines of single phases come together (Di + Fo, for example). From each binary eutectic, a cotectic leads to a ternary eutectic represented by the triangular point Di + Fo + An. A quaternary cotectic then leaves this point. It would normally intersect two other cotectics to produce a quaternary isobaric invariant point, but this does not happen in this system because of the solid solution in the plagioclase. The quaternary cotectic then passes through the system with the plagioclase becoming progressively more albitic. Directions of falling temperature on the flowchart are indicated with arrows (double arrows for reaction curves), and the temperatures of special points are given. The piercing point is indicated by a line drawn across the cotectic line.

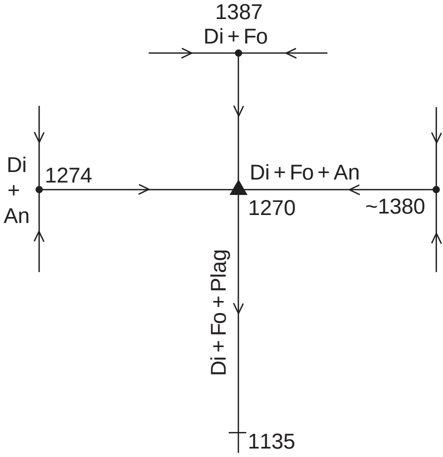


Fig. 10.36 Flowchart for the system diopside–forsterite–anorthite–albite, which is shown in perspective in Figure 10.34. Temperatures are indicated in °C. See text for discussion.

Because of the simplicity of the Di–Fo–An–Ab system, the flowchart does not provide any advantage over the perspective representation in this case. But with increasing complexity, the flowchart becomes more useful. Consider the quaternary system Di–Fo–An–SiO_2 (Fig. 10.37(A)). This system is complicated by the presence of a binary phase, enstatite, which forms a solid solution with diopside. The phase relations in this system can be seen with difficulty in the perspective representation, but this is about the limit of complexity that can reasonably be shown in perspective. The flowchart of this system (Fig. 10.37(B)) is certainly easier to read.

The flowchart in Figure 10.37(B) begins with the isobaric ternary invariant points on the faces of the tetrahedron. Two of these (Fo + An + En and Fo + Di + En) are peritectics, so that from them, two reaction curves extend into the quaternary system to meet at a quaternary peritectic involving Fo + Di + An + En. The other line that intersects at this peritectic is a cotectic coming from the Fo + Di + An face of the tetrahedron. Because of solid solution between diopside and enstatite at these temperatures, a temperature maximum exists along this cotectic represented by the dashed line in

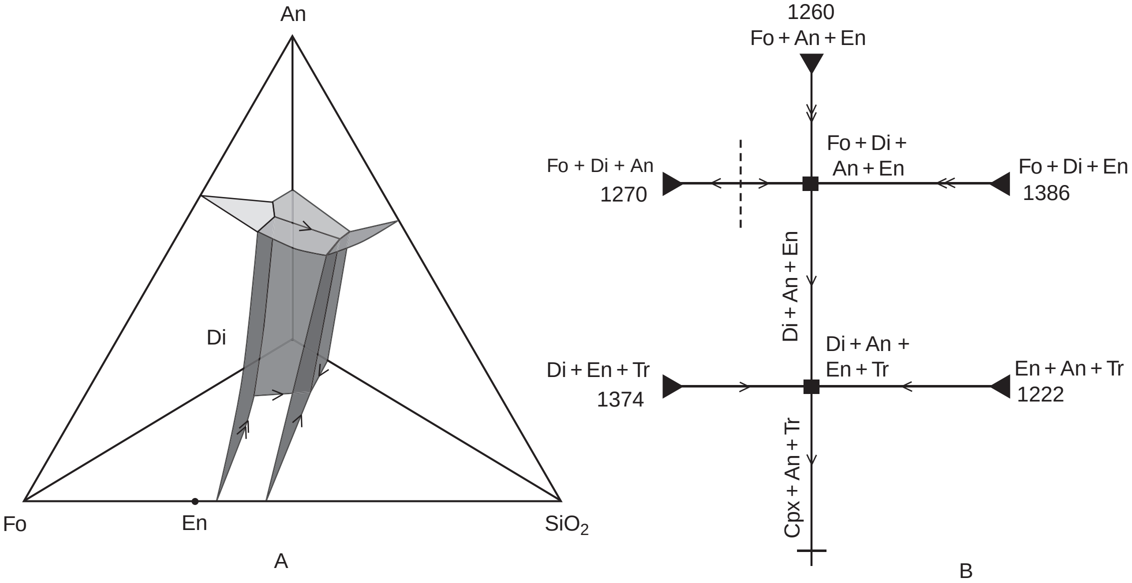


Fig. 10.37 (**A**) Schematic perspective representation of the system diopside–forsterite–anorthite–SiO_2. Small volume of the nonquaternary phase spinel is omitted from the Fo–An join. (**B**) Flowchart for same system. Temperatures are indicated in °C.

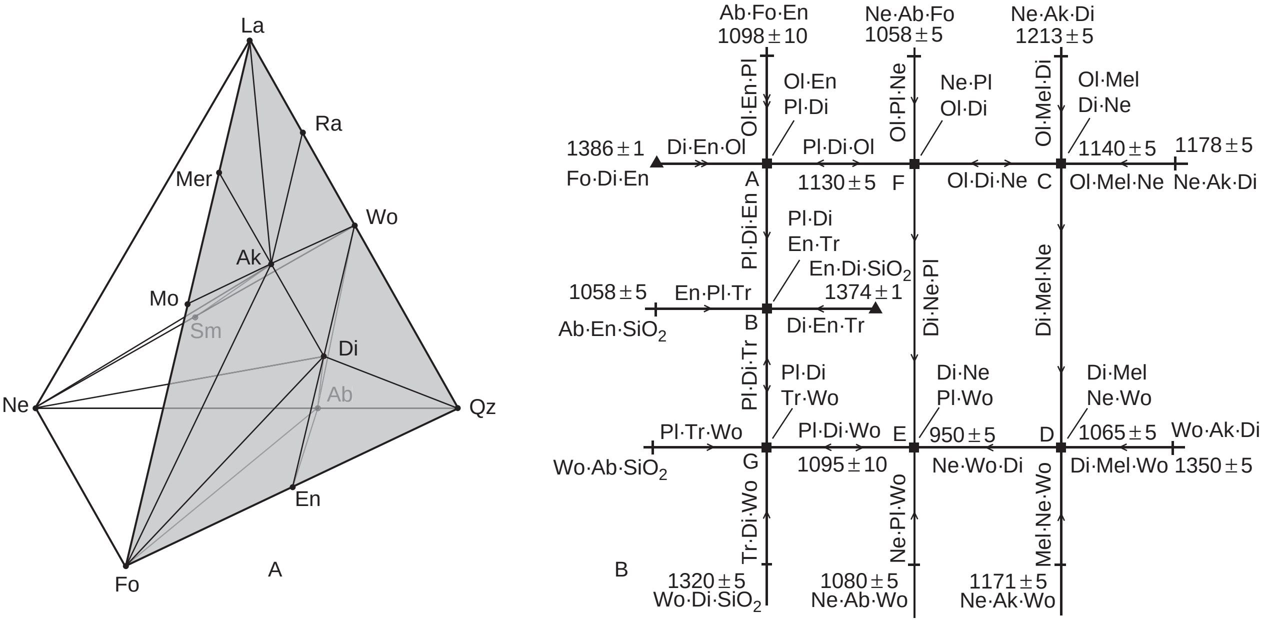


Fig. 10.38 (**A**) Perspective representation of system Ca_2SiO_4 (La)–Mg_2SiO_4 (Fo)–$NaAlSiO_4$ (Ne)–SiO_2 (Qz) with tie lines between coexisting subsolidus phases. All but albite and soda melilite plot on the shaded face. Abbreviations are as follows: larnite, La; rankinite, Ra; quartz, Q; wollastonite, Wo; merwinite, Mer; akermanite, Ak; soda melilite, Sm; monticellite, Mo; diopside, Di; albite, Ab; enstatite, En; nepheline, Ne; forsterite, Fo. (**B**) Flowchart for geologically significant part of this system. Additional abbreviations are olivine, Ol; plagioclase, Pl; melilite, Mel; tridymite, Tr. Temperatures are °C. (After Schairer and Yoder, 1964; published with permission of the Geophysical Laboratory, Carnegie Institution of Washington.)

Figure 10.37(B). The assemblage at the quaternary peritectic is isobarically invariant. Only after all of the forsterite has reacted out does the liquid start cooling toward the next invariant point, where tridymite begins to crystallize. Two other cotectic lines intersect at this point, one involving Di + En + Tr and the other En + An + Tr. Enstatite is reacted out at this second quaternary isobaric invariant point. At still lower temperatures, liquids coexisting with clinopyroxene, anorthite, and tridymite fractionate out of the tetrahedron. In this system, then, the flowchart provides a far clearer indication of the phase relations than does the perspective representation (Problem 10.22).

Finally, we examine the quaternary system Ca_2SiO_4–Mg_2SiO_4–$NaAlSiO_4$–SiO_2, which is commonly referred to as the expanded basalt tetrahedron because it contains liquids that approximate the range in composition from oversaturated to strongly undersaturated basaltic magmas. A large number of phases exist in this system. The stable subsolidus tie lines between these phases are given in Figure 10.38(A); note that all but albite and soda melilite plot on the front right face (shaded) of the tetrahedron. Clearly, if the liquidus boundaries between all phases were included in the diagram, the representation would be unintelligible. However, the flowchart for this system can be read with ease. The geologically significant part of this chart is given in Figure 10.38(B). It is left to the reader to study the diagram. Note that quaternary eutectics have four lines coming into them, whereas peritectics have three lines coming in and one line going out. After familiarizing yourself with the chart, attempt Problem 10.23.

10.23 ADIABATIC PHASE RELATIONS

Our discussion of phase diagrams so far has been limited primarily to isobaric *T–X* diagrams, with only brief mention being made of how pressure may affect the phase relations (e.g. Fig. 10.8). Isobaric *T–X* diagrams are useful to illustrate the processes that occur during the crystallization of a magma at some particular pressure. For discussion purposes, we tracked the crystallization of liquids that started cooling from the liquid field, but in nature, magmas are rarely superheated; that is, they cool from some point between the liquidus and solidus. What we did not consider was how this magma became molten in the first place. The results of a few laboratory melting experiments were shown, where samples of rock had been heated and liquid of eutectic composition formed at grain boundaries between the minerals involved in the eutectic melting (Figs. 10.4, 10.5, 10.10). This type of melting does occur in nature where a source of heat, such as the intrusion of hot basaltic magma, causes melting of lower-melting mineral assemblages, such as quartz and alkali feldspar (Fig. 3.2), but a far more important cause of melting in nature involves no heating but, instead, results from adiabatic decompression. For example, the separation of tectonic plates results in the upwelling of mantle that undergoes decompression melting to form the basaltic lavas of the ocean floor (MORB). Similarly, mantle rising beneath hot spots, such as Hawaii, also undergoes decompression melting. Decompression melting is certainly one of the most important of all igneous processes and has played a major role in differentiating the Earth.

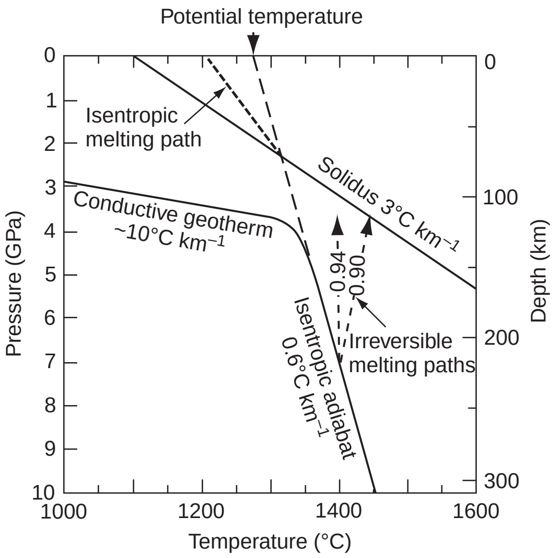


Fig. 10.39 *P–T* plot of an isentropic mantle adiabat whose projection intersects the Earth's surface at a temperature known as the potential temperature (McKenzie and Bickle, 1988). The low (0.6 °C km^{-1}) isentropic adiabatic gradient in a convecting mantle merges with the steeper (~10 °C km^{-1}) conductive geotherm in the lithosphere. The slope of the solidus for most minerals is ~3 °C km^{-1}. If the mantle continues to rise toward the surface along the adiabat, as would happen in regions of lithospheric extension, the adiabat will intersect the solidus and melting will take place. The latent heat of fusion would then cause the temperature to fall below the isentropic adiabat. If the rise of the mantle is irreversible, that is, $TdS > dQ$, and the rise of the mantle is at constant velocity, then the temperature gradient is even shallower and may even reverse. If the density ratio causing the mantle to buoyantly rise is 0.94, the temperature remains constant with depth (dashed line), and if the ratio is 0.90, the temperature actually increases (Ganguly, 2005).

In the central part of a large mass of slowly rising rock in the mantle, low temperature gradients and low thermal conductivity make it unlikely that any significant quantity of heat will escape as the mass rises; that is, the rise will approach being an adiabatic process with no heat being transferred in or out of the system ($\mathrm{d}Q = 0$). With decreasing pressure, reactions will take place, such as the conversion of minerals to lower-pressure assemblages and possibly melting. If we treat these reactions as taking place reversibly, then from Eq. (7.17), $\mathrm{d}Q = T\mathrm{d}S$, and since $\mathrm{d}Q = 0$, $\mathrm{d}S$ must be zero; that is, the rise of the mantle will involve no change in entropy. We describe this process as being *isentropic*.

In Problem 7.6, you were asked to show from the total differential of $S(T, P)$ that

$$\mathrm{d}S = \left(\frac{C_\mathrm{P}}{T}\right)\mathrm{d}T - \alpha\bar{V}\mathrm{d}P \tag{10.21}$$

where C_P is the heat capacity at constant P, α is the isobaric coefficient of thermal expansion, and $\bar{V}$ is the molar volume. Since $\mathrm{d}S = 0$ in the reversible adiabatically rising mantle, then

$$\left(\frac{C_\mathrm{P}}{T}\right)\mathrm{d}T = \alpha\bar{V}\mathrm{d}P$$

On rearranging this expression, we can show that the variation in temperature with pressure in the isentropic adiabatically rising mantle is given (Problem 7.7) by

$$\left(\frac{\partial T}{\partial P}\right)_S = \frac{T\bar{V}\alpha}{C_\mathrm{P}} \tag{10.22}$$

Substituting $\rho g\partial z$ for ∂P (Eq. (1.1)) and rearranging gives

$$\left(\frac{\partial T}{\partial z}\right)_S = \frac{gT\alpha}{C_\mathrm{P}} \tag{10.23}$$

Because all of the terms on the right side of this equation are positive, the temperature in the adiabatically rising mantle must decrease. Assuming the following reasonable values for the mantle: $\alpha = 4\times 10^{-5}\ \mathrm{K}^{-1}$, $C_\mathrm{P} = 10^3\ \mathrm{J\ kg^{-1}\ K^{-1}}$, and $T = 1500\ \mathrm{K}$, then the temperature gradient in the adiabatically rising mantle is ~0.6 °C km^{-1}. Thus, as the mantle rises adiabatically toward Earth's surface, the temperature decreases by only ~0.6 °C km^{-1} as a result of decompression (Fig. 10.39). In the mantle, where convection is possible, the temperature gradient will approach the isentropic adiabat (i.e. ~0.6 °C km^{-1}). On nearing the lithosphere, this very flat gradient merges with the much steeper (~10 °C km^{-1}) conductive geotherm in the lithosphere. This accounts for the sharp inflection in the geotherm at the base of the lithosphere.

The slope of the solidus for most minerals is ~3 °C km^{-1} (Fig. 10.39); that is, five times steeper than the isentropic adiabat. Consequently, if the mantle continues to rise toward the surface, as would happen in a zone of lithospheric extension, the adiabat would intersect the solidus and melting would begin. The latent heat of fusion would cause the temperature to fall below the adiabat as melting progressed and follow a path such as that shown by the dotted line in Figure 10.39. The isentropic adiabat in the mantle below the intersection with the solidus can be extrapolated to the surface to give a temperature that McKenzie and Bickle (1988) refer to as the *potential temperature* (1280 °C in Fig. 10.39). Mantle with a higher potential temperature has the potential for creating larger quantities of melt during decompression than mantle with a lower potential temperature.

The question now is how best to represent this isentropic decompression melting. Clearly, our by now familiar isobaric T–X diagram serves no purpose. Nor would large numbers of isobaric T–X diagrams drawn at different pressures be useful, because in deriving such diagrams the equilibrium phases were determined as a function of T, P, and composition by minimizing the Gibbs free energy according to Eq. (9.4); i.e.

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}P + \sum_i \mu_i \mathrm{d}n_i$$

However, in the case of adiabatic isentropic decompression, temperature is not an independent variable (it is a function

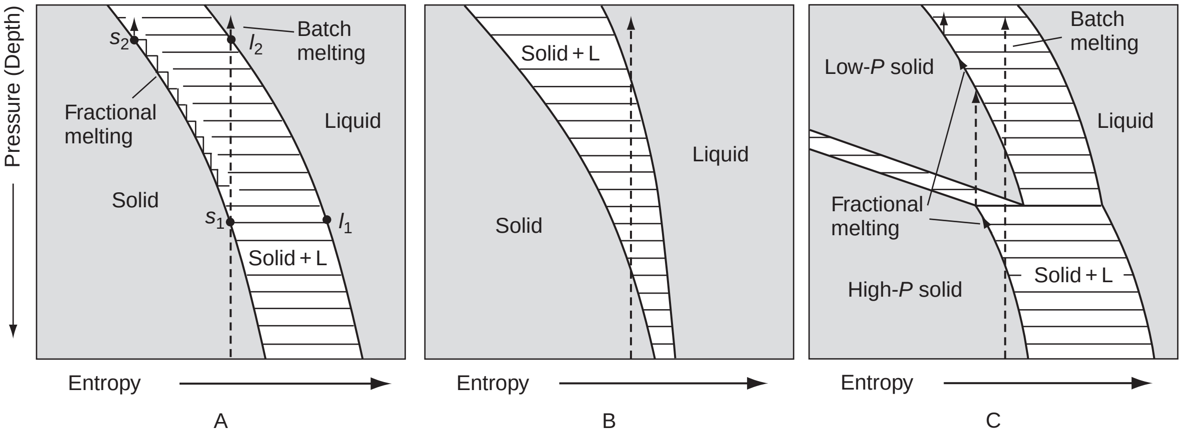


Fig. 10.40 (**A**) *P–S* plot showing a line of constant entropy (dashed line) intersecting the solidus in a one-component system, such as diopside (Stolper and Asimow, 2007) as a result of the isentropic adiabatic rise of the mantle shown in Figure 10.39. The proportion of liquid to solid formed at any given pressure in the field of Solid + L is read using the lever rule. (**B**) Similar to (A), but the solidus and liquidus are closer at high pressure. (**C**) The solidus is offset by intersection with a solid–solid phase transition. See text for discussion. (Drawn from diagrams by Ganguly, 2005, and Stolper and Asimow, 2007.)

of P). In this case, we derive the equilibrium phases as a function of entropy, pressure, and composition. For a system of fixed composition we know that by definition $H = E + PV$ (Eq. (7.8)) and therefore $dH = dE + PdV + VdP$. Also, when only work of expansion is involved in a reversible process, $dE = TdS - PdV$. Combining these two expressions, we obtain $dH = TdS + VdP$, which for a system of variable composition becomes

$$dH = TdS + VdP + \sum_i \mu_i dn_i \tag{10.24}$$

where the chemical potential is defined as $\mu_i = \left(\frac{\partial H}{\partial n_i}\right)_{S,P,n_j}$. Enthalpy then is the thermodynamic potential that must be minimized to achieve equilibrium for any given entropy, pressure and bulk composition (Ghiorso, 1997). Remember that entropy is an extensive property. It is a measure of the capacity of a system for thermal energy, which is exploited by the intensive variable temperature. Thus, a body of adiabatically rising mantle has a given amount of entropy that remains constant as the mass rises. However, to remain constant during decompression, changes amongst the phases take place to minimize the enthalpy. Decompression melting is one of these changes.

In isobaric *T–X* diagrams, we read how the phase relations change with temperature by following an isopleth for a given composition through the isobaric diagram. In the case of adiabatic (isentropic) decompression, we can follow a line of constant entropy through a *P–S* diagram for a given bulk composition (Fig. 10.40). For simplicity, let us consider only a one-component system. Figure 10.40(A) shows the *P–S* phase relations in a one-component system, such as that for diopside (Stolper and Asimow, 2007). The field for liquid diopside, which plots at high entropy, is separated from the field for solid diopside by the two-phase field (solid + liquid), which is ruled with horizontal lines (isobars). The two sides of this field need not be parallel but in the case of diopside they are. The boundaries become steeper at higher pressure (greater depth) because of the greater compressibility of the liquid. The vertical dashed line is the path of an isenthalpic adiabat. As diopside rises along the adiabat, it intersects the solidus and a melt is formed (l_1) which has a higher entropy than the coexisting solid diopside (s_1). As the rise continues, the liquid changes toward l_2 as the solid changes toward s_2. At any given pressure, the amount of liquid and solid is given by the lever rule. Although Figure 10.40 is for a one-component system (diopside), the general phase relations hold for more complex system; that is, the amount of melt increases during decompression so as to keep the entropy constant.

If F is the fraction of melt produced by the decompression, the rate of change of F with respect to P, that is $(\partial F/\partial P)_S$, is of particular importance to the generation of magma. Stolper and Asimow (2007) show that for a one-component system, if ΔS_{fus} remains constant, then the melt productivity is given by

$$-\left(\frac{\partial F}{\partial P}\right)_S = \frac{1}{\Delta S_{\text{fus}}}\left(\frac{C_P^s}{T}\left(\frac{\Delta V_{\text{fus}}}{\Delta S_{\text{fus}}}\right) - V^s\alpha^s\right) \tag{10.25}$$

where the superscript s refers to the solid. This relation shows that the melt production is an inverse function of temperature, which decreases as melting progresses. In this particular case, therefore, the melt production would be expected to increase as pressure decreases.

Some peridotite melting experiments, however, suggest that the melt productivity would decrease as the mantle rises (McKenzie and Bickle, 1988; Langmuir *et al.*, 1992). Take, for example, the phase relations shown in Figure 10.40(B), where the solidus and liquidus are closer together at high pressure than they are at low pressure. Once the adiabat crosses the solidus, only a small increase in pressure is required to produce a large increase in the proportion of liquid. However, at shallower depth, the gap between the solid and the liquid increases, so that a larger change in pressure is required to effect a similar change in the proportion of liquid.

Changes in the productivity of liquid can result from pressure-induced solid–solid phase transitions (Asimow *et al*., 1995). With decreasing pressure, minerals may convert from high- to low-pressure polymorphs (e.g. coesite to quartz), or mineral assemblages may change to adjust to the lower pressure conditions (e.g. garnet- → spinel-peridotite or spinel- → plagioclase-peridotite). Because the low-pressure assemblage has a higher entropy than the high-pressure assemblage (e.g. compare the entropies of coesite and quartz, kyanite and andalusite, and diamond and graphite in Table 7.1), the intersection of the reaction curve with the solidus in a *P–S* diagram has the form shown in Figure 10.40(C). The resulting phase relations show that such a solid–solid reaction causes a decrease in the melt production near the reaction and may, depending on the position of the adiabat, result in solidification. If we follow the rising adiabat in Figure 10.40(C), we see that it first crosses the solidus of the high-pressure solid, and melting begins. However, at slightly lower pressure, the transition to the low-pressure solid is encountered, and above this, the adiabat passes into the field of the low-pressure solid with no melt present at all. The adiabat then crosses the solidus of the low-pressure solid and melting begins again. Along such a path, the production of magma would be episodic, and since it would involve different minerals, the magma composition could be different in the different melting fields.

For the system to remain truly isentropic when the adiabat rises through the liquid + solid field, the liquid must remain in equilibrium with the solids as the amount of melt increases. Such melting is referred to as *batch melting*. However, as the mantle rises and undergoes partial melting, it is possible that melt will begin to separate from the solids (Section 3.10 and Fig. 3.12) and erupt on the surface as basaltic lava. As the liquid leaves the zone of melting, it takes with it entropy, and so the process is not isentropic. If the liquid is removed as soon as it is produced, the melting is said to be *fractional*. In nature, the generation of magma more closely approaches fractional rather than equilibrium or batch melting.

Although fractional melting is not isentropic, we can think of it as progressing in a series of small steps, each of which is isentropic. In Figure 10.40(A), if the first small amount of liquid formed on crossing the solidus is removed, the entropy will be shifted to a lower value. Another small pressure decrease will result in another small degree of isentropic melting. Again, if this liquid is removed and the process repeated, the solids will slowly progress along the solidus as the liquid is removed from the system (Fig. 10.40(A)). Melt productivity can be dramatically affected in systems undergoing fractional melting near a solid–solid reaction. In Figure 10.40(C) an adiabat was carefully chosen so that batch melting resulted in a period of solidification above the high- to low-pressure reaction. If the entropy of the system had been slightly higher, the high–low transition would have caused a decrease in the melt production at the transition, but it would not have caused complete solidification. With fractional melting, however, any system intersecting the high-pressure solidus will shift to lower entropies as melt is removed, which ensures that there will be a period where no melt is present above the transition.

The phase diagrams shown in Figure 10.40 are for a one-component system, so composition is not a variable. Multicomponent systems have too many variables to make graphical representation possible, although binary systems are tractable (see for example, Asimow *et al*., 1995).

Treating the rise of the mantle as an isentropic adiabatic process is an approximation. In reality, the process is bound to be at least partially irreversible, in which case entropy will increase. Analysis of such behavior is complex. Ganguly (2005), however, has shown that if a mantle diapir rises frictionlessly at constant velocity, irreversible adiabatic decompression results in a temperature–depth trajectory that is shallower than the isentropic adiabat, and that when the density ratio between the diapir and the surrounding mantle drops below 0.94, the temperature increases as the diapir rises (Fig. 10.39). He also found that under these conditions melt productivity is enhanced.

10.24 COMPUTER-GENERATED PHASE RELATIONS

When we progressed from ternary to quaternary phase diagrams, it became apparent that graphical representation of phase relations was becoming very difficult and, of course, impossible for more complex systems. In addition, when other restrictions were placed on the system, such as making them isenthalpic, graphical representation became even more difficult. However, our analysis of phase equilibria should not be limited by the number of dimensions we can represent on a piece of paper. The same principles of equilibrium apply whether we are dealing with a simple binary system or a complex magma of ten or more components. If we abandon the graphical approach, the phase relations can be calculated for any given composition and set of conditions. When the number of components is large, these calculations are complex and require the use of computer programs. While a number of programs have been created to do this, one of these has met with more success than others. This is the MELTS program (Ghiorso and Sack, 1995; Asimow and Ghiorso, 1998; Ghiorso *et al*., 2002).

We quote here from the MELTS home page at http://melts.ofm-research.org/: “MELTS is a software package designed to facilitate thermodynamic modeling of phase equilibria in magmatic systems. It provides the ability to compute equilibrium phase relations for igneous systems over the temperature range 500–2000 °C and the pressure range 0–2 GPa.” A later revision of the program, pMELTS, was calibrated specifically for higher pressures and is particularly useful for considering the generation of magmas in the upper mantle. Again, quoting from the MELTS home page: “pMELTS is a revised calibration of the liquid thermodynamic model in MELTS and is optimized for mantle bulk compositions. pMELTS is calibrated from an extended experimental database that includes more extensive data at elevated pressure. pMELTS is optimized for “mantle-like” bulk compositions at near solidus conditions (0–30% melting) and is intended to be applied over the temperature range 1000–2500 °C and the pressure interval 1–3 GPa." The software can be downloaded from the MELTS home page.

The MELTS program was created by calibrating the free energy of a liquid against careful experimental results. The liquid's free energy was expressed as follows:

$$\bar{G} = \sum_{i=1}^{n} X_i \mu_i^o + RT \sum_{i=1}^{n} X_i \ln X_i + 1/2 \sum_{i=1}^{n} \sum_{j=1}^{n} W_{i,j} X_i X_j$$
$$+RT[X_{\mathrm{H_2O}} \ln X_{\mathrm{H_2O}} + (1 - X_{\mathrm{H_2O}}) \ln(1 - X_{\mathrm{H_2O}})] \qquad (10.26)$$

The first term on the right-hand side of the equation gives the contribution each component makes to the solution, independent of mixing (Eq. (9.5)). The next two groups of terms account for the free energy of mixing in a regular solution model, where $W_{\mathrm{i,j}}$ are the interchange energies (Eq. (9.42)). The last group of terms gives the ideal free energy of mixing of water in the liquid (Eq. (9.22)). The free energies of the end-member solids are taken from the compilation of Berman (1988). Internally consistent solid-solution models were created for feldspars, olivine, orthopyroxene, Ca-bearing pyroxenes, rhombohedral oxides, and spinels.

The MELTS program allows one to model the equilibrium or fractional crystallization of a specified bulk composition. This can be done under specified conditions of, for example, constrained *T* and oxygen fugacity, constant entropy (adiabatic), or constant volume (as might occur in a melt inclusion in a mineral). The program works well for basaltic compositions (see Problems 10.24 and 10.25). Although MELTS accounts for the effect of water in the melt, hornblende and biotite are not modeled well by the program, and consequently it is not recommended for intermediate and silicic magmas where these minerals may play an important role.

10.25 PROBLEMS

10.1 **(a)** Using the ideal cryoscopic equation, calculate the liquidus lines for diopside and anorthite in the binary system diopside–anorthite at atmospheric pressure. The melting points for diopside and anorthite are 1665 and 1830 K, respectively, and their heats of fusion are 142.6 and 136 kJ mol^{-1}, respectively.

(b) Recalculate mole fractions into weight percent (mol. wt., Di = 0.2166 kg, An = 0.2782 kg), and plot a binary phase diagram for the system diopside–anorthite (*T* °C versus wt%). What are the temperature and composition of the eutectic?

10.2 **(a)** Compare the calculated phase diagram for the system Di–An from Problem 10.1 with the experimentally determined one in Figure 10.4. What might explain the differences?

(b) Calculate the activity coefficients (γ_i) for diopside and anorthite at the experimentally determined eutectic by comparing calculated and actual temperatures for this composition. What conclusion can you draw about the ideality of the solutions?

10.3 **(a)** On a line between two components *A* and *B*, plot a point representing a magma containing 50 wt% *A* and 50 wt% *B*. If on cooling slightly, 10 wt% of this magma forms crystals of pure *A*, what is the composition of the remaining liquid (90% of total)? Solve graphically.

(b) Graphically determine the bulk composition of a mixture of 20% crystals of *B* and 80% of a liquid with a composition 30% *A* and 70% *B*.

10.4 **(a)** List the sequence of appearance and disappearance of phases involved in the equilibrium cooling of a liquid indicated by the isopleth in Figure 10.10. What is the composition of the rock formed from this magma?

(b) If some of the same magma in part (a) were erupted and rapidly quenched when it is 50% solidified, what composition would the glass (supercooled liquid) have in terms of Ks–SiO$_2$, and what would the phenocrysts be?

(c) Some of the magma erupted in part (b) remains at depth and has all of the crystals removed from it when it is 50% solidified. The remaining liquid then crystallizes under equilibrium conditions. List the sequence of crystallization of this separated liquid, and give the composition and name of the rock that it finally forms.

(d) Discuss the range of rock types that can form from the magma in part (a) by the different crystallization paths in parts (a) and (c).

10.5 Using Eqs. (10.18) and (10.20), calculate the liquidus and solidus in the system $\mathrm{Mg_2SiO_4}$–$\mathrm{Fe_2SiO_4}$ at atmospheric pressure assuming ideal mixing in both the solid and liquid. The melting points of fayalite and forsterite are 1490 and 2163 K, respectively, and their heats of fusion are 92.174 and 122.256 kJ mol^{-1}, respectively. Convert mole fractions to wt% (mol. wt. Fa = 0.20379 kg, Fo = 0.14069 kg), and compare the calculated phase diagram [*T* (°C) versus wt%] with the experimentally determined one in Figure 10.11 (plot both in the same diagram). What can you conclude about the ideality of both solid and liquid olivine solutions?

10.6 Thermochemical data from which heats of fusion can be determined are not available for all minerals; forsterite is one such mineral. However, if the effect of pressure on the melting point and the volume change on melting are known, the enthalpy of fusion can be determined from the Clausius–Clapeyron equation (Eq. (8.3)). The melting point of forsterite at atmospheric pressure (10^5 Pa) is 2163 K, and the volume change on melting is 2.703×10^{-6} m^3 mol^{-1}. If the rise in melting point with pressure is 47.7 K GPa^{-1}, calculate the heat of fusion of forsterite at atmospheric pressure. (Recall that 1 J = 1 Pa m^3.)

10.7 In Figure 10.12, if a magma containing 60% anorthite cools and crystallizes under equilibrium conditions until only one-third of the liquid remains, what is the composition of this residual liquid?

10.8 Discuss the textures resulting from disequilibrium crystallization of magma of composition x in Figure 10.14(D). Assume that some fraction of every composition that crystallizes is preserved. Pay particular attention to where and when exsolution would occur.

10.9 List the sequence of events in the equilibrium crystallization of a liquid containing 90% Ab in Figure 10.16(B).

10.10 Draw sketches contrasting textural differences to be expected in thin sections of feldspars that have crystallized under disequilibrium conditions with the phase relations as shown in Figure 10.17.

10.11 **(a)** On a sheet of triangular graph paper, plot a magma that contains 25% *A*, 50% *B*, and 25% *C*. This magma assimilates an equal weight of xenoliths that contain 80% crystals of *A* and 20% crystals of *C*. What is the resulting bulk composition?

(b) The heat necessary to cause assimilation normally comes from the heat of crystallization of a mineral on the liquidus. If the assimilation in part (a) brings about 20% crystallization of *A* from the contaminated bulk composition, what is the composition of the final liquid?

10.12 **(a)** Describe the equilibrium crystallization of a liquid containing 40% Ne, 40% Ab, and 20% sodium silicate in the system in Figure 10.21(A).

(b) What composition does the liquid have following 50% crystallization?

(c) What solids are present following 50% crystallization, and what are their relative abundances?

(d) What is the instantaneous composition of the solids crystallizing from this liquid following 50% crystallization?

(e) What composition do the solids have when this liquid first reaches the eutectic?

10.13 Consider the equilibrium crystallization of a liquid halfway between liquids *A* and *B* in Figure 10.23(C).

(a) Where will crystallization end, and what will be the composition of the rock formed?

(b) When 50% of the liquid in part (a) has crystallized, what composition do the solids have?

(c) If all of the crystals in part (b) are suddenly removed, trace the equilibrium crystallization of the remaining liquid.

(d) Compare the rock types formed in parts (a) and (c).

10.14 List the sequence of steps involved in the equilibrium crystallization of a magma having a composition of 20% Fa, 20% Lc, and 60% SiO_2 in Figure 10.24.

10.15 **(a)** What composition plagioclase will first crystallize from a melt containing 10% Di, 45% Ab, and 45% An in the system Di–Ab–An (Fig. 10.25(C))?

(b) Assuming equilibrium crystallization of this liquid, what composition will the plagioclase have when the first diopside crystals form?

(c) Assuming equilibrium crystallization, what is the composition of the final liquid to crystallize from this same starting composition?

10.16 In Figure 10.26(B), trace the steps involved in the equilibrium crystallization of a liquid containing 15% Q, 35% Ab, and 50% Or. First determine what the composition of the final feldspar must be.

10.17 List the steps involved in the equilibrium crystallization of a liquid containing 25% Di and 75% En in the pseudobinary diagram of Figure 10.29(B). Note the directions in which exsolution lamellae develop and draw a sketch of the resulting texture.

10.18 Would an andesitic lava with a temperature of 1100 °C and crystallizing pyroxene with Mg/(Mg + Fe) of 0.6 be likely to have crystals of pigeonite or orthopyroxene?

10.19 A diabase dike that is known to have crystallized near the surface of the Earth contains coexisting orthopyroxene ($Ca_4Mg_{67}Fe_{29}$) and augite ($Ca_{35}Mg_{48}Fe_{17}$). Using the Lindsley pyroxene geothermometer, estimate the temperature at which these pyroxenes crystallized. Would you expect the rock also to contain pigeonite, especially on the rims of the orthopyroxene; if so, what composition would you expect it to have?

10.20 Explain why the greater solubility of nonquadrilateral components in augite and orthopyroxene relative to that in pigeonite increases the stability field of augite + orthopyroxene (see Section 10.10).

10.21 By careful inspection of the direction of falling temperature on the boundary curves in Figure 10.33, label all eutectics and peritectics in this system. Using the definition of subalkaline and alkaline rocks in Chapter 6, draw the boundary separating these two groups of rocks in this phase diagram.

10.22 Many basaltic rocks contain plagioclase phenocrysts. In terms of Figure 10.37, this means that they plot on the plagioclase (An) saturation surface. For such rocks it is possible to simplify this quaternary system by projecting the phase boundaries that intersect the anorthite surface from the An apex onto the triangular base Di–Fo–SiO_2. The result is an isobaric ternary diagram indicating the liquidus phase relations in anorthite-saturated liquids. Using the information given in Figure 10.37, draw this ternary projection and label all boundary lines with appropriate arrows indicating directions of falling temperature and cotectic and reaction relations.

10.23 In Figure 10.38(B), a temperature maximum is indicated on the olivine–diopside–nepheline boundary line (between F and C). This temperature maximum occurs where the boundary line pierces the Fo–Di–Ne plane (locate this plane in Fig. 10.38(A)). List the sequence of steps involved in the fractional crystallization of two magmas that both lie on this boundary line but on opposite sides of the temperature maximum. Plot the approximate paths that these fractionating liquids might follow in the perspective representation of Figure 10.38(A). Note that a peritectic must plot outside the tetrahedron formed by the phases that coexist at the peritectic. Moreover, a peritectic must be away from the face of the tetrahedron opposite the phase that is consumed at the peritectic. A eutectic, on the other hand, must plot within the tetrahedron formed by the phases involved. What textural difference would the plagioclase exhibit in the rocks formed from these two liquids?

10.24 In this problem we will use the MELTS program to trace the crystallization of the same liquid as in Problem 10.13 through the system forsterite–anorthite–SiO_2.

(a) A liquid with a composition halfway between points A and B in Figure 10.23(C) will have the following composition:

SiO_2	55.47
Al_2O_3	5.86
MgO	35.44
CaO	3.20
FeO	0.03

Although this system contains no iron, a very small amount of iron has been added to make the program run properly (it does not affect the phase relations). Enter this analysis into the MELTS program, and from the *Intensive Variables* menu select **T, P** and set *Start* temperature to 1700 °C and *Stop* temperature to 1150 °C and temperature increment to 2 ° C; set Start and Stop pressure to 1 bar. From the *Commands* menu select **Find Liquidus**. Next, from the *Commands* menu select **Execute** and allow the program to run until crystallization is complete. What minerals are crystallizing when the last liquid disappears? What point is this in the phase diagram?

(b) In the MELTS folder you will find a file labeled melts.out. Open this as a Word document and set the font to 9 point Times Roman and the output data will format properly. This file contains all of the data from your crystallization run, giving the phases that are present at each temperature step, their amounts and compositions, and physical properties such as density and viscosity. From these data extract the following information for the given temperatures, and create a spreadsheet showing temperature, phases present and their proportions, and the composition of the liquid for the following temperatures: liquidus, first appearance of orthopyroxene, 1452 °C, 1400 °C, 1300 °C, and first appearance of anorthite. Using a computer program, calculate the CIPW norms (see Section 6.3 and the web page http://volcanoes.usgs.gov/staff/jlowenstern/other/software_jbl.html) of the liquids at the selected temperatures and enter these into your spreadsheet. Next, plot these liquids in a ternary diagram of forsterite–anorthite–SiO_2. To do this you will have to recast normative hypersthene as a mixture of forsterite and SiO_2 using the formula weights given in Table 7.1 ($Mg_2Si_2O_6 = Mg_2SiO_4 + SiO_2$).

(c) From the melts.out file determine the composition of the liquid in part (b) when the initial composition is 50% solidified. Decrease the alumina in this analysis by 0.06 and add 0.06 FeO (prevents MELTS from hanging up). Enter this analysis into the MELTS program and by following the same procedure as in part (a) allow this liquid to cool from 1500 to 950 °C. Using the data in the melts.out file complete a similar spreadsheet as in part (b) for the following temperatures: liquidus, first appearance of anorthite, 1150 °C, 1100 °C, 1050 °C, and first appearance of tridymite. With the data in the spreadsheet, complete as much of the forsterite–anorthite–SiO_2 phase diagram as possible. Label the peritectic and eutectic and indicate directions of falling temperature.

(d) Compare your calculated phase diagram with the experimentally determined diagram shown in Figure 10.23(C).

10.25 In Section 4.2, vesicle cylinders in basalt flows were described as forming from bubbly lava whose composition indicated between 30% to 40% crystallization of the host basalt. Goff (1996) gives the following analyses of a host basalt and vesicle cylinder from Petroglyph Canyon, California:

	Basalt	Cylinder
SiO_2	49.50	50.00
TiO_2	1.3	1.9
Al_2O_3	16.9	15.5
Fe_2O_3	2.5	6.9
FeO	6.2	3.9
MnO	0.12	0.16
MgO	7.4	5.4
CaO	10.2	10.2
Na_2O	3.5	3.8
K_2O	0.67	0.82
P_2O_5	0.3	0.38

Enter the analysis of the basalt into the MELTS program and select start and stop temperatures of 1250 and 1050 °C and a pressure of 1 bar. Determine the liquidus temperature and run the cooling of this composition in 2 °C increments. As crystallization progresses, elements such as Ti and P, which do not enter any of the early crystallizing minerals, become enriched in the residual liquid. From inspection of the melts.out file, determine what degree of crystallization is necessary for the TiO_2 content of the residual liquid to rise to 1.9%. How closely do the other elements in the calculated residual liquid compare with their concentrations in the vesicle cylinder?

11 Effects of volatiles on melt equilibria

11.1 INTRODUCTION: COMPOSITION OF VOLCANIC GASES

All magmatic components under specified conditions of temperature, total pressure, and magma composition have a definite vapor pressure, but for most components it is so small, even at the low confining pressures on the Earth's surface, that these components form only condensed phases; that is, solids and liquids. A few components, however, can have high vapor pressures, and if their concentrations are high enough, or the confining pressure low enough, they may form a separate gaseous phase. These are the components referred to as the volatile constituents of magmas. Even the normally condensed components can become volatile if pressures are extremely low (approaching a vacuum). Such pressures existed in the early solar nebula during accretion of the planets, and under these conditions differences in the vapor pressures of the common rock-forming minerals played an important role in determining what minerals could condense to form the planets.

The volatile constituents of magmas, although normally present in only small amounts, can profoundly affect liquidus temperatures, phases that crystallize, and magmatic fractionation trends. They play an important role in the generation of magmas, and as such are probably responsible for the shallow depth of the asthenosphere (Liu *et al*., 2006) and the Earth's unique style of tectonism (Ulmer, 2001). Heat and volatiles released from basaltic magma intruded at the base of the crust cause melting of the crustal rocks, with formation of hydrous intermediate and silicic magmas (Annen *et al*., 2006). The dissolved volatiles lower the density and viscosity of the magma, which allows it to rise more easily. On nearing the Earth's surface, the pressure decrease causes the volatiles to exsolve, which increases the viscosity and causes the magma to stall and crystallize. The release of volatiles can cause volcanic eruptions, which may have devastating consequences, or the volatiles may be released more slowly to form hydrothermal systems, which may form economic mineral deposits. The solubility of volatiles in magmas is clearly an extremely important topic.

The true significance of volatile constituents to magmatic processes has become apparent only through experimental studies on silicate systems and natural rocks. Magmas, on cooling and crystallizing, lose most of their volatiles. The final rock may therefore preserve only a poor record of the volatiles that were originally present. Hydrous minerals, fluid inclusions, and vesicles in volcanic and hypabyssal rocks do provide some evidence. In laboratory studies, however, it is possible to control the composition of the volatiles and to evaluate their effect on phase relations. The results, in turn, can be used to interpret the role volatiles may have played in the origin of a rock.

That magmas contain volatiles, and in some cases, large quantities, is obvious from volcanic explosions such as that of Mount St. Helens in 1980. Sampling of the gaseous emanations of active volcanoes has provided a wealth of information on the common volatile constituents associated with near-surface igneous activity. Samples have been obtained from such different environments as strato volcanoes over subduction zones (Mount St. Helens), shield volcanoes over hot spots in oceanic plates (Kilauea), and submarine fumeroles on mid-ocean ridges. Not all gases liberated from volcanoes, however, have their origin in the magma. Indeed, a very large fraction can be recycled groundwater or ocean water. The steep temperature gradients associated with high-level igneous intrusions are ideal for setting up large hydrothermal convection cells. The circulating, hot solutions can drastically alter the composition of the igneous rocks through which they pass. Moreover, many elements, in particular the chalcophile ones, may be redistributed and concentrated by these solutions to form important mineral deposits.

Gaseous volcanic emanations consist largely of the elements hydrogen, oxygen, carbon, and sulfur, which form H_2O, CO_2, CO, SO_2, H_2S, H_2, S, and O_2. Other minor constituents include N_2, Ar, HCl, HF, and B (BVSP, 1981). Although the abundance of these gases varies considerably from one volcano to another and even in the same volcano at different times, H_2O and CO_2 are always by far the two most abundant gases, constituting from 30 to 80 and 10 to 40 mol%, respectively, of the vapor phase (Anderson, 1975). There is considerable uncertainty, however, as to how much of each of these gases is juvenile (coming to the Earth's surface for the first time) and how much is meteoric (has, at some time, resided in the hydrosphere or atmosphere) and derived from recycled groundwater or contact metamorphism of intruded rocks (CO_2 from limestone, for example). Hydrothermal convection cells are certainly capable of recycling large volumes of water (Davis and Elderfield, 2004). For example, the entire volume of water in the oceans is estimated to pass through the hydrothermal systems along oceanic ridges once every 8 million years, and Stein and Stein (1994) estimate that 34% of the total global oceanic heat flux occurs by hydrothermal circulation. Over geologic time, however, it is generally agreed that the hydrosphere and atmosphere have been formed and maintained largely by juvenile volcanic gases. For some of the gases (oxygen and helium, for example), isotopes provide a simple means of identifying juvenile gases (Section 13.5), and direct evidence of juvenile gases is found in fluid inclusions in mantle-derived nodules.

Although the abundance of the various volatiles in volcanic gases is determined in part by elemental abundances, concentrations of individual gas species are also controlled by reactions between the species. For example, carbon dioxide and carbon monoxide are related by the reaction

$$CO_2 \rightleftharpoons CO + {}^1\!/_2\, O_2 \tag{11.1}$$

for which an equilibrium constant (Eq. (9.55)) can be written as

$$K = \frac{(f\text{CO})(f\text{O}_2)^{1/2}}{(f\text{CO}_2)} = e^{-\Delta G/RT} \tag{11.2}$$

At constant pressure, the equilibrium constant is a function of temperature (Eq. (9.60)). In addition, the gases behave almost ideally at the temperatures and pressures at which they are released from volcanoes. We can therefore rewrite Eq. (11.2) as

$$\frac{p\text{CO}}{p\text{CO}_2} = p\text{O}_2^{-1/2} e^{-\Delta G/RT} \tag{11.3}$$

where partial pressures are measured in bars. If we assume for simplicity that these gases are the only ones present, the sum of their partial pressures must equal the total pressure on the gas (1 bar); that is,

$$p_{\text{CO}_2} + p_{\text{CO}} + p_{\text{O}_2} = 1\ \text{bar} \tag{11.4}$$

The partial pressure of oxygen, however, is extremely small (about 10^{-10} bar) and can therefore be ignored without introducing serious error. Also, for an ideal gas,

$$p_i = X_i P_{\text{total}} \tag{11.5}$$

We can therefore rewrite Eq. (11.3) as

$$\frac{X_{\text{CO}}}{X_{\text{CO}_2}} = p\text{O}_2^{-1/2} \cdot e^{-\Delta G/RT} = \frac{V_{\text{CO}}}{V_{\text{CO}_2}} \tag{11.6}$$

Equation (11.6) shows that the volume proportions of CO and CO_2 in volcanic emanations are a function of the oxygen fugacity and temperature. As will be shown in Section 11.5, the ferromagnesian minerals that crystallize from magmas are capable of maintaining oxygen fugacities at fixed or near fixed values (buffering) at a given temperature. These minerals will therefore determine the ratio of CO and CO_2 in the magmatic gases. For example, at 1150 °C (1423 K) the ΔG for reaction (11.1) is 161 kJ mol^{-1}, and if the oxygen fugacity is buffered at $10^{-11.2}$ bar by the ferromagnesian minerals, the volume ratio of CO to CO_2 would be 0.5; that is, twice as much CO_2 as CO would be present in the volcanic gas.

In passing, it is worth noting that reactions such as Eq. (11.1) are actually used to control oxygen fugacities in experimental laboratory studies. Carefully metered volumes of CO and CO_2 are mixed and allowed to flow into a furnace. Once the gases reach equilibrium (a few seconds), an oxygen fugacity is established that is determined by the volumes of CO and CO_2 and the temperature of the furnace.

Many other equilibria exist in magmatic gases in addition to the one given in Eq. (11.1) (Heald *et al*., 1963). Some of these are

$$H_2O \rightleftharpoons H_2 + {}^1\!/_2 O_2 \tag{11.7}$$

$$CO_2 + H_2 \rightleftharpoons CO + H_2O \tag{11.8}$$

$$CH_4 \rightleftharpoons C + 2H_2 \tag{11.9}$$

$$H_2S + O_2 \rightleftharpoons SO_2 + H_2 \tag{11.10}$$

Clearly, the concentrations of all these volatile species are interrelated and are sensitive to the oxygen fugacity, which in most cases is buffered by the ferromagnesian minerals in the rock. In addition, because equilibrium constants change with temperature, the concentrations of the gaseous species are sensitive to temperature. Pressure can also be of importance, as can be illustrated by considering the interaction of Eqs. (11.7) and (11.10). Expressions for the equilibrium constants for these two reactions can be combined to give the following equation:

$$\frac{(f\text{SO}_2)}{(f\text{H}_2\text{S})} = \frac{K_{11-10}(f\text{O}_2)^{3/2}}{K_{11-7}(f\text{H}_2\text{O})} \tag{11.11}$$

At a specific temperature the equilibrium constants are fixed, and fO_2 is fixed by the buffering of ferromagnesian minerals. Thus, as the fugacity of water increases with depth, the fugacity of H_2S will increase relative to that of SO_2. The major sulfur-bearing gaseous species at depth may well be H_2S despite the prevalence of SO_2 in surface samples of volcanic gases. A detailed analysis of the composition of a Kilauea gas sample by Nordlie (1971) provides an excellent example of the difficulties involved in determining the composition of volcanic gases.

In this chapter, we focus on the effect of volatiles, in particular water, carbon dioxide, sulfur, and oxygen, on the phase relations between solids and liquids. We will consider first the solubility of these volatiles in common magmas, and then study how their presence can alter the phase relations in some of the important systems considered in Chapter 10. Finally, a consideration of fractional crystallization and melting in volatile-bearing systems will reveal the crucial importance of volatiles to the genesis of many igneous rocks.

11.2 SOLUBILITY OF H_2O IN SILICATE MELTS

Chapter 10 dealt with phase diagrams in which components have such low vapor pressures that even at atmospheric pressure only condensed phases form. In such systems, the lowering of the liquidus of a phase, resulting from the addition of another component, is limited only by the liquid

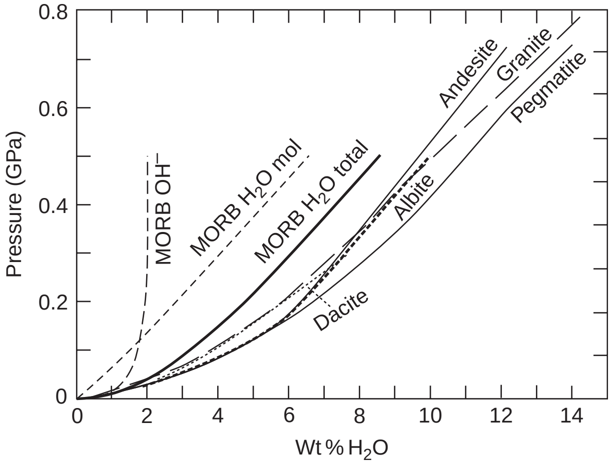


Fig. 11.1 Solubility of water as a function of pressure in mid-ocean ridge basalt (MORB), andesite, albite (heavy dotted line), synthetic granite ($Qz_{28}Ab_{38}Or_{34}$), Mount St. Helens dacite (fine dotted line), and granite pegmatite at temperatures ranging from 1200 °C for the MORB to 900 °C for pegmatite (from data in Dixon *et al*., 1995; Holtz *et al*., 1995; Silver and Stolper, 1989; Moore *et al*., 1995; Burnham, 1979). As shown for the MORB, the total water in solution can be divided into that which is dissolved as OH^- ions and that which is dissolved as molecular H_2O; the OH^- reaches a maximum of about 2 wt% (Dixon *et al*., 1995). Similar distributions of dissolved species are found in other silicate melts.

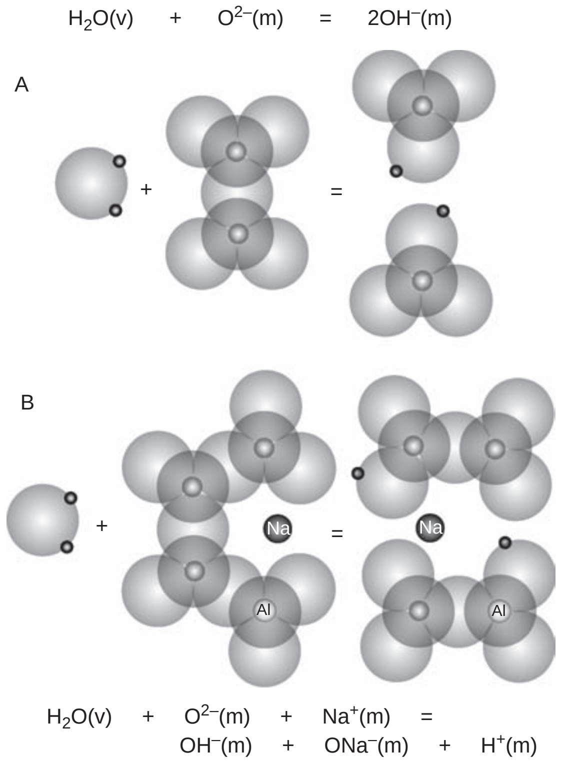


Fig. 11.2 (**A**) Solution of water in a silicate melt by hydrolysis of a bridging oxygen of the SiO_4^{4-} tetrahedral framework. Large atoms are oxygen; smaller atoms in the center of the tetrahedra are silicon and the very small atoms attached to oxygen are hydrogen. (**B**) Solution of water in the albite melt structure involving dissociation of water and movement of the Na^+ ion to maintain charge balance. (After Burnham, 1979.)

becoming saturated in a new crystalline phase. In volatile-bearing systems, however, addition of a component may lead to the formation of a separate volatile-rich phase. Once this happens, further addition of that component to the system does not lower liquidus temperatures but simply adds to the amount of separate volatile phase. In other words, the liquid becomes saturated in that volatile component. Because the amount of lowering of liquidus temperatures is critical to the formation of magmas in a planet that does not have excess heat, it is important to know what determines the saturation level of a volatile component in a silicate melt.

The simplest way of tackling this problem is to think of the formation of a separate volatile-rich phase in terms of a reaction. For example, if the volatile component is H_2O, the reaction would be written as

$$\text{hydrous silicate melt} = \text{anhydrous silicate melt} + H_2\text{O(vapor)} \qquad (11.12)$$

Water in a separate vapor phase has a greater volume than when it is dissolved in the melt. Consequently, the volume change for this reaction is positive, being large at low confining pressures but becoming smaller at higher pressures, because of the greater compressibility of the separate vapor phase. Because the left-hand side of this reaction has the smaller volume, increased pressure favors the formation of hydrous melt. In other words, the solubility of water in a melt is a function of pressure. This conclusion is not surprising in view of our everyday experience with carbonated beverages, where releasing the pressure causes carbon dioxide to start coming out of solution. Water in a separate vapor phase has greater entropy than when dissolved in the melt. Consequently, the entropy change for this reaction is positive. Increasing temperature will therefore favor the right-hand side of this reaction; that is, the solubility of water in magma decreases with increasing temperature. This is true of other volatile constituents in magmas (see Problem 11.3).

The solubility of water in melts has been studied in many synthetic systems and in the common magma types (Fig. 11.1). The saturation limit-versus-pressure relations for all these melts are similar. The solubility at atmospheric pressure is negligible, but with increasing pressure it rises rapidly at first and then at a more moderate, almost linear, rate at pressures above about 0.2 GPa. Mafic melts generally can dissolve less water than felsic ones. For example, at 0.5 GPa and 1100 °C, molten basalt can dissolve up to ~8 wt% H_2O, but molten granitic pegmatite can dissolve up to 11 wt%.

The solubility of water in silicate melts is clearly a function of pressure and to a lesser extent composition and temperature. To understand more fully the solubility of water in silicate melts, it is necessary to analyze how water actually dissolves in the melt. Does it, for example, remain as molecular water and simply fit into cavities in the structure of the melt, or does it react with the melt? The answer to this question has come from analyzing the infrared (IR) spectra of glasses formed by rapid quenching of silicate liquids produced in experiments under controlled conditions. Absorption peaks in the IR spectra show that water enters silicate liquids both as molecular water and by reacting with oxygen to form hydroxyl groups (Silver and Stolper, 1989; Dixon *et al*., 1995). Figure 11.2(A) illustrates one way in which water can enter the melt by a hydrolysis reaction with a bridging oxygen. The bridging oxygen remains attached to one of the tetrahedral groups and by gaining a hydrogen ion is converted to a hydroxyl group. The remaining hydroxyl ion completes the tetrahedral coordination around the other tetrahedron. This form of solution can be represented by the reaction

$$\mathrm{H_2O, molecular\,(m)} + \mathrm{O^{2-}(m)} \rightleftharpoons \mathrm{2OH^-(m)} \qquad (11.13)$$

where (m) represents the melt. Because the linkage of the framework structure is broken (depolymerized) by this solution, it dramatically reduces viscosities (Section 2.4). Of course, other similar reactions are possible that involve nonbridging oxygens.

The solubility of water in silicate melts has been investigated in detail by using molten albite as a simple model for more complex silicate melts (e.g. Goranson, 1938; Burnham and Davis, 1974; Silver and Stolper, 1989). As seen in Figure 11.1, the solubility of water in molten albite (heavy dotted line) is similar to that in other more complex magmas. The infrared spectra of albite glasses show that, at low water concentrations, water dissolves in molten albite primarily as hydroxyl groups, but at higher concentrations it enters in molecular form (Silver and Stolper, 1989). The hydroxyl water, although rising rapidly at first, soon reaches a constant value of ~2 wt%. The molecular water, however, increases almost linearly with pressure and when the total water reaches ~4 wt%, the molecular water constitutes half of the water in solution. This same relation is found in other simple silicate melts and more complex ones, such as the mid-ocean ridge basalt (MORB) shown in Figure 11.1 (Dixon *et al*., 1995).

According to Eq. (11.13), at equilibrium a silicate melt will contain a mixture of molecular water, hydroxyl groups, and oxygen ions. If we assume ideal mixing, their proportions can be expressed in terms of the equilibrium constant for Eq. (11.13), that is,

$$K = \frac{\left(a^{\mathrm{m}}_{\mathrm{OH}}\right)^2}{\left(a^{\mathrm{m}}_{\mathrm{H_2O,mol}}\right)\left(a^{\mathrm{m}}_{\mathrm{O}}\right)} = \frac{\left(X^{\mathrm{m}}_{\mathrm{OH}}\right)^2}{\left(X^{\mathrm{m}}_{\mathrm{H_2O,mol}}\right)\left(X^{\mathrm{m}}_{\mathrm{O}}\right)} \qquad (11.14)$$

Silver and Stolper (1989) show that this simple relation fits the data well until the total water content exceeds 2 wt% and then it diverges progressively because of the leveling off of the concentration of the hydroxyl groups (see Problem 11.1). They consequently modify this equation, taking into account complicating factors. For example, in Eq. (11.14), all anhydrous oxygens are treated as being identical, but clearly many different types of oxygens will exist in a magma (nonbridging, bridging, different polymer groups, etc.) and separate reactions similar to Eq. (11.13) could be written for each type of oxygen. In addition, different types of polymeric groups could result from the hydrolysis. The assumption that mixing is ideal implies that every possible configuration of molecular water, hydroxyl group, and oxygen atoms is energetically identical, that is, equally possible (see Eqs. (9.16)–(9.20)). This is most unlikely, and a better fit to the data can be achieved by using a regular solution model (Section 9.6) where interactions between various species can be assigned individual energies (*W* in Eq. (9.42)).

The resulting equations are complex. Fortunately, a software program, *VolatileCalc*, is available, which is based on the Silver and Stolper melt model (Newman and Lowenstern, 2002). This program can be downloaded from the USGS web site at http://volcanoes.usgs.gov/staff/jlowenstern/other/software_jbl.html. The program, which runs in Microsoft Excel, presents solution models for both H_2O and CO_2 in basalt and rhyolite compositions. It allows one to calculate: (1) the saturation pressures for given H_2O and CO_2 concentrations up to pressures of 0.5 GPa; (2) concentrations of H_2O and CO_2 in melt and vapor during open- or closed-system degassing; (3) isobaric-, isoplethic- (constant composition) and polybaric-solubility curves for basalt and rhyolite; (4) and the fugacities of H_2O and CO_2 vapors at magmatic temperatures (see Problem 11.1).

Figure 11.1 shows that despite the similarity in the solubility of water in a wide range of silicate liquids, there is clearly a compositional effect, with the more silicic melts dissolving more water at any given pressure. Even in the simple granite system, quartz–albite–orthoclase, water solubility is greatest in compositions rich in albite (Holtz *et al*., 1995). From the measurements of water solubility in a wide range of silicate liquids, Moore *et al*. (1998) have created the following empirical expression to account for the solubility in terms of pressure, temperature and composition:

$$2 \ln X^{\mathrm{melt}}_{\mathrm{H_2O}} = \frac{a}{T} + \sum_i b_i X_i \left(\frac{P}{T}\right) + c \ln f^{\mathrm{fluid}}_{\mathrm{H_2O}} + d \qquad (11.15)$$

where the coefficients are as follows: $a = 2565$, $b_{\mathrm{Al_2O_3}} = -1.997$, $b_{\mathrm{FeOt}} = -0.9275$, $b_{\mathrm{Na_2O}} = 2.736$, $c = 1.171$, $d = -14.21$. The fugacity of water is calculated from a *modified Redlich–Kwong* equation of state, which is discussed in Chapter 20 (see Eqs. (20.5) and (20.6)). These fugacities can also be obtained from the *VolatileCalc* program. Although Eq. (11.15) is an empirical fit to a large experimental database, the individual terms have thermodynamic significance. The first term (*a*) is related to the enthalpy of the reaction of water going from the vapor to the melt state. The *b* term relates to the molar volume of H_2O in the melt, which shows a compositional dependence (hence the different values for b_i). The *c* term modifies the linear dependence of X_{H2O} on the fugacity of water. The *d* term is related to the entropy change resulting from water transferring from the vapor to the liquid state and hence is strongly negative (i.e. entropy decreases). Although this expression does not distinguish between water entering the melt in the molecular or hydroxyl forms, it fits the data for total water over a wide range of magma compositions up to pressures of 0.3 GPa (see Problem 11.2).

11.3 SOLUBILITY OF CO_2 IN SILICATE MELTS

Carbon dioxide, the other major volatile constituent in volcanic gases, behaves quite differently from H O in silicate

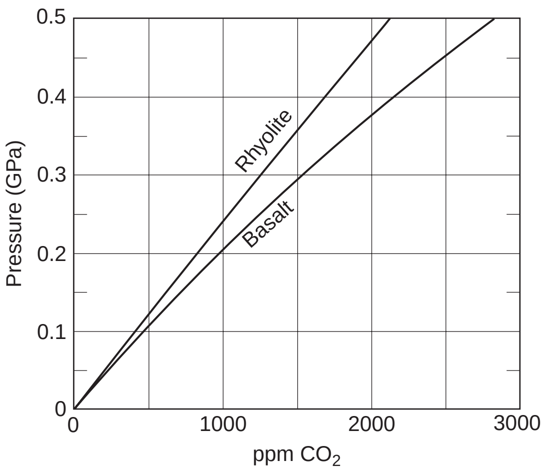


Fig. 11.3 Solubility of CO_2 in basaltic and rhyolitic magma at 1200 °C according to the solution model in the VolatileCalc program (Newman and Lowenstern, 2002).

melts. Infrared spectroscopy shows that CO_2 dissolves in silicate melts by reacting with O^{2-} to form carbonate ions (CO_3^{2-}). This carbonation reaction, which is equivalent to the hydrolysis reaction for water, can be written as

$$CO_2(v) + O^{2-}(m) \rightleftharpoons CO_3^{2-}(m) \tag{11.16}$$

where v and m indicate vapour and melt phases respectively. If the O^{2-} in the melt is a bridging oxygen, as for example in molten albite, the carbonate ion that is formed places a small, highly charged C^{4+} ion near a small, highly charged Si^{4+} or Al^{3+} ion of the tetrahedron. The resulting repulsion means that this reaction is not likely to proceed to the right, at least not at low pressures. Therefore, in framework-dominated melts, the amount of solution of CO_2 by this mechanism is small. Nor is it likely that holes in the structure of these melts will accommodate molecular CO_2, which is confirmed by the IR spectroscopy. According to Bohlen *et al*. (1982), at pressures greater than 1.5 GPa, molten albite does dissolve some CO_2. Addition of a small amount of H_2O to a framework melt dramatically increases the solubility of CO_2 by depolymerizing the melt and opening its structure so that it can accept molecular CO_2 (Eggler, 1973).

In melts containing nonbridging oxygens (i.e. oxygens not joining SiO_4 or AlO_4 tetrahedra), solution of CO_2 by the carbonation reaction (Eq. (11.18)) is feasible. In this case the highly charged C^{4+} ion is not repelled as much by the larger, less-charged ion (Mg^{2+}, Fe^{2+}, Ca^{2+}, for example). Consequently, as the percentage of nonbridging oxygens increases in the melt, so does the solubility of CO_2 (Fig. 11.3). Thus, molten diopside, which contains 50% bridging oxygens, can dissolve more CO_2 than can molten albite, which contains 100% bridging oxygens, and molten forsterite, which contains no bridging oxygens, can dissolve still more (Eggler, 1973, 1974). The solubility of CO_2 is therefore strongly dependent on the silica content of magmas. Dixon (1997) has shown that the solubility of CO in basaltic magmas increases with the degree of silica undersaturation. For example, at 1200 °C and 0.1 GPa, a tholeiitic basalt from Kilauea (SiO_2 = 49.2) can dissolve 543 ppm CO_2, whereas a basanite (SiO_2 = 46.0) can dissolve 807 ppm, and a leucitite (SiO_2 = 44.1) can dissolve 2200 ppm. The solubility of CO_2, even in the more mafic melts, is, however, far less than the solubility of H_2O. For example the MORB shown in Figure 11.1, at 0.5GPa, can dissolve 8.55 wt% H_2O but only 0.0028 wt% CO_2.

When the melt contains both H_2O and CO_2, which normal magmas do, the activities of the volatiles affect each other, and solution in the melt becomes increasingly nonideal. The *VolatileCalc* program takes these interactions into account for the two main magma types, basalt and rhyolite. A more complex model for the solubility of H_2O and CO_2 in a wide compositional range of silicate melt compositions has been devised by Papale *et al*. (2006). They show that although the maximum solubility of CO_2 would be expected for a fluid that consists entirely of CO_2, nonideality causes the CO_2-solubility in some compositions to have a maximum when some H_2O is present.

11.4 SOLUBILITY OF SULFUR IN SILICATE MELTS

Sulfur can be present in the magmatic vapor phase in a number of different species: S_2, SO_2, and H_2S. The relative abundances of these species are determined by the fugacities of S_2, O_2, and H_2, which in turn determine how sulfur dissolves in the melt (Wallace and Carmichael, 1992). If H_2S is abundant, sulfur could dissolve by a hydrolysis reaction with bridging oxygens:

$$H_2S(v) + O^{2-}(m) \rightleftharpoons SH^-(m) + OH^-(m) \tag{11.17}$$

If the magma is dry and the oxygen fugacity low, solution of sulfur will take place with the formation of sulfide in the melt:

$$\tfrac{1}{2}S_2(v) + O^{2-}(m) \rightleftharpoons S^{2-}(m) + \tfrac{1}{2}O_2(v) \tag{11.18}$$

and at high oxygen fugacities it will dissolve as sulfate:

$$\tfrac{1}{2}S_2(v) + O^{2-}(m) + \tfrac{3}{2}O_2(v) \rightleftharpoons SO_4^{2-}(m) \tag{11.19}$$

Solution by the mechanism of Eq. (11.17) seems to be minor, because the solubility of sulfur in most melts decreases with increase in number of bridging oxygens. Solution by the mechanism of Eq. (11.18) seems most likely in magmas under the normal range of oxygen fugacities. At oxygen fugacities above those of the nickel–nickel oxide buffer ($fO_2 > 10^{-5}$ bar; see Section 11.8), however, Eq. (11.19) will describe the solution.

Fincham and Richardson (1954) showed that the solubility of sulfur in silicate melts can be related to the composition of the vapor phase simply by expressing the sulfur content of the melt in terms of what they referred to as the sulfide and sulfate capacities of the magma, which are defined respectively as

$$C_S = \text{wt\% S(m)} \times \left(\frac{fO_2}{fS_2}\right)^{1/2} \tag{11.20}$$

and

$$C_{SO_4} = \text{wt\% S(m)} \times \left(fS_2 \cdot fO_2^3\right)^{-1/2} \tag{11.21}$$

The sulfide capacity is applied when $fO_2 < 10^{-6}$ bar and the sulfate capacity when $fO_2 > 10^{-5}$ bar. At a given temperature and pressure the value C_s or C_{SO_4} is a constant. The logarithm of this constant varies inversely with absolute temperature. Its variation with pressure is unknown.

The explanation for Fincham and Richardson's finding is readily apparent from examining the equilibrium constants for Eqs. (11.18) and (11.19). Consider first the reaction at low oxygen fugacity ($<10^{-6}$). The equilibrium constant is given by

$$K = \frac{(aS^{2-})(fO_2)^{1/2}}{(aO^{2-})(fS_2)^{1/2}} \tag{11.22}$$

At fixed temperature and pressure, K is a constant. In addition, if the composition of the melt is constant other than for sulfur, the activity of O^{2-} will not vary much. Consequently, $(aS^{2-}) \times (fO_2/fS_2)^{1/2}$ must be essentially constant. This then accounts for the functional form of the "sulfide capacity" (C_S). In addition, from Eq. (9.56) we know that $\ln K = -\Delta G/RT$, thus accounting for the variation of C_s with temperature. The equilibrium constant for Eq. (11.19) is

$$K = \frac{(aSO_4^{2-})}{(aO^{2-})(fS_2)^{1/2}(fO_2)^{3/2}} \tag{11.23}$$

Using the same arguments as before, $(aSO_4^{2-}) \times (fS_2 \cdot fO_2^3)^{-1/2}$ must be a constant at a given temperature and composition of the melt. This then accounts for the functional relation of C_{SO_4} in more oxidized melts.

Doyle (1987) has shown that Fincham and Richardson's way of expressing the sulfur contents of melts works so well that the proposed solution mechanisms are probably accurate. It also explains why in magmas with low oxygen fugacities (Eq. (11.20)), increasing oxygen fugacity decreases the sulfur content of the melt. By contrast, at high oxygen fugacities (Eq. (11.21)), increasing oxygen fugacity increases the sulfur content of the magma. Tholeiitic basalts typically have low oxygen fugacities and thus contain sulfur in solution as sulfide. Many andesites, however, are more oxidized and have sulfur dissolved as sulfate. Because of the difference in the effect of fO_2 on C_S and C_{SO_4}, we can expect that magmas intermediate between basalt and andesite, having fO_2 near that of the nickel–nickel oxide buffer, will have a minimum sulfur-bearing capacity.

Fincham and Richardson (1954) showed that not only is the solubility of sulfur in a silicate melt dependent on the composition of the vapor phase but that the composition of the melt plays a major role, especially its ferrous iron content. Indeed, the solubility of sulfur in silicate melts is so strongly correlated with the FeO content that Eq. (11.18) could be rewritten as

$$\tfrac{1}{2}S_2(v) + FeO(m) \rightleftharpoons FeS(m) + \tfrac{1}{2}O_2(v) \tag{11.24}$$

The equilibrium constant for this reaction can be written as follows:

$$K = \frac{(aFeS)(fO_2)^{1/2}}{(aFeO)(fS_2)^{1/2}} \tag{11.25}$$

from which we see that for a given temperature and sulfur fugacity, increasing the oxygen fugacity decreases the activity of FeS in the melt and consequently the amount of sulfur held in solution. This effect is made more pronounced by the fact that increasing the oxygen fugacity also decreases the activity of FeO in the melt (because of conversion to Fe_2O_3). Lowering the activity of FeO in the melt, for example by crystallizing Fe-bearing oxides or silicates, also decreases the capacity of the melt to carry sulfur.

In addition to being a volatile constituent in silicate melts, sulfur has the ability to form immiscible sulfide liquids (MacLean, 1969; Haughton *et al*., 1974; Naldrett, 1989). These liquids consist largely of iron and sulfur with small amounts of oxygen (they crystallize to pyrrhotite and minor magnetite), but if copper or nickel are present in the magma, these elements partition strongly into the sulfide melt. Large magmatic ore deposits associated with mafic and ultramafic rocks, such as those of the Sudbury district of Ontario (Naldrett, 1969, 1989), are thought to have formed by the segregation of such sulfide liquids.

The separation of an immiscible sulfide liquid from a magma can be illustrated with the aid of the pseudoternary system FeS–FeO–SiO_2 (Fig. 11.4). Not all phases in this system plot within the ternary diagram. Pyrrhotite, for example, contains more sulfur than FeS (troilite), and the oxygen content of wüstite depends on the oxygen fugacity and temperature. For purposes of illustration, however, no serious errors are introduced by considering the system a ternary one that coexists with a vapor phase consisting of sulfur and oxygen. For a complete discussion of this system, see MacLean (1969). Figure 11.4 shows the vapor-saturated phase relations in this system for liquids in equilibrium with solid iron; that is, in addition to the phases shown in Figure 11.4, a vapor phase and solid iron are always present. The composition of the vapor phase varies considerably throughout the system, depending on the phases present. The total pressure is near atmospheric.

The phase diagram is dominated by three two-liquid fields and one three-liquid field. The three liquids involved are a sulfide-rich one with a composition near FeS, a silica-rich one, and an iron-rich silicate one. The two-liquid field involving sulfide liquid and iron-rich silicate liquid is the geologically important part of this diagram. The arcuate boundary of this field reflects the increasing solubility of sulfur in silicate melts as a function of their FeO content. At the highest silica content, this liquid can contain only ~1 wt% FeS in solution, but as the FeO content increases, the solubility of FeS increases.

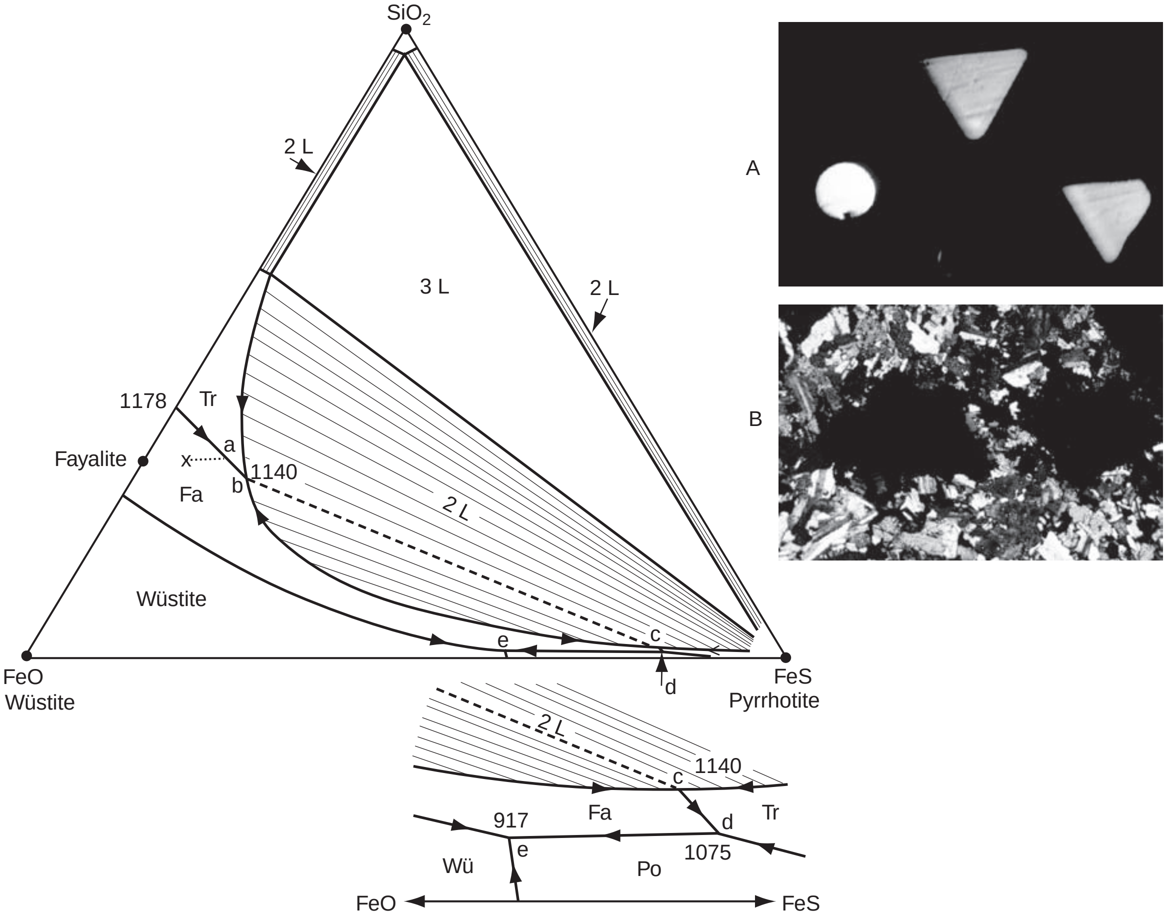


Fig. 11.4 Liquidus phase relations in the vapor-saturated pseudoternary system FeS–FeO–SiO$_2$. Inset shows details of the low-temperature sulfur-rich part of this system. Temperatures in °C. (After MacLean, 1969.) (**A**) Reflected light photomicrograph of a small immiscible droplet of sulfide melt (now pyrrhotite) and two octahedra of magnetite in the glassy residue of the Holyoke tholeiitic basalt, Connecticut (width of field 60 μm). (**B**) Large immiscible sulfide globules (opaque) in a gabbro forming the nickel ore in the Sudbury Complex, Ontario (width of field 3 cm).

For example, when the FeO/SiO$_2$ ratio in the melt is the same as that in fayalite, the melt can dissolve up to almost 20 wt% FeS. In still more FeO-rich melts, FeS becomes completely soluble and the immiscibility field disappears.

To illustrate how an immiscible sulfide liquid might form from a magma that initially was undersaturated in sulfide, consider the equilibrium crystallization of liquid x in Figure 11.4. This liquid first crystallizes fayalite (+iron), which changes the liquid composition to point a on the cotectic where tridymite begins to crystallize. The liquid then descends the cotectic to the isobaric invariant point b at 1140 °C. At b, droplets of immiscible sulfide liquid with composition c nucleate and grow. Further crystallization of fayalite and tridymite decreases the amount of liquid b and increases the amount of liquid c, but the compositions and temperatures of these liquids remain unchanged. Eventually, when all of liquid b has disappeared, the sulfide liquid at c is free to cool again (see enlargement of phase relations in lower part of Fig. 11.4). This it does by following the fayalite–tridymite cotectic from c down to the isobaric invariant point d at 1075 °C. Point d is a peritectic where tridymite reacts with the liquid to form fayalite and pyrrhotite. Because the initial magma composition x lies in the triangle SiO$_2$–Fa–FeS, all liquid is consumed at this point. If the immiscible sulfide liquid had separated from the silicate fraction during differentiation, the liquid at d would have continued cooling to the ternary eutectic e at 917 °C, where fayalite, pyrrhotite, and wüstite (+iron) would have crystallized.

Although the phase relations in Figure 11.4 are for rather reducing conditions (in equilibrium with iron), they are not very different from those under more oxidizing conditions. The two-liquid field has essentially the same shape, but the tridymite–fayalite cotectic is replaced by a tridymite–magnetite one, and magnetite, tridymite, and pyrrhotite crystallize at the ternary eutectic. Figure 11.4 does not show any of the other important chemical fractionations that occur during the differentiation of a magma, such as the variation in the Mg/Fe ratio. The abundance of these other constituents, however, affects the solubility of sulfur in magma and hence the size of the immiscibility field (Doyle, 1987). Most basaltic magmas reach this immiscibility field at some stage in their crystallization (inset (A) in Fig. 11.4). If unmixing occurs early in the crystallization history of a magma, the dense sulfide liquid may be able to separate from the silicates and segregate to form an ore body (inset (B) in Fig. 11.4). This early-separating liquid would take with it most of the nickel present in the magma. If unmixing occurs late, the sulfide liquid has greater difficulty separating from the silicates. In addition, because nickel enters early-crystallizing ferromagnesian minerals, especially olivine, late-separating sulfide liquids may contain no nickel and therefore be of little economic value.

11.5 EFFECT OF H_2O ON MELTING IN SILICATE SYSTEMS

Having examined the solubility of common volatile constituents in magmas, we will now evaluate their effect on liquidus phase relations, beginning with water. The system albite–H_2O has been the most carefully studied volatile-bearing silicate system (Burnham and Davis, 1974; Burnham, 1979; Silver and Stolper, 1989), and the effects of water in this system are generally applicable to other silicate melts. Qualitatively, the effect of dissolved water on the liquidus of albite is easy to predict. From the cryoscopic equation (Eq. (10.7)), we know that any lowering of the activity of albite in the melt lowers the liquidus temperature. Thus, the more water that is dissolved in the melt, the lower will be the liquidus of albite. However, the solubility of water in the melt is strongly dependent on pressure. We would expect, therefore, that if water is available to dissolve in the melt, the melting point of albite should decrease with increasing pressure, which it does.

To quantify this relation, we must know the solubility of water in albite melt as a function of temperature and pressure. If the albite melt is saturated in water and the vapor phase is pure H_2O, the activity of H_2O in the melt must be unity. The vapor phase is not actually pure H_2O, but it is so nearly so that no serious error is introduced by assuming that $a^{m}_{H_2O} = 1.0$. The darker shaded, curved surface in Figure 11.5(A) shows the water saturation surface for albite melt as a function of temperature and pressure (Burnham, 1979; Silver and Stolper, 1989). This surface is contoured with lines of constant mole fraction of water in the melt. Hydrous albite melts can exist only on the high-pressure side of this surface. Should a melt cross this surface by having the pressure drop, water would exsolve from the melt and form gas bubbles (vesicles).

For albite to be in equilibrium with a melt, it is necessary for the chemical potential of albite in both phases to be the same; that is,

$$\mu^{m}_{Ab} = \mu^{s}_{Ab} \tag{11.26}$$

where m and s indicate melt and solid respectively. Using Eq. (9.34), we can rewrite Eq. (11.26) as

$$\mu^{m*}_{Ab} + RT\ln a^{m}_{Ab} = \mu^{s*}_{Ab} + RT\ln a^{s}_{Ab} \tag{11.27}$$

which on rearranging gives

$$0 = \Delta G^{*}_{mAb} + RT\ln\frac{a^{m}_{Ab}}{a^{s}_{Ab}} \tag{11.28}$$

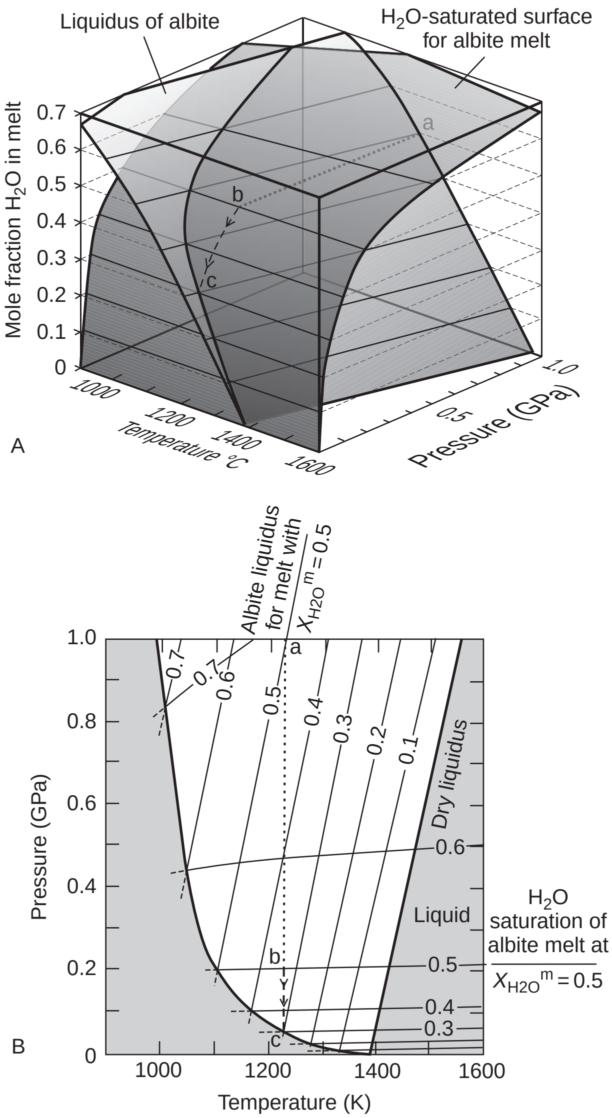


Fig. 11.5 (**A**) Temperature–pressure– $X^{m}_{H_2O}$ plot of the water saturation limit of molten albite (dark shaded surface). This surface is contoured with lines of constant mole fraction of water in the melt. The light shaded surface is the liquidus of albite, which is also contoured with lines of constant mole fraction of water in the melt. The line of intersection of these two surfaces is the water-saturated solidus of albite. This line and contours of constant mole fraction of water on the saturation surface and liquidus surface are projected onto the *P*–*T* plane in (**B**). Other points in figure are discussed in the text. (Drawn from data in Burnham, 1979, and Silver and Stolper, 1989.)

where ΔG^{*}_{mAb} is the free-energy change of melting of pure albite at the given *T* and *P*. But according to Eq. (7.34), $\Delta G^{*}_{mAb} = \Delta G^{\circ}_{mAb} + P\Delta V_{mAb}$, where ΔG°_{mAb} is the free-energy change of melting of pure albite at temperature *T* and 1 bar, and ΔV_{mAb} is the volume change on melting of albite. After dividing through by *RT*, Eq. (11.28) can be rewritten as

$$0 = \frac{\Delta G^{\circ}_{mAb} + P\Delta V_{mAb}}{RT} + \ln\frac{a^{m}_{Ab}}{a^{s}_{Ab}} \tag{11.29}$$

The first group of terms in Eq. (11.29) can be evaluated from data in thermodynamic tables (see Burnham, 1979). The activity of albite in the solid is unity as long as we are dealing only with the system albite–H_2O. The activity of albite in the hydrous melt clearly depends on the amount of water dissolved in the melt. From the Gibbs–Duhem equation (Eq. (9.7)) for this two-component system,

$$X_{Ab}^{m}d\mu_{Ab}^{m}+X_{H_2O}^{m}d\mu_{H_2O}^{m}=0 \tag{11.30}$$

For μ_i we can substitute $\mu_i^*+RT\ln a_i$ (Eq. (9.34)), giving

$$X_{Ab}^{m}d\ln a_{Ab}^{m}+X_{H_2O}^{m}d\ln a_{H_2O}^{m}=0 \tag{11.31}$$

At low concentrations of water ($X_{H_2O}^{m}<0.5$), the dissociation of water results in the $a_{H_2O}^{m}\propto(X_{H_2O}^{m})^2$ (Burnham, 1979). If we take logarithms of both sides of this expression and differentiate, we obtain $d\ln a_{H_2O}^{m}=2d\ln X_{H_2O}^{m}$, which can be substituted into Eq. (11.32) to give

$$X_{Ab}^{m}d\ln a_{Ab}^{m}+2X_{H_2O}^{m}d\ln a_{H_2O}^{m}=0 \tag{11.32}$$

But $X_{H_2O}^{m}=1-X_{Ab}^{m}$, which on substituting into Eq. (11.32) and simplifying gives

$$\frac{da_{Ab}^{m}}{a_{Ab}^{m}}=\frac{2dX_{Ab}^{m}}{X_{Ab}^{m}} \tag{11.33}$$

On integrating, this gives

$$a_{Ab}^{m}=\left(X_{Ab}^{m}\right)^2=\left(1-X_{H_2O}^{m}\right)^2 \tag{11.34}$$

This simple expression, then, relates the activity of albite in the melt to the mole fraction of water in the melt as long as $X_{H_2O}^{m}\leqslant 0.5$. A similar but more complex expression can be derived for melts with higher concentrations of water (see Burnham, 1979).

We are now in a position to calculate the albite liquidus for any given pair of values of temperature, pressure, or mole fraction of water in the melt. For example, if we choose a pressure of 0.2 GPa and a temperature of 1200 K, the value of $\left(\Delta G_{mAb}^{\circ}+P\Delta V_{mAb}\right)/RT$ is 0.91. According to Eq. (11.29), then, the activity of albite in the melt is 0.4, which Eq. (11.34) indicates corresponds to a mole fraction of albite in the melt of 0.63. The mole fraction of water in the melt under these conditions would then be 0.37. It is possible, therefore, to calculate the liquidus surface for albite and contour it with lines of constant mole fraction of water in the melt, as shown in Figure 11.5(A) by the light-shaded curved surface.

At the base of Figure 11.5(A) is the liquidus of albite where the mole fraction of water in the melt is zero. This is commonly referred to as the "dry liquidus." It rises from a temperature of 1391 K at atmospheric pressure to just below 1600 K at 1.0 GPa. As the mole fraction of water increases in the melt, this liquidus shifts to progressively lower temperatures. But to do this, the melt must be able to dissolve water, and we have already seen that this is strongly dependent on pressure and that melts can exist only on the high-pressure side of the water saturation surface in Figure 11.5(A) (darker shaded surface). As a result, only that part of the albite liquidus lying on the high-pressure side of the water saturation surface is stable. The line of intersection of the albite liquidus surface and the water saturation surface is known as the *water-saturated solidus* of albite. Only solid albite plus vapor is stable below this.

The phase relations shown in perspective in Figure 11.5 (A) are commonly projected onto the *P–T* plane, as shown in Figure 11.5(B). To read information from such a projection, it is necessary to have a clear understanding of the complete T–P–$X_{H_2O}^{m}$ diagram. To help illustrate this, consider the path followed by a hydrous magma as it rises to the surface. Be certain to locate the path in both the perspective and projection diagrams.

Let us assume that an albite melt is in equilibrium with a solid albite source at the base of the crust where the pressure is 1 GPa. Enough water is present to produce a mole fraction of water in the melt of 0.5 (point a in Fig. 11.5). The temperature of this melt must be 1230 K. We will assume that the magma rises rapidly toward the surface of the Earth so that its temperature remains essentially constant (we will ignore adiabatic cooling) and its mole fraction of water remains constant at 0.5. It will therefore follow the dotted line a–b. As soon as the magma starts to rise, it leaves the liquidus surface, and any solid albite suspended in it would dissolve. As the magma rises, it approaches the water saturation surface, which for a melt with a mole fraction of water of 0.5 occurs at a pressure of 0.2 GPa (point b in Fig. 11.5). At this point, a separate vapor phase forms bubbles in the magma. Note that there is still no crystallization of albite. Further rise of the magma and lowering of pressure requires that the melt descend the water saturation surface in Figure 11.5 (dashed arrowed line, b–c). This progressively lowers the mole fraction of water in the melt and increases the volume of vapor phase. During this stage, the initial vapor bubbles grow into large pockets of gas, which are likely to float toward the upper part of the magma chamber, where they can form pegmatites or hydrothermal solutions. As the rising magma descends the water saturation surface in Figure 11.5, the lowering of the mole fraction of water in the melt brings the magma closer to the albite liquidus. At point c, where $X_{H_2O}^{m}=0.3$ and the pressure is 0.05 GPa, the magma reaches the water-saturated albite solidus and crystallizes to form albite and vapor.

Another useful way of presenting the albite liquidus in the hydrous system is with lines of constant activity of water. Two such lines are already present in Figure 11.5(B), that for the dry melt, where the $a_{H_2O}^{m}=0$, and that for the water-saturated solidus, where $a_{H_2O}^{m}=1$. Values of the activity of water between these extremes can be calculated from Eqs. (11.31) and (11.36). Figure 11.6 shows the same *P–T* projection of the albite liquidus as in Figure 11.5(B), but now it also has lines of isoactivity of water in the melt. Note that these lines do not lie in the planes of constant mole fraction of water in the melt but all radiate from the anhydrous atmospheric melting point of albite.

Isoactivity lines are particularly useful in considering the melting of albite in the presence of a mixed vapor phase, such

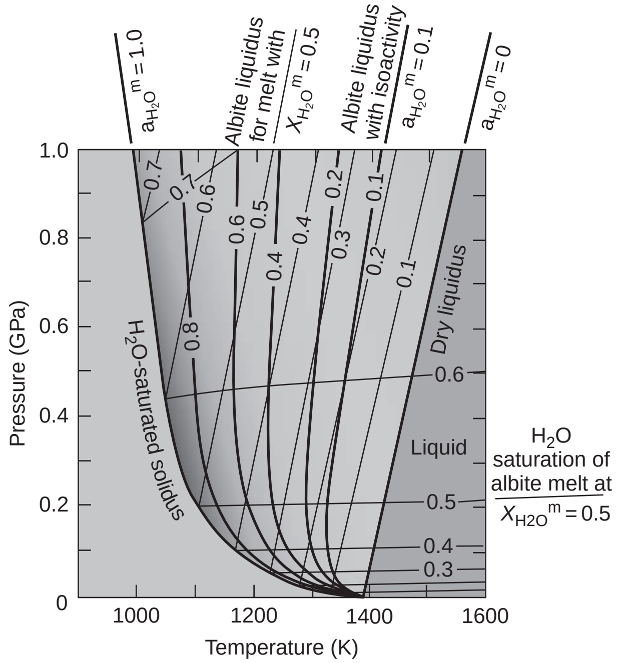


Fig. 11.6 Projection of the liquidus of albite onto the *P–T* plane showing lines of constant activity of water. Contours of constant mole fraction of water are also shown on the liquidus and water saturation surfaces (from Burnham, 1979).

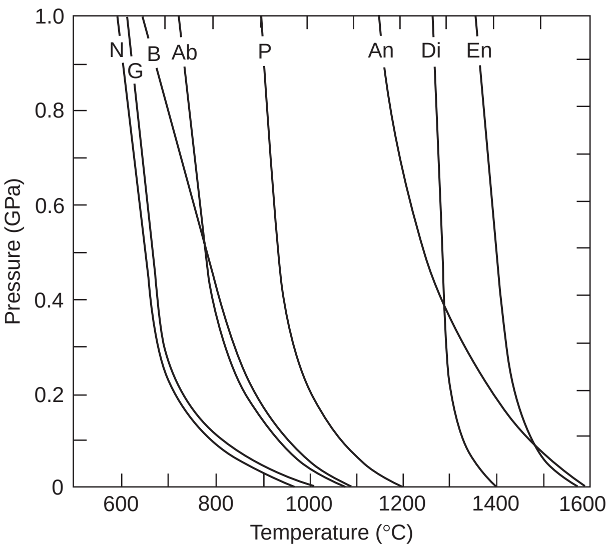


Fig. 11.7 Beginning of melting curves under water-saturated conditions for nepheline syenite (N), granite (G), tholeiitic basalt (B), albite (Ab), peridotite (P), anorthite (An), diopside (Di), and enstatite (En). (Data from various sources.)

as H_2O–CO_2. At low to moderate pressures, CO_2 is insoluble in albite melt, and thus it behaves essentially as an inert gas that reduces the activity of water in the vapor phase and in the coexisting melt. If a large amount of CO_2 were present, as might be the case if impure limestones were undergoing decarbonation reactions, the activity of water would be low. If the activity of water were only 0.1, melting of albite under a total pressure of 1 GPa would not occur until the temperature exceeded 1500 K. At this same pressure, if the activity of water were 1.0 (no CO_2 in the environment) melting of albite would occur at the much lower temperature of 1000 K.

The melting behavior of albite in the presence of H_2O has been dealt with in some detail here because it provides a good model for the behavior of most hydrous silicate melts. Although the liquidus temperatures of different minerals vary, the effect of dissolved water in lowering liquidus temperatures is similar. Even rocks that consist of a number of minerals and melt over a range of temperature show similar lowering of liquidus and solidus temperatures. Figure 11.7 shows the water-saturated solidi – also known as the *beginning of melting curves* – for a number of minerals and common rocks. Most of these solidi have similar negative slopes. The diopside and enstatite curves are steeper than the others, as might be expected in light of the discussion of the way in which water enters silicate melts (Section 11.2). In contrast, the slope on the anorthite curve is flatter, and basalt, which contains a significant amount of anorthite, also has a flatter solidus. Peridotite, which is composed essentially of olivine and pyroxene, is steep. At higher pressures than shown in this diagram, each of the solidi passes through a temperature minimum and then begins to rise with further increase in pressure. At these high pressures, the vapor phase is so compressed that it contributes very little to the volume change for the reaction from hydrous to less hydrous melt (Eq. (11.12)). Instead, the volume decrease on crystallizing begins to dominate and causes melting points to increase with increasing pressure (see Fig. 23.15).

One of the most important conclusions to be drawn from the data in Figure 11.7 is that water has a remarkable fluxing ability in lowering the melting points of silicates. For each of the materials in Figure 11.7 solution of a few weight percent water lowers solidus temperatures by several hundred degrees. For example, 5 wt% H_2O lowers the solidi of granitic and basaltic magmas by about 200 °C. We can conclude, therefore, that if water is present as a pure vapor phase in a region undergoing melting, the first melt forms on the water-saturated solidus. For granitic melts this is at about 600 °C at a pressure of 1 GPa. Such conditions are commonly attained in the highest grades of regional metamorphism. It is not surprising, then, to find evidence of partial melting in such rocks (see *migmatites* – mixed igneous and metamorphic rocks in Section 23.6). The amount of melt formed at the water-saturated solidus depends on the amount of water available, which may be quite small. The first formed magma mops up any available water. Once this water has entered the magma, further melting occurs only if the temperature rises, and this produces water-undersaturated magma. In addition, if the vapor phase is not pure H_2O, melting does not start at the water-saturated solidus but occurs only at the solidus for the particular activity of water (Fig. 11.6).

Figure 11.7, while showing the amount of lowering of solidus temperatures resulting from solution of H_2O in the melts, does not indicate how the compositions of melts vary with temperature. Of course, with congruently melting single minerals, such as albite, the only variation possible is in the

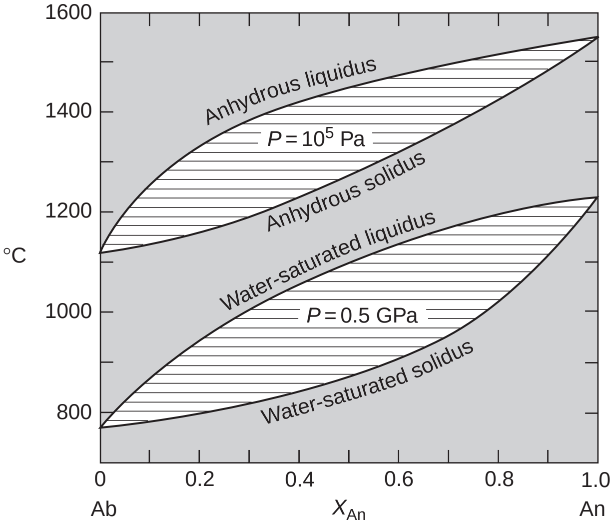


Fig. 11.8 System albite–anorthite at atmospheric pressure (10^5 Pa) and dry, and water-saturated at 0.5 GPa (after Yoder *et al.*, 1957).

content of water. But when melting involves more than one mineral, as in a rock, the composition of the liquid can change significantly with pressure. This is due to the different solubilities of water in the melts formed from the different minerals. If solubilities are similar, there will be little change in the composition of the melts; if the solubilities are very different, the changes will be great.

Because albite and anorthite are both framework silicates with the same number of exchangeable cations, their ability to dissolve water is similar, and thus they show similar lowering of melting points with increasing water pressure (note that their water-saturated solidi in Fig. 11.7 are similar). Consequently, the phase diagram for the plagioclase feldspars under water-saturated conditions at high pressure is almost identical to the anhydrous diagram at atmospheric pressure, except that at high pressure the dissolved water shifts the phase fields to lower temperatures (Fig. 11.8).

By contrast, the water-saturated solidus of diopside decreases less rapidly with increasing water pressure than does that of anorthite (Fig. 11.7), and thus the eutectic between these minerals shifts toward anorthite at high water pressures (Fig. 11.9). Yoder (1965) has shown that this eutectic (actually cotectic) in the pseudobinary system diopside–anorthite shifts from a composition of 42 wt% anorthite in the anhydrous system at atmospheric pressure to 78% anorthite in the water-saturated system at 1 GPa. Yoder has speculated that the plagioclase-rich composition of this eutectic under these conditions may be responsible for the composition of the rock-type *anorthosite*.

Similar changes in the extent of liquidus fields can be illustrated from ternary systems, such as that of forsterite–diopside–SiO_2 (Fig. 11.10). In this system, SiO_2 is the only mineral with a framework structure, and thus its melt is capable of dissolving more water than those of forsterite, diopside, or enstatite, each of which shows similar water solubilities. As a result, the phase relations involving SiO

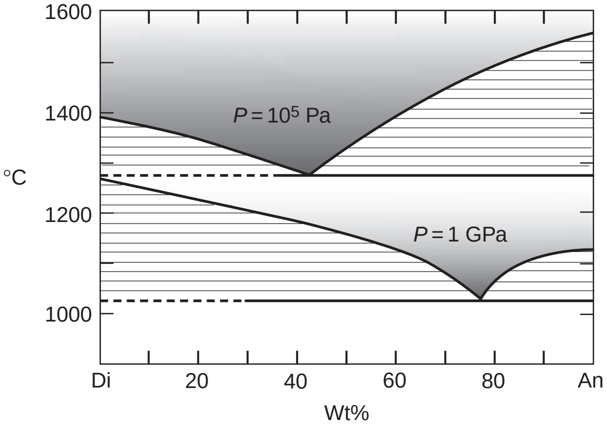


Fig. 11.9 System diopside–anorthite at atmospheric pressure (10^5 Pa) and dry, and water-saturated at 1.0 GPa. Alumina entering diopside makes diopside-rich parts of this system deviate slightly from binary behavior (after Yoder, 1965).

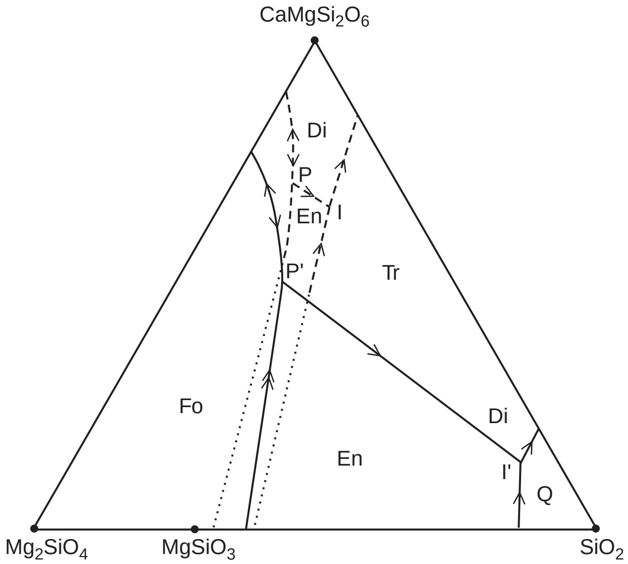


Fig. 11.10 System forsterite–diopside–silica. Dashed lines show position of phase boundaries on the liquidus in the anhydrous system at atmospheric pressure, and the solid lines are for the water-saturated liquidus at 2 GPa (after Kushiro, 1969).

are very sensitive to water pressure, but those between the ferromagnesian minerals are not. The peritectic between forsterite, diopside, and enstatite in the anhydrous system at atmospheric pressure (*P* in Fig. 11.10) changes composition only slightly to *P′* in the water-saturated system at 2 GPa. By contrast, the invariant point *I* in the anhydrous system at atmospheric pressure, which is a peritectic involving diopside, pigeonite, and tridymite, shifts by almost 60 wt% SiO_2 to *I′* in the water-saturated system at 2 GPa, where it involves diopside, enstatite, and quartz (Kushiro, 1969). Kushiro has suggested that the silica-rich composition of this invariant point under hydrous conditions may be responsible for the silica-rich nature of many rocks in the calcalkali series, which are typically more hydrous than are those in the tholeiitic series.

Even in systems containing only framework silicates, the composition of phase boundaries can shift if the melts of these phases have different solubilities of water. In the system quartz–albite–orthoclase–H_2O (Figs. 10.26 and 10.27), for example, silica-rich melts can dissolve water by either depolymerizing the framework structure (Eq. (11.13)), or by finding holes in the framework for molecular water, but in feldspar-rich melts, there is an additional solution mechanism. In albite-rich melts, one out of every four tetrahedra is an AlO_4^{5-} group which requires the juxtaposition of an Na^+ ion to maintain charge balance (Fig. 11.2(B)). When water dissociates into OH^- and H^+ ions and enters such a framework melt, the disrupted bridging oxygen need not become a hydroxyl group as in Figure 11.2(A) but, instead, can remain an oxygen ion that maintains charge balance by attracting the Na^+. Charge balance on the AlO_4 tetrahedron is maintained by one of its oxygens being hydrolyzed by addition of the H^+ released from the dissociation of water. The remaining hydroxyl ion from the dissociation of water is added to complete the other silica tetrahedron. This mechanism of solution presumably is not favored for the orthoclase-rich part of the system, because this part has the lowest solubilities of water (Holtz *et al*., 1995; Papale *et al*., 2006).

With increasing pressure, the greater solubility of water in albite-rich melts causes a greater depression of the liquidus in the albite-rich parts of the quartz–albite–orthoclase–H_2O. The minimum in the anhydrous 1-atmosphere system (Fig. 10.26), which changes to a eutectic in the lower-temperature hydrous system, consequently shifts toward albite as the pressure increases (Fig. 10.27).

Addition of water to anhydrous systems can also bring about changes in the phase relations by stabilizing hydrous minerals, such as amphiboles and micas. Hydrous minerals break down at high temperatures to anhydrous minerals plus a vapor phase. At low to moderate pressures this stability limit is raised by increasing pressures, but at very high pressures the anhydrous phases are again more stable. The stability limit of a hydrous phase can be calculated from Eq. (8.10) (see Problem 8.8 for the calculation of the stability limit of muscovite + quartz). Regardless of the mineral or group of minerals involved in a dehydration reaction, the stability limit in the presence of a pure H_2O vapor phase plots as a curved line with positive slope in a *P*–*T* diagram (Fig. 11.11). The beginning of melting curve of a rock, however, plots as a curved line with negative slope in such a diagram. At some point, therefore, these curves will intersect (A in Fig. 11.11), and above this pressure the hydrous mineral is in equilibrium with a melt.

If the vapor phase in equilibrium with a hydrous mineral is not pure H_2O (a mixture of H_2O and CO_2, for example), the stability limit of the hydrous phase decreases as the activity of water in the vapor decreases (Fig. 11.11). Water isoactivity lines can be drawn on the dehydration reaction surface, which resemble mirror images of the isoactivity lines on the solidus surface. Thus, as the activity of water decreases, the univariant lines for the dehydration reaction rotate in a counterclockwise direction, whereas the lines on the solidus rotate in a clockwise direction. The point of intersection of the solidus with the stability limit of the hydrous mineral therefore moves to progressively higher pressures with decreasing activity of water. Indeed, below a critical activity of water, the lines diverge and the hydrous mineral can never be stable at magmatic temperatures.

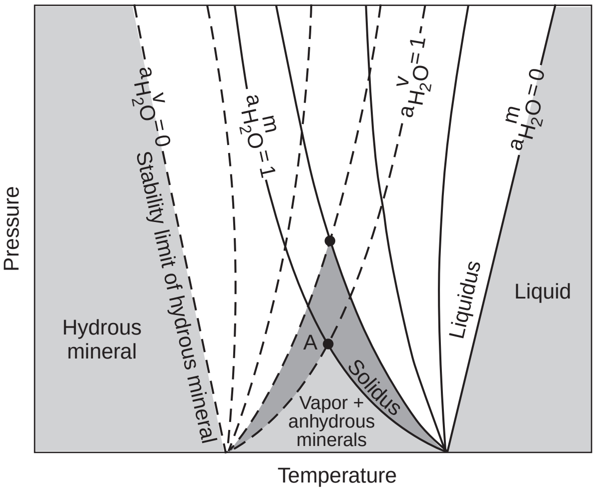


Fig. 11.11 Plot of the stability limit of a hydrous mineral in the presence of a vapor phase with variable activity of water; dashed lines indicate constant activity of water in the vapor. Solid curves are the vapor-saturated beginning of melting of a rock at different activities of water.

The stability limits of the hydrous minerals vary with composition and with associated minerals. In general, the stability limits of muscovite, biotite, and hornblende extend well above the solidus of granites, as long as the activity of water is moderately high. In the presence of excess water, hornblende is stable up to the liquidus temperatures of some basaltic magmas – camptonite, for example, typically has phenocrysts of hornblende. At pressures above 3 GPa, however, the stability limit of hornblende drops well below the solidus of basalts. Hornblende can therefore play a magmatic role only in the crust or uppermost mantle. Phlogopite, by contrast, is stable to higher temperatures and much higher pressures than is hornblende. Phlogopite may therefore be present in mantle rocks to considerable depth, where it could provide a source for the potassium in alkaline mafic magmas.

11.6 FRACTIONAL CRYSTALLIZATION OF HYDROUS MAGMA

At depth in the Earth, most magmas are probably neither water-saturated nor completely dry. Even when melts form on the water-saturated solidus, the limited supply of volatiles in most regions in the Earth is likely to cause melts to become undersaturated in water as the volume of melt increases. Once magma starts ascending, however, confining pressures decrease, and at some level the magma is likely to become vapor-saturated – most lavas, for example, contain vesicles. The pressure, and consequently the depth, at which vapor saturation occurs depends on the magma's initial content of volatiles, its saturation limit, its degree of crystallization, and the water content of minerals that crystallize from it (or carbonate content of minerals in CO_2-rich magma).

The degree of solidification of a magma at the time a vapor phase forms is of importance to a number of processes, such as explosive volcanism and pegmatite formation. Explosive volcanism is clearly of importance because of its potential hazard to humans. The formation of pegmatites is of interest because of their economic value. Pegmatites provide most of the feldspar used in the ceramics industry, and their large muscovite crystals were used extensively as electrical insulators in the manufacture of electronic tubes. Pegmatites also provide ores of many rare elements, such as Li, Be, Cs, Nb, Ta, W, Zr, U, Th, and the rare earths; they are also the source of many gemstones. The level of interest in pegmatites and much useful information concerning these rocks can be found on the web site of the Pegmatite Interest Group (PIG), which is hosted by the Mineralogical Society of America at www.minsocam.org/msa/special/Pig/.

Extensive field investigations of pegmatites, many of which are well exposed by mining, have made them some of the best known of all intrusive igneous bodies (Cameron *et al*., 1949). They have also been investigated experimentally (Jahns and Burnham, 1969; London, 1987, 2005). We will deal with them here to illustrate the factors involved in the fractional crystallization of hydrous magma.

The main evidence for the role of volatiles in the formation of pegmatites is their extremely coarse grain size, and their minerals, which indicate the presence of H_2O (mica), B (tourmaline), CO_2 (calcite), F_2 (topaz, apatite, fluorite), and Cl_2 (apatite). Strictly speaking, *pegmatite* is a textural term describing rocks with exceptionally coarse and variable grain size. Most pegmatites are of granitic composition, but nepheline syenite ones are common in alkaline igneous complexes, and even ones approaching gabbroic compositions can be found. Used loosely, without qualifiers, the term is normally understood to indicate a granitic composition. The conclusions we will draw here, however, will apply to pegmatites of any composition.

Crystals with diameters in excess of a meter are not uncommon in pegmatites (Fig. 11.12(A)). As will be shown in Chapter 12, the growth of such large crystals is possible only if few nucleation centers form and nutrients are able to move rapidly through the magma. The depolymerization of silicate melts brought about by high concentrations of water dramatically increases diffusion rates, and in the vapor phase these rates are still higher. Large crystals commonly nucleate on the walls of pegmatites (see Fig. 11.15(B)), and as they grow inward, they become larger, forming cone-shaped crystals; this is particularly characteristic of alkali feldspar crystals. Not all the material in pegmatites is coarse grained. Indeed, one particular rock type that characteristically occurs in pegmatites, known as *aplite*, consists of millimeter-size grains of quartz and feldspar with an allotriomorphic granular texture (Fig. 11.12(B) and (C)). The term *saccharoidal* is also used in reference to the sugary appearance of this rock. Aplite may form dikes cutting pegmatites, but most occurs as zones or irregular patches among the coarser crystals (Fig. 11.13). Many aplites are modally layered on the centimeter-scale (Fig. 11.12(B)). The layering is commonly cuspate and resembles, on a coarser scale, the layering seen in botryoidal mineral specimens such as malachite. The layering may also show discordances and evidence of deformation (Fig. 11.12 (C)). In granitic pegmatites, quartz and feldspar are commonly intergrown in graphic intergrowths (Figs. 10.5 and 11.12(A)). Evidence of replacement of country rock is common. Xenoliths of country rock in pegmatites often are so strongly replaced by alkali feldspar and quartz that they are only barely recognizable as xenoliths (Fig. 11.14). In silica-undersaturated alkaline pegmatites similar textures involve albite, nepheline, and sodalite.

Most pegmatites form lenticular bodies (Fig. 11.15), but more continuous dike- and sill-like forms are common. They may occur in the upper parts of batholiths and stocks or intrude the overlying rocks. Some occur in metamorphic rocks where they have no obvious related magmatic source (Fig. 11.16). At low metamorphic grades, veins of quartz and carbonate are common, but once the middle to upper amphibolite grade is reached (Section 16.3), veins contain alkali feldspar and become pegmatitic. Their formation appears to be caused by the temperature rising above the beginning of melting of low-temperature hydrous granitic fractions (see Fig. 16.6). Most pegmatites are zoned in both grain size and composition. Most have quartz cores (Figs. 11.12 and 11.15) and concentrations of potassium-bearing minerals toward the top.

If pegmatites are formed by the intrusion of magma, a number of features are hard to explain. Unlike almost all other igneous rocks, pegmatites are never porphyritic. They certainly contain crystals that range enormously in grain size, but there is never one population of crystals that could be considered phenocrystic. Pegmatites never have chilled margins. Indeed, some of the coarsest crystals in pegmatites can be seen to have nucleated on the contacts and grown into the pegmatite (Figs. 11.15(B) and 12.19). Aplite is a fine-grained rock, and in some pegmatites it does form at the contact, but it is just as likely to occur within a body amongst patches of coarse pegmatite (Fig. 11.13). Its grain size is not related to rapid quenching at the margin of pegmatites. Although pegmatites commonly contain elongate crystals, such as those of tourmaline in Figure 11.15, these crystals never show an alignment that would be caused by flow of magma. Instead, they are commonly oriented perpendicular to the walls of the pegmatite and appear to have grown out into a static environment. One could argue that the tourmaline crystals shown in the roof and floor zones of the pegmatite in Figure 11.15(A) are not quite perpendicular to the wall and hence might indicate flow. However, it will be noted that the roof and floor crystals have the same direction of tilt, which would not be the case for flow. The tilt in this pegmatite was caused by the entire body undergoing a slight top-to-left shear due to later deformation. Unlike other igneous bodies, pegmatites commonly have no feeder. Mining operations have provided

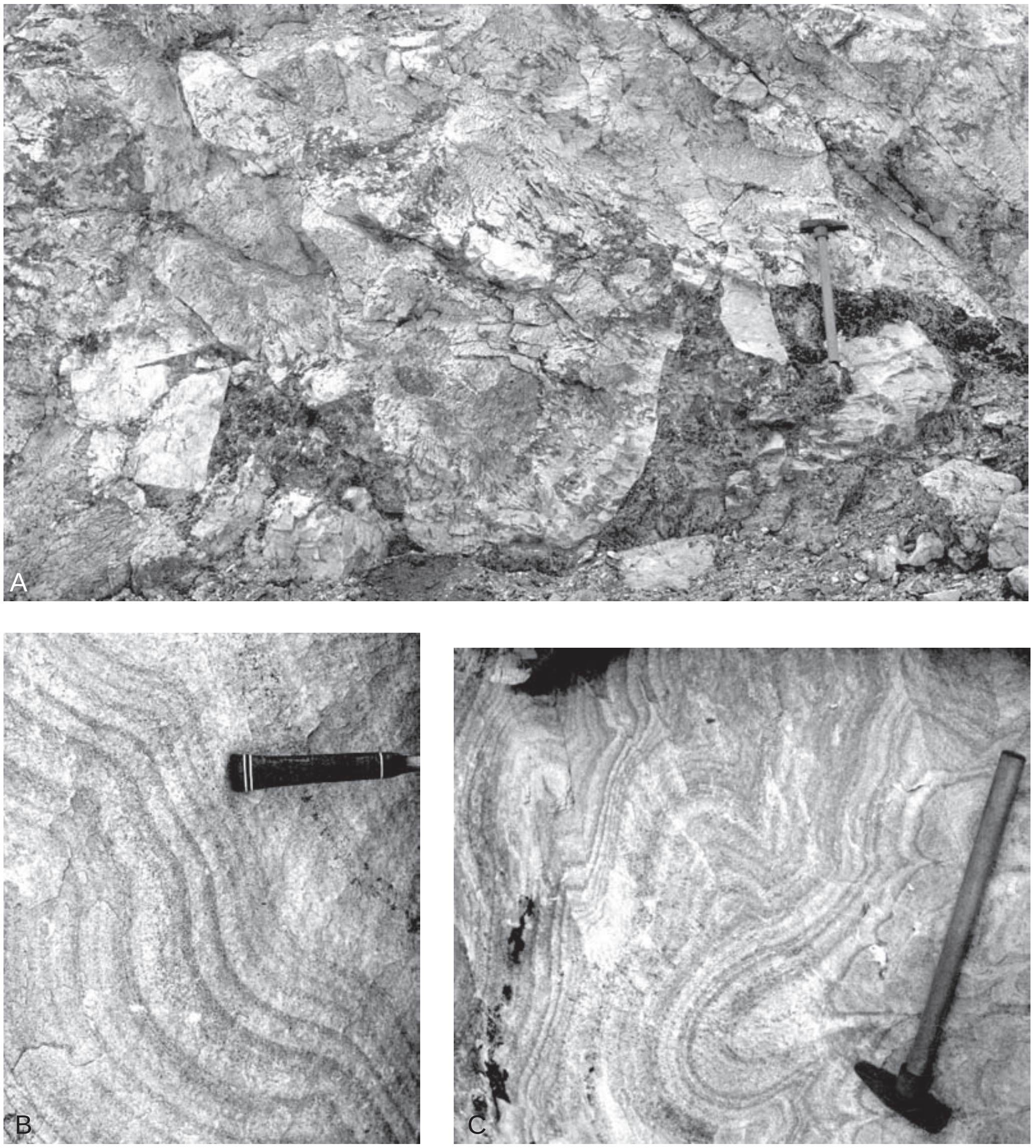


Fig. 11.12 Hale pegmatite, Middletown pegmatite district, Connecticut. (**A**) Core zone with meter-sized microcline crystals (sledgehammer with meter-long handle for scale) extending into massive quartz (dark); around the core, microcline forms a graphic intergrowth with quartz. (**B**) Banded aplite from steeply dipping upper contact of pegmatite. Gradational zoning is due to variations in abundance of fine-grained muscovite. Note that in each layer the abundance of muscovite suddenly increases and then gradually decreases toward the interior of the pegmatite (lower left). Near the hammer handle a zone contains a few larger quartz crystals growing perpendicular to the layering. The contact of the banded aplite with the pegmatite of (A) is abrupt and concordant. (**C**) Complex zoning in aplite with core of pegmatite to left.

excellent three-dimensional views of many pegmatites, which commonly form isolated pods (Cameron *et al*., 1949). Pegmatites are clearly not formed from hot crystal-laden viscous magmas that are intruded into cold country rocks. They must form in rocks that are at the same temperature and are probably formed by segregation of volatile-rich fractions that already exist in the surrounding rock, regardless of whether that host be a crystallizing igneous rock or a partially melted metamorphic rock.

Jahns and Burnham (1969) propose that the development of a separate vapor phase in the magma is responsible for the development of pegmatites. At the temperatures and pressures under which most pegmatites form, this vapor phase is a supercritical fluid with a density similar to that of water at room temperature and pressure. If a vapor phase appears early enough in the crystallization history of a magma, it will be able to segregate easily and form large pegmatites. The vapor phase is also capable of altering and replacing early-formed minerals (Fig 11.14) to produce assemblages that would otherwise be difficult to explain by simple magmatic crystallization.

Once a separate vapor phase forms, large crystals start growing because of the higher diffusion rates through the vapor. The size of crystals depends, in large part, on how interconnected the vapor phase is. In addition, as soon as this phase forms, certain elements, in particular those with large ionic radius and charge, leave the highly polymerized silicate melt in favor of the less polymerized vapor phase. This

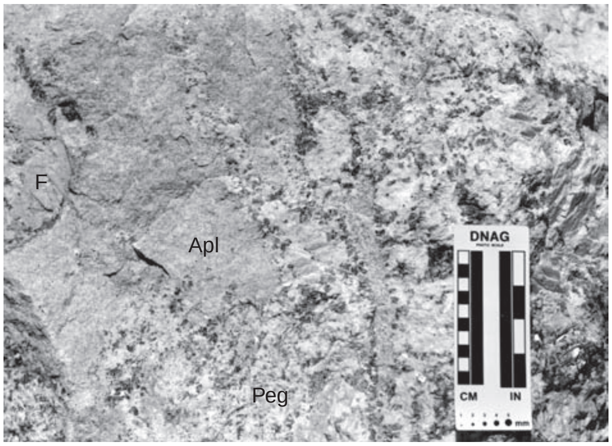


Fig. 11.13 Pegmatite and aplite from Gotta Quarry, Middletown pegmatite district, Connecticut. Note juxtaposition of single coarse alkali feldspar crystal on left with fine-grained aplite.

Fig. 11.14 Xenolith of amphibolite in granitic pegmatite almost totally replaced by alkali feldspar and quartz; La Verendrye Park, Quebec.

partitioning is so strong that elements that may be present in trace amounts (ppm) in the magma become highly concentrated in the vapor phase. It is from such solutions, for example, that large beryl and tourmaline crystals grow (Fig. 11.15). Some of these volatile constituents have significant fluxing effects on the magma. Experiments at 0.1 GPa show that the water-saturated solidus of a quartz–albite–orthoclase mixture is lowered by 60 °C by addition of 2 wt% B_2O_3 and by 160 °C by the addition of 2 wt% F (Manning and Pichavant, 1983).

The pressure on the vapor phase may weaken surrounding rocks and cause them to fracture, and by so doing release a pulse of hydrothermal fluid from which minerals will be deposited as the solution cools. The accompanying drop in pressure on the pegmatitic magma brings about rapid crystallization. Consider, for example, a pegmatitic magma on the water-saturated solidus of albite at point c in Figure 11.5(A). Rupturing of the pegmatite walls and loss of the volatile pressure would drop the magma below the solidus. The suddenness of such an event brings about rapid crystal nucleation and the development of fine-grained aplite. Repeated losses of pressure could be responsible for the formation of banded aplite (Fig. 11.12(B)). Jahns and Burnham (1969) claimed that once a separate vapor phase forms in a granitic melt, potassium preferentially enters that phase, from which large potassium feldspar and mica crystals would form. Loss of potassium to the vapor phase causes the residual melt to become enriched in albite, which then crystallizes to form aplite, which generally is a more sodic rock than the host pegmatite.

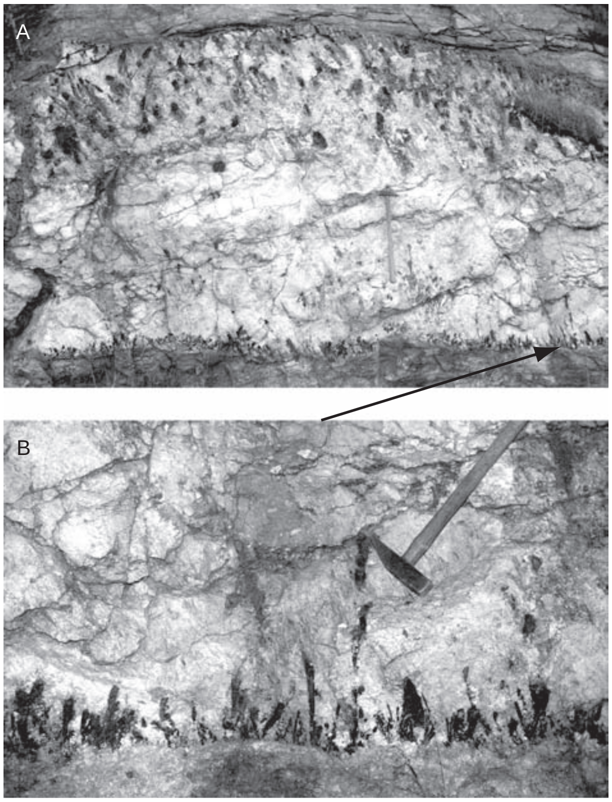


Fig. 11.15 (**A**) Three-meter-thick lenticular zoned pegmatite, Lewiston, Maine. Dark crystals are tourmaline; note that tourmaline crystals at top are larger than ones at base. Quartz-rich core is overlain by a zone rich in muscovite (reflecting light) and underlain by a zone of graphic granite. (**B**) Detail of lower contact of pegmatite in (A) showing how tourmaline crystals have nucleated on the contact and grown into the zone of graphic granite.

Despite the common acceptance of the Jahns and Burnham (1969) model for pegmatite formation, London (2005) has shown that subsequent experimental studies have not supported the claim that potassium preferentially enters the vapor phase. He has also argued that many textures in pegmatites are more easily explained by rapid crystal growth resulting from the undercooling of the melt by 75 to 100 °C than they are by slow cooling of a melt. Figure 11.17, for example, shows a photomicrograph from a narrow granophyre dike that was formed by the rapid quenching of a granitic magma in the Ossipee ring complex of New Hampshire (Fig. 4.52). This dike is so fine grained that it was used by the native people to make arrowheads. The rock's textures, however, resemble those of a pegmatite except for the scale, millimeters instead of meters. The rock is porphyritic in the sense that it contains larger crystals of feldspar in a fine-grained groundmass. However,

Fig. 11.16 Pegmatite cutting interlayered quartz–plagioclase–biotite–hornblende gneiss and calcsilicate of the Bucksport Formation at Pemaquid Point lighthouse, Maine. The country rocks underwent a Buchan-type metamorphism (Section 16.3); that is, low pressure–high temperature. Pegmatites are common where the rocks were heated into the sillimanite zone (see beginning of melting curve in Fig. 16.6). Pegmatites hosted in metamorphic rocks commonly show a pinch and swell structure known as boudinage (after the French word *boudin* meaning sausage) which results from stretching of the competent pegmatite in the less competent metamorphic rock during deformation.

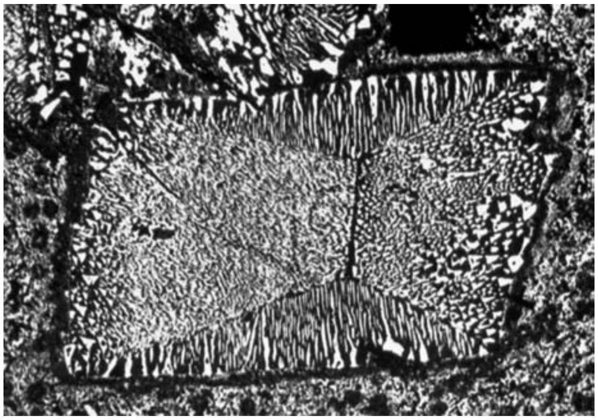


Fig. 11.17 Photomicrograph of small phenocryst of alkali feldspar with quartz inclusions (dark) forming a granophyric (graphic) intergrowth in a narrow dike from the Ossipee Ring Complex, New Hampshire (width of field ~1 mm). Feldspar growth was too rapid for rejected silica in the surrounding melt to diffuse away, and consequently quartz grew as rods perpendicular to the feldspar faces. Growth sectors in the feldspar crystal are brought out by the orientation of the quartz rods; rods oriented at a high angle to the plane of the section appear as dots or triangles. (Sample collected by B. Calagero.)

crystallization was so rapid that, as the feldspar crystals grew, diffusion was unable to remove excess silica from the melt in front of the advancing crystal faces and, consequently, rods of included quartz grew perpendicular to the crystal faces to produce a granophyric texture. Except for scale, the millimeter-size feldspar crystal in Figure 11.17 is identical to the 2-m-diameter crystal in Figure 11.12, both containing rods of quartz perpendicular to the faces of the crystals.

Although "conventional wisdom" claims that crystals in igneous rocks grow to large sizes as a result of slow cooling (e.g. compare plutonic versus extrusive grain sizes), we will see in Chapter 12 that grain size in igneous and metamorphic rocks is determined by the number of nuclei formed and not by cooling rates. If a large number of nuclei are formed, and nothing acts subsequently to dissolve them, the resulting rock will be finer grained than if a smaller number of nuclei had formed (see discussion of Fig. 12.13). The coarse grain size of pegmatites indicates that very few nuclei are formed, whereas a large number of nuclei must be formed when aplite crystallizes. This inability for pegmatites to form nuclei is also indicated by the common growth of crystals out from the walls of pegmatites (see tourmaline crystals in Fig. 11.15). Such crystals find it easier to grow from a previously existing crystal than to nucleate a new crystal within the pegmatite (see Fig. 12.19). Of course, nutrients for crystal growth in

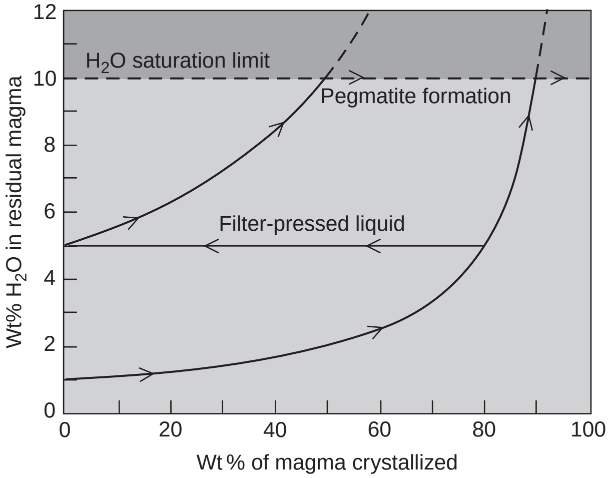


Fig. 11.18 Plot of the change in concentration of water in a granitic magma as a result of crystallization of anhydrous minerals. The water saturation limit of 10 wt% is for a pressure of 0.4 GPa.

pegmatitic fluids must be able to diffuse distances of the order of decimeters to meters to get to these growth sites. Although nucleation can be inhibited by rapid cooling, the separation of residual liquid from crystal mush can produce nuclei-free liquids that eventually crystallize pegmatitic scale material (see discussion of Fig. 12.19). It is also possible that some of the volatile constituents that are enriched in pegmatitic melts not only lower the viscosity of the melt (Dingwell *et al.*, 1996) but also depolymerize it to the extent that crystal nucleation becomes difficult (London, 1987; Swanson and Fenn, 1992).

These, then, are some of the processes involved with the formation of pegmatites. Their effectiveness depends on how early in the crystallization history the magma becomes water saturated. If saturation occurs early, bubbles are free to rise and segregate to form large pegmatites; if saturation occurs late, restricted movement of the vapor phase may permit only small *miarolitic* cavities to form. Also, bubbles rising great distances through magma may be able to scavenge more of the extremely minor constituents of the magma than bubbles that travel only a few meters. Pegmatite formation, then, depends on how rapidly a fractionating hydrous magma reaches the water-saturation limit.

Let us assume that a totally liquid magma initially contains a weight fraction of water Wt°. When crystallization begins, the assemblage of minerals that forms has an average weight fraction of water C. The weight fraction of water in the residual magma at any stage of crystallization (Wt) can be determined by writing a mass balance equation, in which the weight of water in the residual liquid plus the water that has entered the minerals must equal the amount of water in the initial magma. If the weight fraction of magma that has crystallized is F, the mass balance equation can be written as

$$\mathrm{Wt}(1-F) + C(F) = \mathrm{Wt}^\circ \tag{11.35}$$

On rearranging and multiplying by 100, the wt% H_2O in the magma at any stage of crystallization is

$$\text{wt\% } \mathrm{H_2O} \text{ in magma} = 100 \times \frac{\mathrm{Wt}^\circ - C(F)}{1-F} \tag{11.36}$$

We can use Eq. (11.36) to trace the change in concentration of water in a magma during crystallization and to predict when the magma would become water saturated. Formation of bubbles of vapor as a result of crystallization is commonly referred to as *resurgent boiling* (also, *second boiling*) to contrast it with the boiling that would occur if the melt were raised to very high temperatures.

If a granitic magma initially contained 1 wt% H_2O and the only crystals that formed during the early stages of solidification were quartz and feldspar, which contain no water ($C = 0$), the concentration of water in the magma would increase along the path shown in Figure 11.18 as crystallization proceeded. If this magma were crystallizing at a depth where the pressure was 0.4 GPa, we know from Figure 11.1 that the water saturation limit would be 10 wt%. This concentration would be reached after 90% of the magma had crystallized. In the residual 10 wt% liquid, then, bubbles of the vapor phase would nucleate and grow as crystallization went to completion. This would occur so late in the crystallization that little segregation of the vapor phase would be possible, and consequently, no large bodies of pegmatite would form.

During crystallization of this magma, compaction of crystals could displace residual liquid into the upper part of the magma chamber, where it would collect and form a new batch of fractionated magma. For purposes of illustration, consider that this filter pressing occurred when 80% of the original magma had crystallized. The new batch of fractionated magma would contain 5 wt% H_2O (Fig. 11.18), and it would reach the 10 wt% H_2O saturation limit when only 50% crystallized. This would be early enough for considerable segregation of the vapor phase to occur, and large pegmatites would result.

In conclusion, then, large pegmatites form only in magmas that initially are rich enough in water to become saturated before they are more than two-thirds crystallized; at higher degrees of crystallization segregation of the vapor phase becomes increasingly difficult. The initial water content necessary to satisfy this condition depends on the water saturation limit of the magma, which, in turn, is a function mainly of pressure. Granitic magma, for example, requires about 4 wt% H_2O at 0.5 GPa but only 1.5 wt% at 0.1 GPa.

11.7 EFFECT OF CO_2 ON MELTING IN SILICATE SYSTEMS

At low to moderate pressures the low solubility of CO_2 in silicate melts makes this gas a relatively inert component. Indeed, in many experimental studies CO_2 is added to hydrous systems as an inert component to lower the activity of H_2O. With high pressure (>1.5 GPa), however, CO_2 does become soluble, especially in the less polymerized mafic

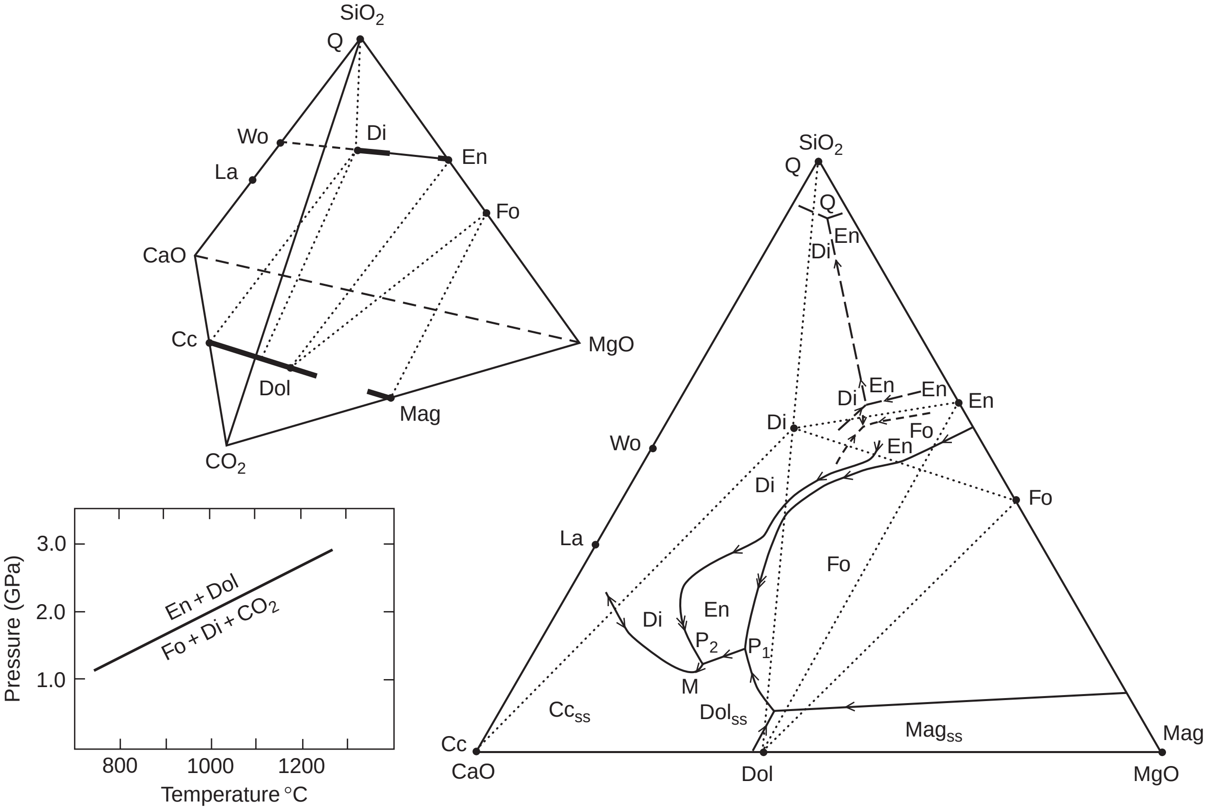


Fig. 11.19 System CaO–MgO–SiO_2–CO_2. Vapor-saturated liquidus phase relations at 3.0 GPa are projected onto the ternary CaO–MgO–SiO_2 face from CO_2. Insets show the compositions of phases in the quaternary system and the decarbonation reaction enstatite (En) + dolomite (Dol) = forsterite (Fo) + diopside (Di) + CO_2 (after Eggler, 1976). Also shown in the ternary projection with short dashes is the position of the Fo–En–Di eutectic in the vapor-free system at 3.0 GPa, and with long dashes, the Fo–En–Di peritectic and the Di–En–Q invariant point at 2.0 GPa (after Kushiro, 1969).

melts. Consequently, at high pressures, CO_2 does affect liquidus temperatures, and through carbonation reactions, it influences the minerals that crystallize. In no system are these effects more dramatic than in the system CaO–MgO–SiO_2 (Eggler, 1976; Wyllie and Huang, 1976). This system is of particular interest because it models simple peridotites and is therefore of importance when considering melting in the upper mantle.

The compositions of minerals that form in this simplified peridotite system in the presence of CO_2 are shown in the quaternary plot CaO–MgO–SiO_2–CO_2 of Figure 11.19. Apart from the carbonates, calcite, dolomite, and magnesite, all minerals plot on the ternary face CaO–MgO–SiO_2. Melts in this system, however, contain variable amounts of CO_2 and thus plot in the tetrahedron. Although the phase relations in the quaternary system can be illustrated more simply when projected onto the CaO–MgO–SiO_2 face, it is important to keep in mind that the boundary lines shown in this ternary diagram are projected from within the tetrahedron.

The ternary projection in Figure 11.19 shows the CO_2-saturated liquidus relations in the system CaO–MgO–SiO_2–CO_2 at 3.0 GPa. The most significant difference between this diagram and its equivalent one for low pressures (Fig. 11.10) is in the behavior of forsterite and diopside. The assemblage olivine + augite is characteristic of all basaltic rocks formed under crustal conditions regardless of their degree of silica saturation. Above a pressure of 2.9 GPa, however, this assemblage is no longer stable in the presence of a CO_2 vapor phase; instead, it is converted to enstatite + dolomite according to the reaction

$$\begin{array}{l}\underset{\text{forsterite}}{Mg_2SiO_4} + \underset{\text{diopside}}{CaMgSi_2O_6} + \underset{\text{vapor}}{2CO_2} \\ \quad \rightleftharpoons \underset{\text{enstatite}}{Mg_2Si_2O_6} + \underset{\text{dolomite}}{CaMg(CO_3)_2}\end{array} \tag{11.37}$$

As with other carbonation or hydration reactions (Eq. (8.9), Problem 8.8), the equilibrium curve for this reaction has a positive slope that increases with pressure, and the assemblage with the gas is stable on the high-temperature, low-pressure side. This equilibrium curve is shown in Figure 11.19.

Because of the carbonation reaction, the cotectic between olivine and clinopyroxene that exists in this system at low pressure (Fig. 11.10) is replaced at 3.0 GPa by two cotectics involving either enstatite + diopside or enstatite + forsterite (Fig. 11.19). The narrow primary field of enstatite which now separates the fields of diopside and forsterite extends across the diopside–forsterite join, which means that extremely silica-undersaturated larnite-normative melts can be in equilibrium with enstatite. Indeed, the enstatite field extends to very low silica concentrations. At the same time, the melts become enriched in carbonate (up to 40% CO_2) and hence plot well into the quaternary system.

Let us follow the path of a fractionally crystallizing CO_2-saturated liquid that initially has a composition on the enstatite–forsterite boundary in the ternary subsystem diopside–forsterite–enstatite. Crystallization of enstatite and forsterite drives this liquid down the cotectic between these two minerals, first crossing the diopside–forsterite join and then the diopside–dolomite join. Note that by the time the liquid crosses the diopside–dolomite join, the boundary curve between enstatite and forsterite, when extrapolated, no longer intersects the enstatite–forsterite join. Forsterite must, therefore undergo reaction with the liquid along the low-temperature part of this curve. This reaction line is terminated at an isobaric quaternary peritectic, P_1, where it intersects the field of dolomite solid solution. At this point all forsterite reacts out to leave only enstatite and dolomite solid solution to continue crystallizing down to the next peritectic, P_2, where enstatite is eliminated by the reaction to produce diopside and dolomite solid solution. From this point, the residual melt fractionates down to a minimum, M, at 1110 °C where diopside and a carbonate solid solution with Ca/Mg ratio of 70:30 crystallize.

The magnitude of the effect of CO_2 on the melting of peridotite compositions at 3 GPa can be seen by comparing the compositions of the liquids at the peritectic P_1 and the eutectic between forsterite, enstatite, and diopside in the volatile-free system under the same pressure (Fig. 11.19). Heating of an olivine-bearing peridotite mantle under CO_2-saturated conditions at 3.0 GPa would produce a melt of peritectic composition P_1 containing 17% SiO_2 and 27% CO_2. Under volatile-free conditions, this same mantle would first form a melt at the forsterite–diopside–enstatite eutectic. Under hydrous conditions, the first melt would form at the forsterite–enstatite–diopside peritectic (Figs. 11.19 and 11.10). Thus melting of identical mantle peridotite under these three different sets of conditions could generate a critically silica-undersaturated carbonate melt (+ CO_2 vapor), a silica-saturated melt (no vapor), and a silica-oversaturated melt (+ H_2O vapor). A vapor containing both CO_2 and H_2O could produce melts with a wide range of compositions between these extremes depending on the composition of the vapor. Moreover, the range of magma compositions that could be generated from these melts is still greater if they were to crystallize fractionally. Magma formed at P_1 in Figure 11.19 could fractionate down to a *carbonatitic* magma at the minimum, and a melt at the olivine–enstatite–diopside peritectic under hydrous conditions could fractionate down to a *granitic* residue.

Also of particular interest is the suddenness with which the phase relations under the CO_2-saturated conditions change with increasing pressure (Fig. 11.20). At low pressures, CO_2 has only a slight effect in lowering solidus temperatures relative to those in the volatile-free system, because of its limited solubility in the melt. Above about 2.5 GPa, however, the solidus begins dropping more rapidly with increasing pressure, and between 2.7 and 2.9 GPa it drops 350 °C. This dramatic lowering results from the increased solubility of CO_2 in the melt as the decarbonation reaction is approached. This seems consistent with the model for silicate melts (Section 11.2), where the structure of the melt mimics the structure of the crystalline phases that form from it. One can imagine that as the carbonation reaction is approached, an increasing fraction of the melt would have a carbonate structure.

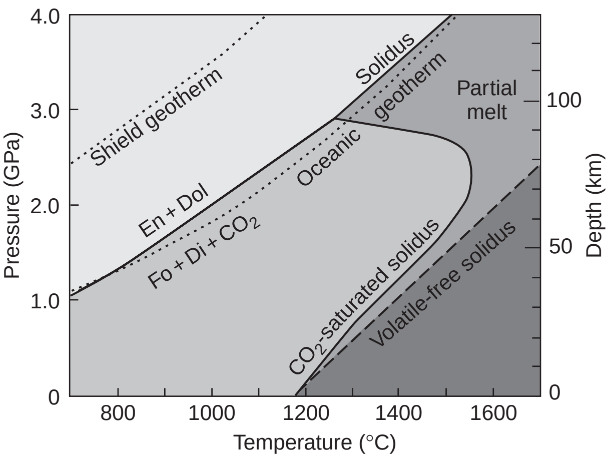


Fig. 11.20 Schematic representation of the solidus for a peridotite containing 0.2 wt% CO_2. Below 2.9 GPa CO_2 exists as a vapor; above this pressure it is present as dolomite because of the carbonation reaction of Fo + Di + CO_2 = En + Dol. Also shown are geotherms for shield and young oceanic regions. (After Eggler, 1976; Wyllie and Huang, 1976.)

As the pressure increases from 2.7 to 2.9 GPa and the solidus temperature drops precipitously, the composition of the first-formed melt shifts from silica-saturated to silica-undersaturated. Melts ranging in composition from olivine tholeiite, through alkali olivine basalt, melilitite, kimberlite, to carbonatite can all form from a common source peridotite over a short interval of depth. Also, carbonated melts formed in the deeper zones will, on rising to depths of about 90 km where the pressure drops below 3.0 GPa, start crystallizing and liberating CO_2 vapor, which may, then, be responsible for the explosive emplacement of these magmas.

This experimental work of Eggler (1976) and Wyllie and Huang (1976) confirms work by others that CO_2 cannot exist as a vapor phase in the mantle below depths of about 100 km beneath continents because estimated geotherms lie well below the carbonation reaction (Fig. 11.20). If CO_2 is present it must be in the form of dolomite. Beneath young oceans, however, the geotherm may lie on the high-temperature side of the carbonation reaction and then CO_2 would exist as a vapor phase. Under such conditions, the geotherm would cross the solidus at a depth of approximately 90 km. The small amount of melt that would be formed could well explain the presence of the low-velocity zone in such regions.

11.8 ROLE OF OXYGEN FUGACITY IN PHASE EQUILIBRIA

Oxygen, by volume, is by far the most abundant component in silicate melts. The amount in vapors in equilibrium with these melts, however, is extremely small, and its partial pressure, measured in bars, typically ranges from 10^{-8} to 10^{-11}. Despite these low values, small changes can bring about major changes in mineralogy, which in turn can change the path of magmatic fractionation. These changes can take place even when a separate vapor phase is not present through changes in the fugacity of oxygen. Visualize, for example, a magma with no separate vapor phase in a sealed container which is connected with an external reservoir of oxygen through a membrane that is porous only to oxygen. By changing the pressure and temperature on the reservoir, we can change the fugacity of oxygen in the reservoir. Oxygen will then diffuse either into or out of the magma chamber through the semi-permeable membrane until the fugacities of oxygen in the magma and the reservoir are the same.

Changes in mineral assemblages resulting from changes in oxygen fugacity are caused largely by changes in the oxidation state of iron (also Mn and possibly Ti). For example, the minerals hematite and magnetite are related by the reaction

$$2Fe_3O_4 + 0.5O_2 \rightleftharpoons 3\,Fe_2O_3 \tag{11.38}$$

Ferromagnesian silicates can also be involved, as for example in

$$3Fe_2SiO_4 + O_2 \rightleftharpoons 2\,Fe_3O_4 + 3SiO_2 \tag{11.39}$$

For any such reaction an equilibrium constant can be written, which, for Eq. (11.39), for example, would be

$$K = \frac{(a_{\mathrm{Mt}})^2 (a_{\mathrm{Q}})^3}{(a_{\mathrm{Fa}})^3 (fO_2)} = \exp\left(\frac{-\Delta G}{RT}\right) \tag{11.40}$$

Knowing the value of ΔG for the reaction, K can be calculated as a function of temperature (not very sensitive to P; see Section 9.8). If the activities of magnetite, quartz, and fayalite are known, the fugacity of oxygen can be determined for any temperature. Because quartz is always pure SiO_2, the activity of quartz will be unity, as long as the crystalline phase is present. Magnetite and fayalite may also be present in their pure forms. If not, their composition must be known before their activity can be determined. If we assume for the moment that all of these minerals are present as pure phases, the oxygen fugacity in equilibrium with quartz, fayalite, and magnetite (QFM) is given by rearranging Eq. (11.40) to

$$\ln fO_2 = \frac{\Delta G^{\mathrm{QFM}}}{RT} \tag{11.41}$$

A plot of $-\log_{10} fO_2$ versus T for Eq. (11.41) is given in Figure 11.21.

Equation (11.41) shows that if a rock contains quartz, fayalite, and magnetite, the oxygen fugacity in equilibrium with this assemblage is determined by the temperature. At constant temperature, then, the oxygen fugacity is fixed. This assemblage of minerals is therefore said to be an *oxygen buffer* – the *QFM buffer*. Different mineral assemblages buffer the oxygen fugacity at different values. In addition to the QFM buffer, Figure 11.21 shows the buffer for coexisting hematite and magnetite (HM) (Eq. (11.38)), and for the reaction involving magnetite and wüstite (MW)

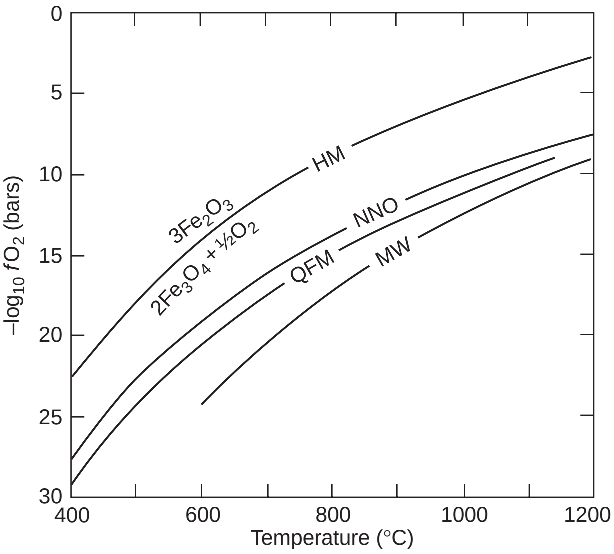


Fig. 11.21 Plot of $-\log fO_2$ (bars) versus temperature for a number of common oxygen buffering reactions: hematite–magnetite (HM), nickel–nickel oxide (NNO), quartz–fayalite–magnetite (QFM), and magnetite–wüstite (MW). See text for reactions.

$$3FeO + 0.5O_2 \rightleftharpoons Fe_3O_4 \tag{11.42}$$

and nickel–nickel oxide (NNO)

$$Ni + 0.5O_2 \rightleftharpoons NiO \tag{11.43}$$

The buffers QFM, MW, and NNO have been used widely in experimental work to control oxygen fugacities because they cover the range of fugacities normally encountered in the crust and upper mantle. Most igneous rocks, in fact, are formed under fugacities near the QFM buffer (Carmichael and Ghiorso, 1986).

Through reactions such as those plotted in Figure 11.21, oxygen fugacities place restrictions on the minerals that can coexist and crystallize from melts. Similarly, there are no conditions under which hematite and fayalite can crystallize together. Likewise, if the oxygen fugacity in a magma is buffered by some external reservoir at a value equal to that of the NNO buffer, fayalite could not crystallize.

To appreciate the effect that oxygen fugacity can have on crystallizing melts, we will examine the simple binary system FeO–Fe_2O_3 (Fig. 11.22). This system contains the phases iron-oxide liquid, iron (α and γ), wüstite, magnetite, and hematite, all of which can contain variable amounts of oxygen depending on the oxygen fugacity. The system has liquidi for each of the crystalline phases, two eutectics – one between hematite and magnetite, the other between wüstite and γ-iron – a peritectic (P) involving magnetite and wüstite, and solvi between each pair of juxtaposed crystalline phases. In this respect, the diagram is not unlike binary systems already discussed in Chapter 10. It differs, however, in that for any composition of phase or pair of phases it is possible to specify, for a given temperature, the oxygen

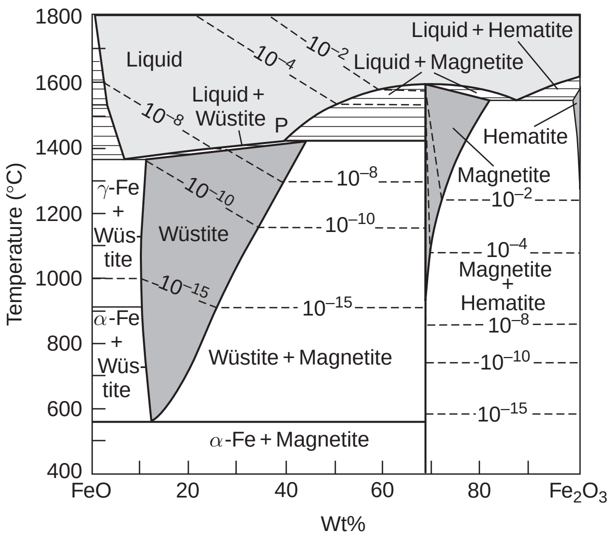


Fig. 11.22 System FeO–Fe_2O_3. Dashed lines are lines of constant oxygen fugacity measured in bars. (After Darken and Gurry, 1945; Muan, 1958.)

fugacity that must exist. The phase diagram can thus be contoured with lines of equal oxygen fugacity (isobars). In fields where two phases coexist, such as hematite and magnetite, the oxygen isobars are horizontal lines, because such an assemblage, regardless of the relative abundances of the phases, is an oxygen buffer for a given temperature. In fields of a single phase, such as wüstite or liquid, the isobars are sloping lines because oxygen fugacity here is a function of both temperature and composition.

Because changes in oxygen fugacity can affect phase assemblages just as easily as changes in temperature (and pressure) in systems such as that of FeO–Fe_2O_3, the oxygen fugacity must be specified before we can interpret what will happen to a composition with changing temperature. Two extreme cases can be considered: one where the bulk composition of the system remains constant – no oxygen is allowed to enter or leave the system; the other is where the oxygen fugacity is maintained at a constant value – this requires that oxygen be able to enter or leave the system. Conditions ranging between both extremes are found in nature.

Consider first the cooling of a melt containing 50% Fe_2O_3 under conditions of constant bulk composition. At 1600 °C the oxygen fugacity in equilibrium with this melt is 10^{-3} bars (Fig. 11.22). On cooling, the melt reaches the magnetite liquidus at 1530 °C where the oxygen fugacity is 10^{-4} bars. The melt then descends the liquidus to the peritectic, P, where magnetite reacts with the liquid to produce wüstite. Once all liquid is consumed by the reaction, magnetite and wüstite are free to cool. As the temperature drops, the oxygen fugacity continues to fall, and wüstite changes composition to progressively lower oxygen contents. At 560 °C wüstite breaks down to α-iron and magnetite, which then cools to room temperature.

Next, consider the cooling of this same melt under conditions of constant oxygen fugacity from a temperature of 1530 °C on the magnetite liquidus where the oxygen fugacity is 10^{-4} bar. As heat is removed, magnetite crystallizes, but the melt cannot change its composition because only this one composition of melt on the magnetite liquidus is stable at an oxygen fugacity of 10^{-4} bar. As more magnetite crystallizes from the melt, the bulk composition changes along the 10^{-4} isobar until it eventually coincides with the magnetite composition, at which point no liquid remains. Note that under these conditions the melt never descends to the peritectic, nor does any wüstite form. Once all of the liquid has disappeared, the temperature again begins to fall. To remain on the 10^{-4} isobar, the magnetite becomes richer in oxygen until at 1080 °C it reaches the solvus with hematite. The temperature then again remains constant while magnetite is converted to hematite. Once all magnetite is consumed, the hematite is then free to continue cooling. In this second example, then, the bulk composition of the system has to change from its initial value of 50% Fe_2O_3 to its final value of 100% Fe_2O_3 in order to maintain the oxygen fugacity at 10^{-4} bar. This change is effected by oxygen being added to the system from an external source.

Oxygen fugacity is also important in silicate systems involving iron, as for example in the system SiO_2–FeO–Fe_2O_3 (Fig. 11.23). The binary diagram of Figure 11.22 forms one of the bounding systems to this ternary. Addition of SiO_2 to this binary introduces the phases tridymite and fayalite. The points of intersection of the lines of constant oxygen fugacity with the liquidus in the binary system (Fig. 11.22) can be extended across the liquidus surface in the ternary system (Fig. 11.23). Note that fayalite melts incongruently to iron plus liquid. Also, the peritectic in the binary system FeO–Fe_2O_3 involving magnetite and wüstite becomes a cotectic after descending a short distance into the ternary system. It is important to keep in mind that this ternary system is part of the larger system Si–Fe–O_2 (inset in Fig. 11.23). When dealing with crystallization at constant oxygen fugacity, oxygen must be added to or removed from the system. For this reason, we must know the location of the oxygen apex in the larger system.

We will again consider crystallization under the two extreme conditions of constant bulk composition and constant oxygen fugacity. Let us first consider the crystallization, under constant bulk composition, of a melt with a composition X that plots in the primary field of fayalite (enlarged inset in Fig. 11.23). Fayalite, on crystallizing, causes the composition of this melt to move directly away from fayalite, thus enriching it in oxygen. This, in turn, causes the oxygen fugacity to increase from 10^{-11} to $10^{-9.4}$ bar by the time the liquid reaches the cotectic with magnetite. Crystallization of magnetite and fayalite drives the liquid down the cotectic to the ternary eutectic with tridymite, where the oxygen fugacity is 10^{-9} bar.

If this same melt is to crystallize at constant oxygen fugacity, it is first necessary to determine in which direction the temperature decreases on the isobars on the liquidus surface of fayalite. This can be done by drawing a line from fayalite to the oxygen apex; this crosses the isobars at temperature maxima. Crystallization of fayalite from melt X under

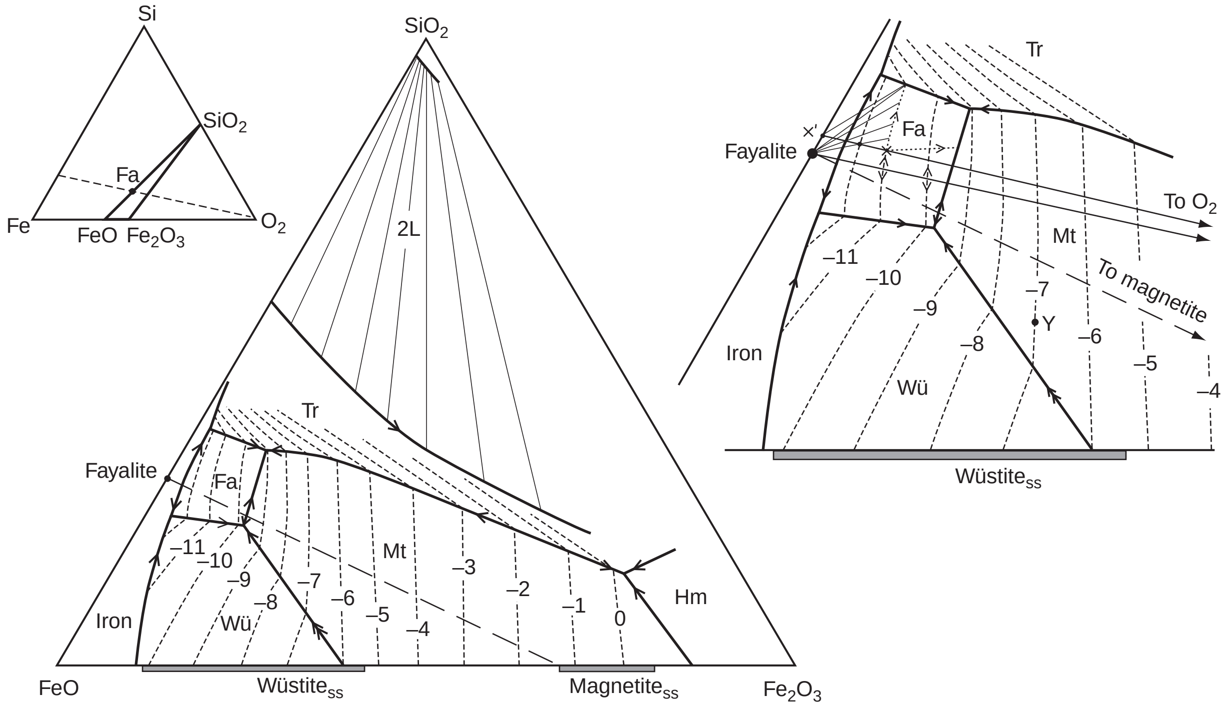


Fig. 11.23 System SiO_2–FeO–Fe_2O_3. Triangular inset indicates position of system in terms of Si–Fe–O_2. (After Muan, 1955.)

an oxygen fugacity of 10^{-11} bar will, therefore, cause the liquid to proceed along the isobar toward the tridymite field. Once this happens, however, a line drawn from the liquid to fayalite no longer passes through the original bulk composition X. The bulk composition, therefore, must change, by either adding or subtracting oxygen. To determine which, a line is drawn through X and the oxygen apex; points along this line have a constant ratio of silicon to iron but variable oxygen contents. The point of intersection of this construction line with the line joining the liquid composition to fayalite is the new bulk composition.

Continued crystallization of fayalite and cooling of the liquid along the 10^{-11} isobar eventually brings the liquid to the cotectic with tridymite. However, as tridymite and fayalite crystallize, the liquid cannot descend the cotectic, for this would increase its oxygen fugacity. Instead, the bulk composition must continue to lose oxygen as it changes its composition along the line extending from X away from the oxygen apex. When the bulk composition reaches the join between fayalite and tridymite (X′), the liquid is totally crystallized, and the system consists only of fayalite and tridymite.

The sequence of crystallization and the compositions of the fractionating liquids are strikingly different in these two modes of crystallization. In nature, then, very different rock types could result from a single parental magma, depending on whether crystallization occurs under buffered or unbuffered conditions. Although conditions approaching both extremes have been identified in natural occurrences, most magmas follow some intermediate path between these extremes.

Because oxygen fugacity plays such an important role in determining what minerals can form in a rock, it is commonly possible to determine from the mineralogy of a rock the oxygen fugacity that existed at the time the rock formed (or at least last equilibrated). A number of mineral assemblages have been used for this purpose but none with more success than that involving coexisting magnetite and ilmenite (Buddington and Lindsley, 1964).

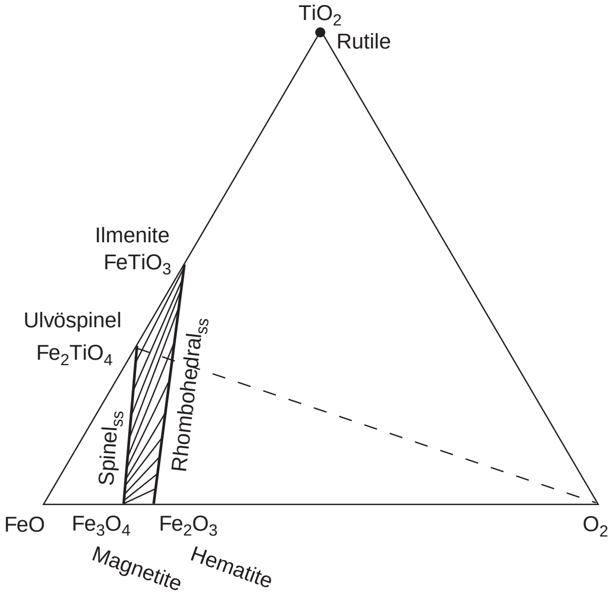


Fig. 11.24 Compositional range of hematite–ilmenite solid solutions (rhombohedral phase) and magnetite–ulvöspinel solid solution (spinel phase) in terms of TiO_2–FeO–O_2. Tie lines indicate compositions of typical coexisting rhombohedral and spinel phases. Dashed line indicates addition of oxygen (increasing fO_2) to ulvöspinel.

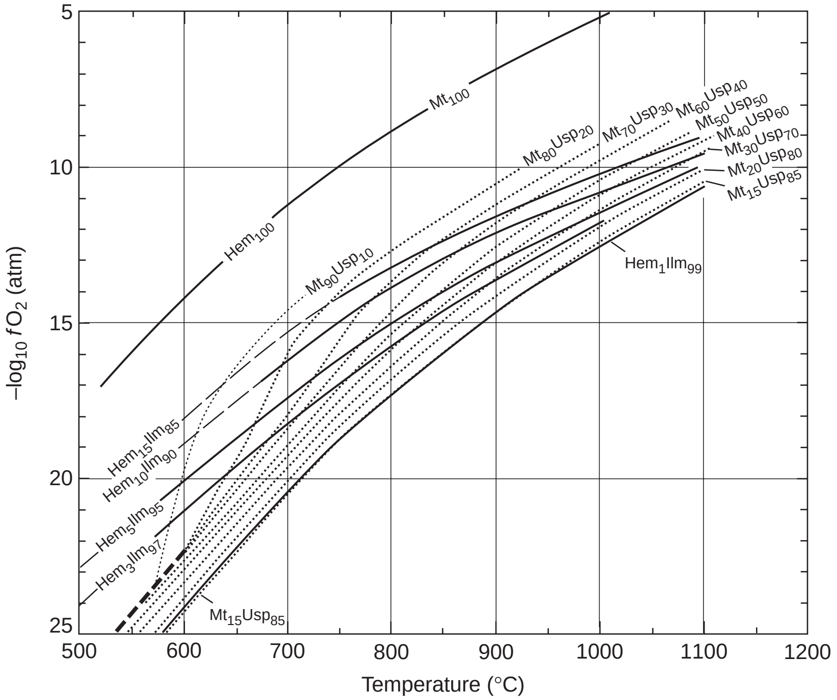


Fig. 11.25 Projection of the compositions of hematite–ilmenite and magnetite–ulvöspinel solid solutions in terms of oxygen fugacity and temperature projected onto the fO_2–T plane. Compositions of phases are in mole percent. Compositions of coexisting rhombohedral and spinel phases uniquely define the oxygen fugacity and temperature at which the phases equilibrated. (After Buddington and Lindsley, 1964.)

Many rocks contain both an ilmenite–hematite solid solution (rhombohedral phase) and a magnetite–ulvöspinel solid solution (spinel phase). These phases plot in the system FeO–TiO_2–O_2 (Fig. 11.24), and they can be related through the following oxidation reactions:

$$\begin{array}{lll} \text{spinel phase} & \text{rhombohedral phase} & \text{spinel phase} \\ 3Fe_2TiO_4 + 0.5O_2 \rightleftharpoons & 3FeTiO_3 & + Fe_3O_4 \\ \text{(ulvöspinel)} & \text{(ilmenite)} & \text{(magnetite)} \end{array} \tag{11.44}$$

$$\begin{array}{ll} 2\,Fe_3O_4 + 0.5O_2 \rightleftharpoons & 3Fe_2O_3 \\ \text{(magnetite)} & \text{(hematite)} \end{array} \tag{11.45}$$

From these reactions and Figure 11.24, it is clear that at low oxygen fugacities, the spinel phase is rich in ulvöspinel and it coexists with an ilmenite-rich rhombohedral phase. As the oxygen fugacity increases (moving toward O_2 in Fig. 11.24) the spinel phase becomes richer in magnetite, and at high oxygen fugacities the spinel phase is rich in magnetite and the rhombohedral phase contains considerable hematite. The compositions of these coexisting phases are also dependent on temperature (total pressure has negligible effect). Thus it is possible, in rocks containing both of these phases, to determine the oxygen fugacity and the temperature at the time the minerals last equilibrated.

To apply this oxygen geobarometer and geothermometer, it is necessary to know the compositions of both the spinel and rhombohedral phases. These would normally be obtained from electron microprobe analyses. The spinel phase is expressed in terms of mole percent magnetite and ulvöspinel and the rhombohedral phase in terms of mole percent hematite and ilmenite. With these compositions, the oxygen fugacity and temperature can be read directly from the graph in Figure 11.25. Over much of the range of conditions the oxygen fugacity can be determined to within one order of magnitude and the temperature to ±30 °C.

While simple in principle, difficulties can be encountered in applying this geothermometer. Both the spinel and rhombohedral phases develop exsolution lamellae on cooling (ilmenite$_{ss}$–hematite$_{ss}$ and ulvöspinel$_{ss}$–magnetite$_{ss}$). In plutonic rocks where cooling is slow, these exsolution lamellae may be expelled from their host grains, which can make determining the original composition of the phases impossible. Many temperatures determined for plutonic rocks are consequently lower than expected. Postmagmatic oxidation can also change phases. For example, the spinel phase is commonly oxidized on cooling to a mixture of magnetite and ilmenite. The ilmenite forms lamellae in the magnetite and resembles an exsolution texture. However, the solid solution between magnetite and ilmenite at high temperatures is negligible. Analyses of magnetite grains containing such lamellae of ilmenite must be recalculated to the original spinel composition (Buddington and Lindsley, 1964).

11.9 PROBLEMS

11.1 Water can dissolve in silicate melts either as molecular water or by combining with oxygen atoms in the melt to form hydroxyl groups by the reaction given in Eq. (11.13). The equilibrium constant for this reaction expresses the ratio between the activities of hydroxyl groups, molecular water, and anhydrous oxygens in the melt. If we assume that these three species mix ideally, the equilibrium constant can be written in terms of mole fractions as shown in Eq. (11.14).

(a) Assuming ideal behavior, calculate from the equilibrium constant the wt% of molecular water and the wt% of hydroxyl in albitic melts with total water contents ranging from 0 to 12 wt% H_2O. The equilibrium constant is 0.17. The mole fractions are defined as follows:

$$X_{H_2O}^{total} = \left(\frac{wt\%_{H_2O}}{18}\right) \div \left(\left(\frac{wt\%_{H_2O}}{18}\right) + \left(\frac{(100 - wt\%_{H_2O})}{32.78}\right)\right)$$

$$X_{H_2O,mol} = \left(\frac{wt\%_{H_2O,mol}}{18}\right) \div \left(\left(\frac{wt\%_{H_2O}}{18}\right) + \left(\frac{(100 - wt\%_{H_2O})}{32.78}\right)\right)$$

Since $X_{H_2O}^{total} = X_{H_2O,mol} + \frac{X_{OH}}{2}$ then

$$X_{OH} = 2\left(X_{H_2O}^{total} - X_{H_2O,mol}\right) X_{O}^{melt} = 1 - X_{H_2O}^{total}$$

The denominator of 18 is the formula weight of water and the denominator of 32.78 is the formula weight of albite written for one oxygen; that is, 262.224/8.

This problem is most easily solved using a spread sheet. Set up columns with the weight percent of total water in the melt (0 – 12), X of total water, X of oxygen in melt, X of molecular water, mole fraction of hydroxyl, and finally, equilibrium constant. The X of molecular water is unknown, so to start with, insert a very small number (e.g. 0.001). The X of hydroxyl is calculated from the X of total water and the X of molecular water (see equation above). In the column for the equilibrium constant, write the equation for the constant based on the mole fractions in the previous columns. Now, using the Solver function in Microsoft Excel, adjust the values of the X of molecular water so that the value of the equilibrium constant becomes 0.17. Because the equilibrium constant has a squared term, Solver can find more than one solution. Be careful to insert a very small number for the initial value of the X of molecular water.

Recalculate the X of molecular water and hydroxyl to wt%, and graph them versus the total wt% water in the melt.

(b) Assuming that molecular water, hydroxyl, and anhydrous oxygen mix ideally is a simplification. A more realistic measure of these components in silicate melts can be obtained using the *VolatileCalc* program. After downloading this program, which runs in Excel, select the option for rhyolitic melts (this approaches closely the composition of albite). Then select *Calculate isobars*, and enter a pressure of 5000 bars; then enter a temperature of 1150 °C. From the output, plot a graph of the wt% of molecular water and wt% hydroxyl versus the wt% total water in the melt.

Compare this graph with the one obtained for ideal behavior in part (a). Using the data presented in Figure 11.1 for MORB as an example of the way in which the total dissolved water is proportioned between molecular water and hydroxyl in silicate melts, which graph (a or b) fits the data best?

11.2 One of the samples used by Moore *et al.* (1998) to calibrate their empirical equation (11.15) expressing the solubility of water in silicate melts as a function of *T*, *P*, and composition is a MORB with the following composition:

SiO_2	50.8
TiO_2	1.84
Al_2O_3	13.7
FeO^t	12.4
MnO	0.22
MgO	6.67
CaO	11.5
Na_2O	2.68
K_2O	0.19
P_2O_5	0.19

Using this composition and Eq. (11.15) calculate the saturation wt% of H_2O at 1200 °C up to a pressure of 0.2 GPa. In converting mole fractions of water from Eq. (11.15) to wt% of water, use a molecular weight for the anhydrous MORB of 62.83 grams. This is different from the way in which molecular weight was calculated in Problem 11.1, which was calculated on the basis of one oxygen. Values for the fugacity of water can be obtained from the *VolatileCalc* program.

How do the results compare with the measured values of the total H_2O in this MORB given in Figure 11.1?

11.3 **(a)** Using the *VolatileCalc* program, calculate the saturation solubility of CO_2 in a basalt and a rhyolite magma at 1200 °C. Assume that both magmas contain no water. Use the *Calculate Saturation Pressure* option and the default compositions for the magmas. Enter a range of values for the CO_2 concentration and graph the resulting saturation pressures for the two magmas.

(b) Repeat part (a) for the rhyolite magma but use a temperature of 800 °C. By comparing your answer with that in part (a), what can you conclude about the effect temperature has on the solubility of a volatile constituent in a magma?

11.4 If a tholeiitic magma that initially is nearly saturated with an immiscible sulfide liquid and has an oxygen fugacity close to that of the QFM buffer rises to a high level in the Earth's crust where its oxygen fugacity is raised to that of the NNO buffer, describe what might happen to the sulfide content of this magma.

11.5 Trace the sequence of events in the complete fractional crystallization of a liquid in the system SiO_2–FeO–FeS having a composition that plots at the top right of the letter F of Fa marking the field of fayalite in Figure 11.4. How does the variation in the sulfide content of the immiscible silicate liquid conform to the general pattern of solubility of sulfur in silicate melts?

11.6 Calculate the mole fractions of albite and water in an albite-saturated melt at 1300 K and 1.0 GPa where the value of $(\Delta G_{\mathrm{mAb}} + P\Delta V_{\mathrm{mAb}})/RT = 1.1$.

11.7 From inspection of Figure 11.6, if the mole fraction of water in an albite melt on the liquidus is 0.2 and the pressure is 0.4 GPa, what is the temperature of the melt and the activity of water in the melt?

11.8 **(a)** From inspection of Figure 11.6, if the vapor phase in a metamorphic rock containing only albite is buffered so that $a_{\mathrm{w}} = 0.4$ and the pressure is 0.8 GPa, at what temperature would the first melt appear, and what mole fraction of water would it contain?

(b) If the magma separates from its source, so that the activity of water is no longer buffered, and the magma rises toward the surface at constant temperature, at what pressure would the melt become water-saturated, and at what pressure would it eventually solidify?

11.9 Create a graph similar to that in Figure 11.18 giving curves for magmas that initially contain 1, 2, 3, 4, 5, 6, 7, and 8 wt% H_2O. Assume that only anhydrous phases crystallize.

11.10 At what degree of solidification would a granitic magma that initially contains 2 wt% H_2O reach a saturation level of 10 wt% H_2O if the crystallizing mineral assemblage consists of 80 wt% quartz and alkali feldspar and 20 wt% muscovite. (The muscovite contains 4.2 wt% H_2O.)

11.11 **(a)** If a magma that initially contains 2 wt% H_2O crystallizes only quartz and alkali feldspar during the first 50% of its crystallization but then crystallizes a mixture of quartz, alkali feldspar, mica, and hornblende that contains an average of 1 wt% H_2O during the final 50% of its crystallization, at what stage of crystallization would the magma reach a saturation level of 10 wt% H_2O?

(b) If no hydrous phases had crystallized from this magma, at what stage of crystallization would the magma have become water-saturated?

11.12 **(a)** A granitic magma initially containing 2 wt% H_2O first crystallizes phenocrysts of biotite containing 3.5 wt% H_2O. Following 20% solidification of the magma, quartz and alkali feldspar join the biotite, and the crystallizing phases now contain only 0.1 wt% H_2O. Trace the concentration of H_2O in the magma as a function of the degree of solidification, and indicate at what stage of solidification the magma would reach a saturation level of 10 wt% H_2O.

(b) By comparing the answers to part (a) and Problem 11.11 discuss the significance of early versus late crystallization of hydrous phases to the development of water-saturated magmas.

11.13 The boundary line between the fields of magnetite and wüstite in Figure 11.23 changes from a reaction to a cotectic as it descends into the ternary system.

(a) What composition does the wüstite have when the reaction relation changes to a cotectic?

(b) From Figure 11.22 determine at what oxygen fugacity wüstite of this composition is in equilibrium with magnetite.

(c) Knowing this oxygen fugacity, determine the point on the magnetite–wüstite boundary in Figure 11.23 where this change from reaction to cotectic occurs.

11.14 Describe the crystallization of a melt of composition *Y* in the enlarged inset of Figure 11.23 under conditions of **(a)** constant bulk composition, and **(b)** constant oxygen fugacity. **(c)** If a magma of composition *Y* were to crystallize fayalite in its residual liquid, what might you conclude about its conditions of crystallization?

11.15 If an andesite contains phenocrysts of magnetite and ilmenite that have compositions of 45 mol% ulvöspinel and 15 mol% hematite, respectively, what were the oxygen fugacity and temperature at the time of eruption?

12 Crystal growth

12.1 INTRODUCTION

With the exception of volcanic glasses and a few rare rock types, all igneous and metamorphic rocks are composed of crystalline mineral grains. The compositions of these grains are readily explained in terms of thermodynamic phase equilibria, but their growth is controlled by kinetic factors that are still poorly understood. The texture of a rock, which bears testimony to the rock's origin, is determined largely by crystal growth. Grain size, which is also determined by crystal growth, affects the rheology of rocks and, thus, plays important roles in determining flow rates in the mantle, which in turn, affect globally so many other geological processes through plate tectonics. Investigating crystal growth in anything other than a vapor or low-temperature aqueous solution at atmospheric pressure is experimentally difficult. A few studies, however, have shed significant light on high-temperature crystal growth of silicates that operate on the timescale typical of cooling lavas (see reviews by Kirkpatrick, 1975; Lofgren, 1980; Cashman, 1990; and Sunagawa, 1992). Growth rates in plutonic igneous rocks and metamorphic rocks, however, are so slow that they will probably always remain outside the reach of the experimentalist. Only through understanding the principles are we likely to gain insight into these slower growth processes (Dowty, 1980; Brandeis *et al*., 1984). Considerably more is now being learned about the growth of crystals through the use of a variety of tools including, in addition to the optical microscope, the electron microscope, X-ray and cathode luminescence topography, and electron and ion beam microanalysis.

The basic principles of crystal growth are equally applicable to igneous and metamorphic rocks – only the medium in which growth occurs is different, so they are both dealt with in this chapter. Crystal growth can occur only after thermodynamic intensive variables exceed, by some finite amount, the equilibrium conditions for the formation of that phase. This overstepping (undercooling, overheating, supersaturating) provides the energy to form a nucleus on which growth can occur and the driving force to sustain that growth. Once formed, a nucleus provides a sink to which the crystallizing components diffuse. The distance over which diffusion transports material depends on the rates of diffusion and the time available. Material that is too far from a nucleus must form a separate nucleus. The fact that rocks rarely have a grain size exceeding a centimeter (pegmatites are an exception), provides an approximate scale over which diffusion operates. The number of nuclei that form depends on the ease of nucleation and the time available for material to diffuse and grow onto these nuclei. The interplay between the rate of nucleation and the rate of crystal growth determines the eventual grain size of the rock.

Crystal growth is a complex process involving numerous steps, each of which proceeds at its own rate, with the slowest one determining the overall growth rate of the crystal. Different minerals growing together may have different rate-controlling steps, as may different faces on a single crystal. Gradients of temperature, composition, and stress affect these rates. A rock's texture is a complicated record of these different rate processes.

12.2 NUCLEATION

In Chapter 10, phase diagrams were used to show what minerals should form from particular compositions at given pressures and temperatures. It was tacitly assumed that, when a liquid, for example, is cooled below the liquidus a mineral crystallizes. At the liquidus itself, however, the ΔG of crystallization is zero, and therefore the rate at which crystals form will also be zero. Even when the temperature drops below the liquidus, experiments show that crystals do not form immediately. Indeed, a significant degree of undercooling is necessary before crystallization begins. Conversely, with solid-state reactions that take place with rising temperatures, a certain degree of overheating is necessary before crystals of the reaction products start forming.

The necessity for overstepping the equilibrium condition is related to energies involved in forming the first minute crystals of the new phase that are known as *nuclei*. Fluctuations in the composition of the reactants momentarily bring about clusters of the components necessary to form the new phase. Most of these clusters are unstable, as will be shown below, and they disappear. Only those reaching a critical size survive and grow. Nuclei may form randomly in the host from which they are growing, in which case nucleation is said to be *homogeneous*, but if it takes place on an already existing surface, such as another crystal or a gas bubble, it is said to be *heterogeneous*. Olivine crystals in a slowly cooling magma, for example, may nucleate homogeneously, but the tourmaline crystals in the pegmatite illustrated in Figure 11.15 clearly nucleated heterogeneously on the roof and floor of the pegmatite. Nucleation of metamorphic minerals is likely to occur heterogeneously, and even in magmas, it may be the dominant mode of nucleation.

The reason that small clusters of the new phase have difficulty in surviving is related to the development of an interface between the embryo and the medium in which it is growing. Any surface between two phases involves an interfacial energy, the magnitude of which depends on the degree of mismatch of structures across the boundary. When a crystal is large, this surface energy constitutes only a small fraction of the total free energy of the crystal and can be ignored, but when the crystal is very small, the surface energy is a significant fraction.

Let us consider all of the energies involved in the growth of a new crystal. First, there is the free-energy change involved in forming the new phase, independent of surface energies. This is the free-energy change with which we are already familiar and which can be obtained from tables of thermodynamic data. We will refer to this as the free energy change of the volume, ΔG_{vol}. Second, there is the free energy associated with the creation of a new interface (ΔG_{intf}). Third, the growth of a crystal may involve the development or release of strain (ΔG_{strain}). A crystal of ice growing in a milk bottle left out on a winter's day is capable of cracking the bottle. Clearly, the crystallization of this ice would require a greater driving force than would ice growing in an unconfined space. In other cases, the medium from which a crystal grows may be strained, and the growth of the crystal releases the strain energy. This certainly occurs when a structurally deformed rock recrystallizes.

For a nucleus to grow, the overall free-energy change associated with its formation must be negative. This free-energy change can be expressed as

$$\Delta G_{\text{growth}} = \underset{(-)}{\Delta G_{\text{vol}}} + \underset{(+)}{\Delta G_{\text{intf}}} + \underset{(+\text{ or }-)}{\Delta G_{\text{strain}}} \tag{12.1}$$

The main driving force for growth of the crystal comes from ΔG_{vol}, which becomes increasingly negative the more the equilibrium conditions are overstepped. The free-energy change associated with the creation of a new interface is always positive and therefore acts to prevent crystallization. The strain energy either inhibits or enhances growth, depending on whether it is positive or negative. For ΔG_{growth} of a nucleus to be negative, the equilibrium conditions must be overstepped sufficiently to make $-\Delta G_{\text{vol}}$ greater than $\Delta G_{\text{intf}} + \Delta G_{\text{strain}}$. This can be done, for example, in the case of nucleation of a crystal in magma by supercooling the magma. In a metamorphic rock, it might be necessary to superheat the rock or deform the rock (thus making ΔG_{strain} negative). In both igneous and metamorphic environments, ΔG_{vol} can also be changed by changing the composition of the host to cause supersaturation.

Let us, for purposes of illustration, consider the homogeneous nucleation of a single olivine crystal in a basaltic magma at constant temperature, pressure, and composition of magma. Under these conditions, the value of ΔG_{vol} will be (crystal volume$/\bar{V}$) $\times\ \Delta\bar{G}_{\text{vol}}$ where $\Delta\bar{G}_{\text{vol}}$ is the molar free energy for the formation of the olivine and $\bar{V}$ is its molar volume. Under the conditions considered here, $\Delta\bar{G}_{vol}$ is fixed, and because the crystal is growing in a liquid, ΔG_{strain} is zero. If we assume that the embryonic cluster of olivine is essentially spherical, the free-energy change associated with the formation of an embryo of radius r will be

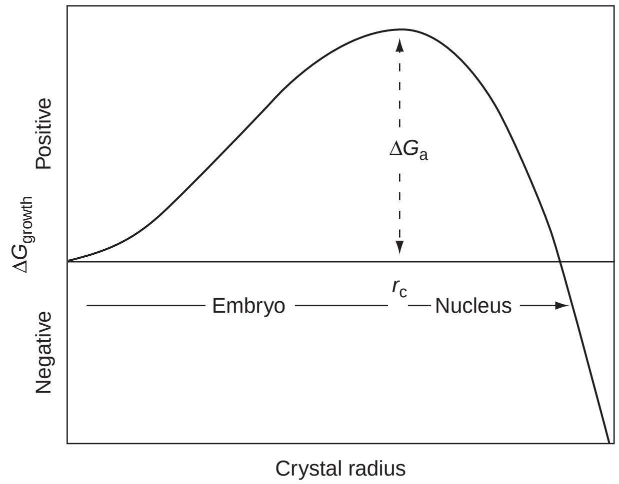


Fig. 12.1 Relation of free energy of growth of an embryo to its radius. The embryo forms a stable nucleus once it reaches a critical radius, r_{c}. Increase in radius beyond r_{c} lowers free energy. ΔG_{a} is the free energy of activation needed to stabilize a nucleus.

$$\Delta G_{\text{growth}} = \overset{\text{volume}}{\left(\frac{4}{3}\pi r^3/\bar{V}\right) \times \Delta\bar{G}_{\text{vol}}} + \overset{\text{interface}}{4\pi r^2 \Delta G_{\text{intf}}} \tag{12.2}$$

Recall that $\Delta\bar{G}_{\text{vol}}$ is negative and ΔG_{intf} is positive. The functional form of Eq. (12.2) is shown in Figure 12.1. For very small values of r, an increase in radius causes an increase in the free energy of growth; this, therefore, is an unnatural process and the embryo dissolves. Above a critical value of the radius, r_{c}, increasing r decreases the free energy of growth, and thus a stable nucleus is formed that will grow. The critical radius occurs where the ΔG_{growth} passes through a maximum. Thus, by setting the derivative of Eq. (12.2) with respect to r equal to zero, we can show that the *critical radius* is given by $-2\Delta G_{\text{intf}}\,\bar{V}/\Delta\bar{G}_{\text{vol}}$ (see Problem 12.1).

The value of ΔG_{growth} at the critical radius is commonly referred to as the activation energy, because it is the energy required to create a nucleus that is large enough to grow (see Section 5.6 and Fig. 5.14 for discussion of activation energy). By substituting the value for the critical radius into Eq. (12.2), we find that the activation energy is

$$\Delta G_a = \frac{16\,\pi \Delta G_{\text{intf}}^3\,\bar{V}^2}{3\Delta\bar{G}_{\text{vol}}^2} \tag{12.3}$$

Because nucleation is an activated process, its rate (R_0) might be expected to obey an *Arrhenius*-type relation (see Eq. (5.53)) of the form

$$R_0 = K \exp\left(\frac{-\Delta G_{\text{a}}}{RT}\right) \tag{12.4}$$

Nucleation rates in magmas are strongly dependent on the degree of undercooling. The rate at which critical nuclei form in a cooling magma depends strongly on the degree of undercooling. As the temperature falls below the liquidus, the increasing negative value of ΔG_{vol} increases the likelihood that nucleation will occur. Falling temperature, however, makes diffusion more difficult and decreases the chances of embryos forming with critical radius. The nucleation rate at a given temperature therefore depends on which of these effects dominates.

We know from Eq. (7.33) that $(\partial \Delta G/\partial T)_P = -\Delta S$, and $\Delta H_{\mathrm{m}}/T_{\mathrm{m}}$ can be substituted as an approximation for ΔS, where ΔH_{m} is the enthalpy of melting and T_{m} the melting point of the pure mineral. Upon integrating Eq. (7.33), we obtain

$$\Delta G_{\mathrm{vol}} = \frac{\Delta H_{\mathrm{m}}(T_0 - T)}{T_0} \tag{12.5}$$

which indicates that ΔG_{vol} is a linear function of the degree of undercooling ($T_0 - T$). Diffusion, on the other hand, changes exponentially with temperature (Eq. (5.53)). With falling temperature, then, the increased negative value of ΔG_{vol} at first causes the nucleation rate to increase, but with continued lowering of temperature, the diffusion term comes to dominate and the nucleation rate passes through a maximum and then decreases to zero, at which point glass would result (Fig. 12.2). Curves of this type have been experimentally determined, for example, for quartz, alkali feldspar, and plagioclase nucleating in granitic and granodioritic melts. The maximum nucleation rates for these crystals occurred at 200 to 400 °C of undercooling (Swanson, 1977).

Dikes provide a natural occurrence where the degree of undercooling would be expected to increase toward the margins because of increased cooling rates. The number of nuclei would therefore be expected to increase toward the margins until the maximum nucleation rate is reached and then decrease again as higher degrees of supercooling make it more difficult for nuclei to form and the melt cools to a glass. Such a variation is illustrated in Figure 12.13.

Crystal nucleation rates for metamorphic reactions that take place on heating behave quite differently from magmatic ones because, while the rising temperature makes the value of ΔG_{vol} more negative, the rising temperature also increases diffusion rates. Thus the nucleation rate steadily increases with degree of overstepping of the equilibrium conditions. More nuclei of a mineral are therefore likely to form in a contact metamorphic rock near the igneous intrusion where temperatures are high, rather than farther away where they are low. Because grain size is inversely related to the number of nuclei formed, the grain size would be expected to increase away from the intrusion (other factors being equal) to a distance where the temperature had never risen above that necessary for nucleation. Figure 12.3 illustrates such a variation in the contact metamorphic zone at the margin of a basaltic dike from County Mayo, Ireland. Grossular garnets formed in the limestone country rock by heat liberated from the dike show a steady decrease in number and increase in size away from the contact.

With metamorphic nucleation, strain energies must also be taken into account. The most common case involves rocks that were deformed prior to heating, so that when new minerals grow or the original grains recrystallize strain energy is released. From Eq. (12.1) it is clear that with slow heating, the most strained parts of a rock will be the regions where $\Delta G_{\mathrm{growth}}$ will first become negative; that is, nucleation will occur first where strain is greatest. Moreover, the greater the

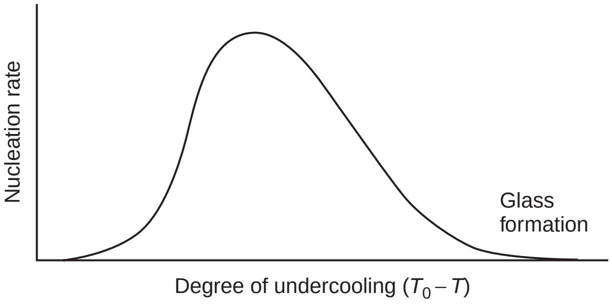


Fig. 12.2 Variation in rate of nucleation in a melt as a function of degree of undercooling.

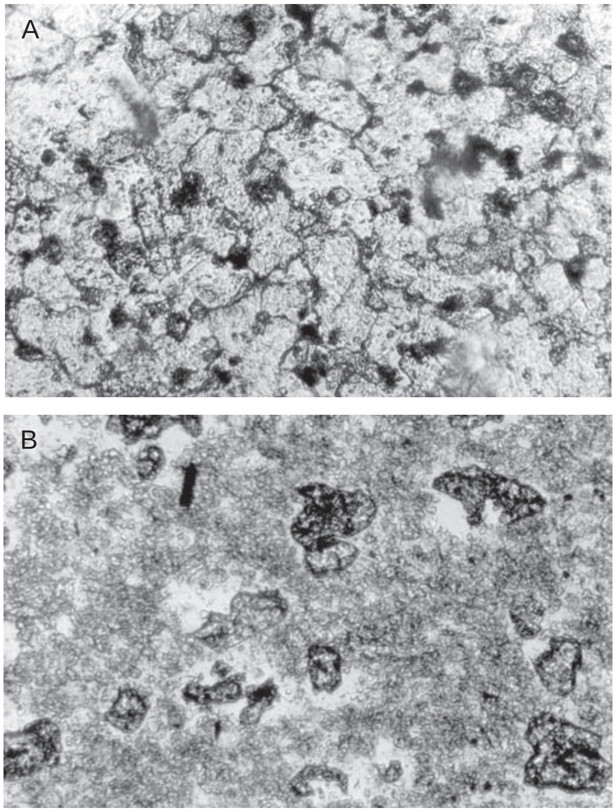


Fig. 12.3 Variation in the abundance and size of grossular garnet crystals in metamorphosed limestone near a diabase dike, County Mayo, Ireland. Near the contact [(**A**) 0.3 m from contact] garnet crystals (high relief and dark) are small but present in large numbers because of a high nucleation rate. Farther from the contact [(**B**) 2.8 m from contact] garnets are larger but fewer in number. The width of each field is 0.7 mm.

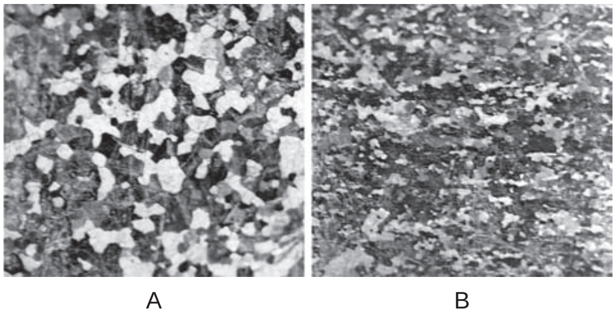


Fig. 12.4 Annealed aluminum metal. Both specimens were first deformed by being passed through a roller, but the spacing between the rollers for specimen B was half that for specimen A. The greater deformation in specimen B created more nucleation sites when the samples were recrystallized by heating to high temperature. Specimen B is therefore finer grained. Width of each field is 2 cm.

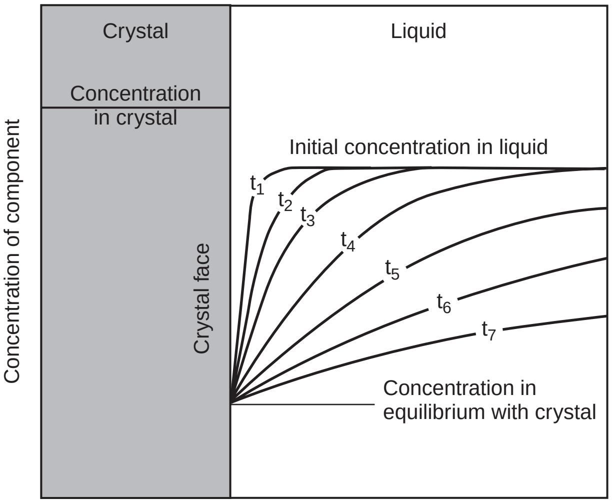


Fig. 12.5 Concentration gradients of a component at successive times (t_i) in advance of a crystal face whose growth rate is determined by diffusion of nutrients through the surroundings. The concentration of the component immediately in contact with the crystal face is the equilibrium concentration. The concentration in the crystal is shown as being constant, but it may vary during growth (see Fig. 12.15).

strain in a rock, the greater will be the number of nuclei to form, and consequently, the finer grained the rock will be.

The effect of strain on nucleation can be illustrated with a metallurgical example. Figure 12.4 shows two pieces of aluminum metal that came originally from the same bar. Both were flattened by being passed through a roller, but specimen B was flattened to half the thickness of specimen A by moving the rollers closer together. Specimen B was therefore more strained than specimen A. Both specimens were then held at high temperature in a furnace where they recrystallized (annealed). Because of the greater strain in specimen B, more nuclei formed and the annealed metal is much finer grained than the less strained specimen.

The same effects are to be expected in metamorphic rocks, with more deformed rocks recrystallizing to a finer grain size. Indeed, one of the finest-grained metamorphic rocks, known as *mylonite*, is formed in fault zones where strain is at a maximum. The grain size of these rocks was originally interpreted to result from grinding and milling in the fault zone (hence the name – *mylos* is Greek for mill). Although some comminution may occur in this way, the fine grain size is now interpreted to be due largely to crystallization at many nucleation sites.

12.3 CRYSTAL GROWTH RATES

Once a stable nucleus has formed it continues to grow and forms a crystal whose size is determined by the concentration of nutrients in the surroundings and the proximity of neighboring nuclei. Four steps are involved in the growth process. First, nutrients must diffuse to the nucleus through the medium in which the crystal is growing; diffusion rates vary considerably depending on whether the medium is a gas, fluid, silicate melt, or solid rock. Second, nutrients, on arriving at the nucleus, may have to react and arrange themselves into building units that are acceptable to the crystal. Third, building units must then attach themselves to the crystal surface; this may involve nucleation of new surfaces or the growth of dislocations. Finally, attachment of the building units produces heat of crystallization and an increase in concentration of components not entering the crystal. Both of these must be dissipated before further growth can occur. The rate at which a crystal grows is determined by the slowest of these steps. We will now examine the time dependency of each of these processes, assuming that the temperature and concentration of nutrients in the environment at some distance from the crystal remain constant during the period of interest.

12.3.1 Diffusion-controlled growth

If diffusion controls the rate at which a crystal face grows, it simply means that this is the slowest of the chain of steps involved in the growth of that face. The other steps still occur, but because they take place more rapidly they have to wait for diffusion to bring nutrients into the chain. Diffusion, then, is described as being the rate-determining process.

The distance a face advances, say in the x direction, in a given time depends on the flux of nutrients brought to that face by diffusion. When the crystal first starts to grow the region immediately in contact with the face becomes depleted in nutrients, and a steep concentration gradient develops. With time, however, the gradient becomes shallower (Fig. 12.5), the flux of nutrients brought to the crystal face decreases, and the crystal grows more slowly.

The relation between the diffusional flux and the concentration gradient is given by Fick's law (Eq. (5.50)). Because the flux of material reaching the crystal decreases as the concentration gradient decreases, the growth rate is given by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = K\left(\frac{D}{t}\right)^{1/2} \tag{12.6}$$

where K is a constant (see discussion of Eq. (5.60)). Integration of this equation gives t in terms of x ($t = x^2/4K^2D$), which, when substituted back into Eq. (12.6), gives

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{k_D}{x} \tag{12.7}$$

where k_D is a constant for the diffusion process that combines all of the other constant terms. This equation expresses the growth rate in terms of a property that we can observe in a rock – grain size. Equation (12.7) indicates that, for diffusion-controlled growth, the larger a crystal becomes, the slower it grows.

12.3.2 Phase boundary reaction-controlled growth

In the growth of complex silicates, addition of atoms one at a time may not be possible. For example, local charge balance may require that certain groups of ions be attached simultaneously. If the reaction to form these building units determines the rate of crystal growth, there is no reason for the growth rate to change with time. The growth rate would therefore have the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k_R \tag{12.8}$$

where k_R is a constant for the reaction process.

12.3.3 Surface nucleation-controlled growth

Once nutrients are organized into acceptable building units, they must attach themselves to the surface of the crystal. This happens in a number of different ways (Fig. 12.6(A)). For example, on a perfectly planar surface a building unit could attach itself only by starting a new surface layer. Such attachment, however, satisfies only a few bonds and actually increases the proportion of surface area to volume. Steps on the surface present more favorable sites for attachment and corners are still better. As a result, precipitation on a face is likely to fill in and complete all irregularities before a new surface layer will nucleate. This is why crystals that are able to grow freely in a medium, such as phenocrysts in a magma or some porphyroblasts in metamorphic rocks, tend to be bounded by well-developed crystal faces; they are said to be *euhedral*. Attachment of the first building block to a new layer is a statistical process – the larger the face, the more chance there is that an attachment will occur somewhere on that face. We can conclude therefore that the growth rate for surface nucleation-controlled growth must be proportional to the surface area of the crystal. If all of the faces on the crystal are growing at the same rate, the x direction can be thought of as the radius of the crystal, and then the area of the growing faces would be approximately $4\pi x^2$ (assuming a sphere). The growth rate could then be expressed as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k_S x^2 \tag{12.9}$$

where k_S is a constant for surface nucleation-controlled growth.

12.3.4 Screw dislocation-controlled growth

Attachment of nutrients to the crystal may also occur on screw dislocations. Dislocations are imperfections in crystal structures, and the type known as a *screw dislocation* (Fig. 12.6(B)) is of particular interest in crystal growth, because the step it produces on the growing surface is self-perpetuating, with each new building block being added to form a spiral ramp. These steps are identical to the steps on a spiral staircase. A number of these dislocations may exist on a single face. The rate at which these dislocations grow, if they are the rate-determining step, is independent of time and the size of the crystal but will depend only in some manner on the concentration of nutrients. The growth rate determined by this mechanism is therefore a constant

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k_{SD} \tag{12.10}$$

where k_{SD} is a constant for screw dislocation-controlled growth.

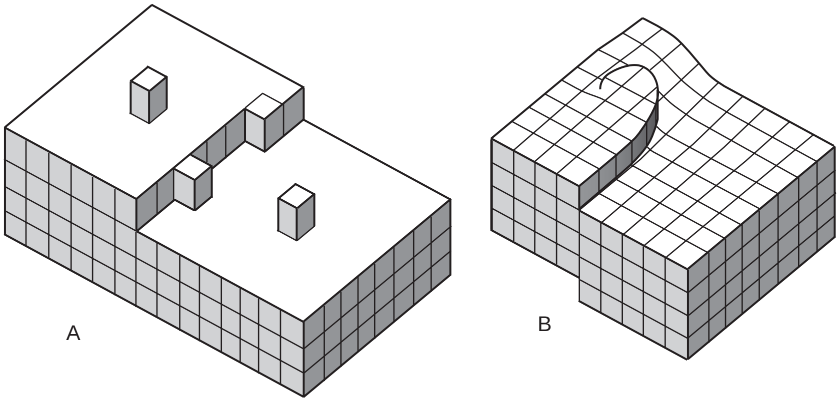


Fig. 12.6 (**A**) Possible attachment sites on the surface of a crystal. Sites that provide the largest number of bonds are likely to be occupied first (i.e. corners > edges > faces). (**B**) The edge produced by a screw dislocation is self-perpetuating, as growth develops a spiral ramp.

12.3.5 Dissipation of heat of crystallization and impurities-controlled growth

When nutrients transfer from a melt onto a crystal face the latent heat of crystallization causes the temperature of the melt on the face to increase. At the same time, components in the melt that do not enter the crystal become concentrated in the melt at the crystal face and, according to the cryoscopic equation (Eq. (10.6)), lower the liquidus temperature, which in turn would decrease the degree of supersaturation. Both heat and material must therefore diffuse away from the crystal surface before growth can continue. When this process determines the growth rate of the crystal, any irregularity that protrudes beyond the general surface of the crystal will extend into cooler and more supersaturated melt and will grow more rapidly. A smooth crystal face is therefore unstable. Instead, the crystal develops a *dendritic* or *skeletal* form. Such habit is common in komatiitic lavas (Fig. 12.7(A)) and basaltic pillow lavas but also in silica-rich residual patches of melt in more slowly cooled rocks where magnetite, ilmenite, and sphene commonly grow with skeletal form (Fig. 12.7(B)).

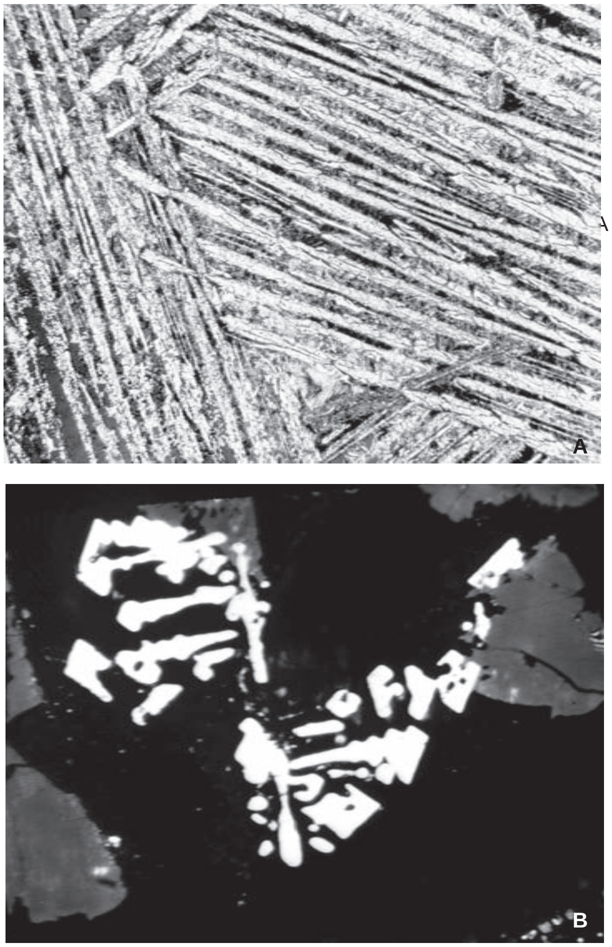


Fig. 12.7 Dendritic crystals of (**A**) olivine in komatiite from Monroe Township, Ontario (width of field, 1.3 cm), and (**B**) magnetite in the silica-rich residue of a diabase dike, New Haven, Connecticut (width of field, 0.33 mm).

Dendritic growth of a crystal cannot be described in terms of the simple advance of a crystal face because of the complex branching nature of the crystal-melt boundary. However, the rate of advance of the tips of dendrites would be expected to remain constant, because these parts of the crystals are always growing into melt of constant composition and temperature.

Growth rates of crystals from melts, like nucleation, are strongly dependent on the degree of undercooling (compare Figs. 12.2 and 12.8). As the temperature drops below the liquidus, growth rates increase rapidly to a maximum and then decrease slowly at higher degrees of undercooling. The position of the maximum varies for different minerals and for different composition melts (Dowty, 1980). For example, anorthite in a melt of anorthite composition has a maximum growth rate at ~200 °C of undercooling (Fig. 12.8), whereas plagioclase of An_{48} composition has a maximum at 50 °C of undercooling. In contrast, intermediate plagioclase growing in the system diopside–anorthite–albite has a maximum growth rate at ~140 °C of undercooling. Olivine and pyroxenes growing in liquids of their own composition require only 20 to 30 °C of undercooling to reach a maximum growth rate. Diopside growing in the system diopside–anorthite–albite, however requires 60 to 100 °C of undercooling to reach a maximum growth rate. Quartz, plagioclase and alkali feldspar in granitic and granodioritic melts have maximum growth rates at 100 to 200 °C of undercooling.

The maximum growth rates vary with mineral and composition of melt. Growth rates of crystals in their own melt tend to be high because building components do not have to travel through the melt, and liquidus temperatures are high. Growth rates of crystals in complex melts, such as magma, are lower because of multicomponent diffusion through the

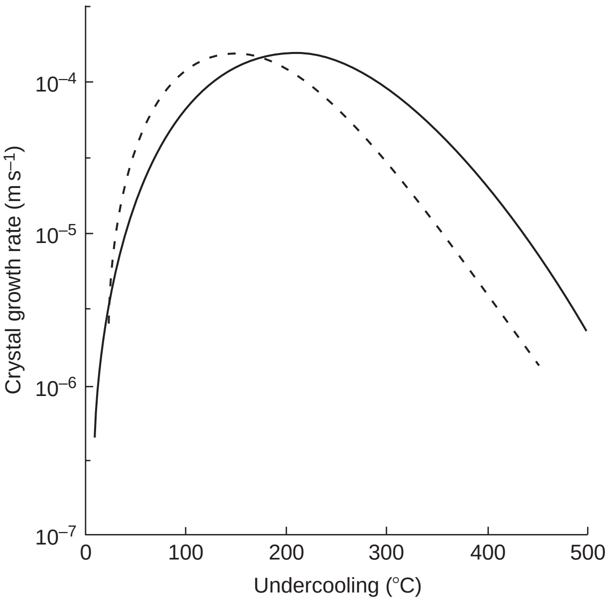


Fig. 12.8 Growth rate of anorthite crystals in anorthite melt as a function of undercooling. Solid line is from Kirkpatrick *et al.* (1979) and the dashed line from Abe *et al.* (1991).

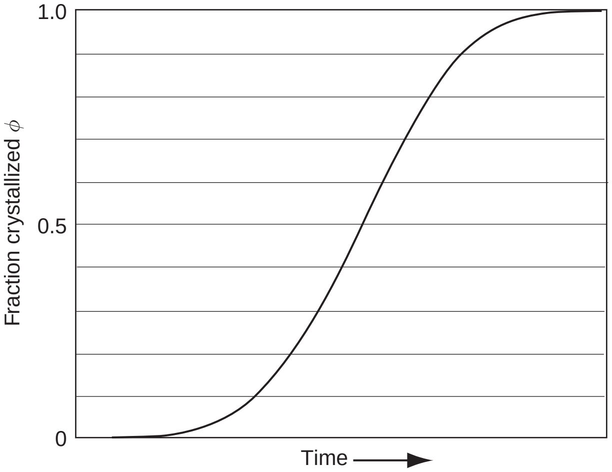


Fig. 12.9 Sigmoidal crystallization path for a one-component system crystallizing at a constant temperature with homogeneous nucleation according to the Avrami equation (see text for discussion). Magmas are thought to follow a similar path.

melt and generally lower liquidus temperatures. Maximum growth rates of plagioclase and diopside in the system diopside–anorthite–albite are ~10^{-7} and ~10^{-5} m s^{-1} respectively (Dowty, 1980). The maximum growth rate of quartz and alkali feldspar in granitic magma is ~10^{-10} m s^{-1} (Swanson, 1977).

The transformation of liquid into a crystalline solid is often described by the *Avrami equation*, which combines the effects of both nucleation and crystal growth into a single expression (Kirkpatrick, 1981; Marsh, 1998). This equation applies to one-component systems solidifying at constant temperature with homogeneous nucleation. The fraction crystallized (ϕ) after any time t is given by

$$\phi = 1 - \exp(-Kt^m) \tag{12.11}$$

where K is a function of the crystal shape, and the nucleation and growth rates, and m describes the time dependence of nucleation and growth (Cashman, 1990). The resulting crystallization graph has a characteristic sigmoidal shape (Fig. 12.9). Although the equation applies to only one-component melts, more complex systems, such as magmas, are believed to follow similar crystallization paths (Marsh, 1998).

12.4 CRYSTAL MORPHOLOGY DETERMINED BY RATE-DETERMINING GROWTH PROCESSES

Experiments by Lofgren (1974) on plagioclase and by Kirkpatrick (1974) on plagioclase, pyroxenes, and melilite have shown that at small degrees of undercooling of the melt large euhedral crystals are formed that resemble phenocrysts in igneous rocks. At greater degrees of undercooling, crystals tend to grow with skeletal form and are more *acicular*; that is, their length to breadth ratio increases. At still greater degrees of undercooling, crystals have a branching, *dendritic* form. Finally, at the greatest degrees of undercooling, the liquid passes the glass-transition temperature (Calas *et al*., 2006) and radiating crystalline fibers form a *spherulitic* texture.

These changes are interpreted to result from a change in the rate-determining step from surface nucleation-controlled growth at small degrees of undercooling to growth controlled by dissipation of impurities at high degrees of undercooling. At the lowest temperatures, growth rates are slow because of the low diffusion rates. As a result, the layer of liquid on the crystal face that is enriched in those components not entering the crystal is thin, and small wavelength perturbations are large enough to penetrate it. This results in the crystal face advancing as a series of closely spaced fibers. At higher temperatures, and consequently smaller degrees of undercooling, the thickness of the zone enriched in components not entering the crystal increases. Larger perturbations are therefore necessary to penetrate this layer, thus explaining the dendritic morphology and at still higher temperatures the skeletal forms. At the smallest degrees of undercooling, where surface nucleation becomes the rate-determining process, diffusion has sufficient time to dissipate unwanted material, and the crystal faces grow as planar surfaces. Textural variations (Fig. 12.10) similar to those obtained in the experiments can be found in pillows (Bryan, 1972).

The cooling rates and degrees of undercooling in the experiments and in the pillows are very large compared with those that must occur in large plutonic bodies. Could different rate-determining steps operate if the degree of undercooling were only a few degrees instead of tens of degrees? While experiments at very small degrees of undercooling are impractical, evidence of the rate-determining steps under these conditions has been obtained from rocks themselves.

Large diabase dikes are sufficiently simple bodies that their cooling histories can be interpreted with some degree of confidence (Chapter 5). Near their centers, magma may take many years to cool through the interval over which crystal nucleation and growth occur. Dikes, therefore, have the potential of providing data for time spans that are much longer than those achievable in laboratory experiments.

In studying the textural variations across wide, flash-injected diabase dikes, Gray (1978) found that the total number of nuclei of a single mineral (actually, the number of crystals) in a unit volume of diabase, N, is related to the distance from the contact, x, by

$$N \propto x^n \tag{12.12}$$

where n is a constant that depends, in part, on the growth-controlling mechanism during nucleation. Thus, from this simple relation it may be possible to determine the growth-controlling mechanism at the much smaller degrees of undercooling that existed in the centers of these dikes.

On cooling below the liquidus of a mineral, a magma becomes supersaturated in that phase, as long as no nuclei form. The degree of supersaturation, s, increases with the amount of undercooling, ΔT (Fig. 12.11). Once nucleation begins, however, the degree of supersaturation decreases until it is just sufficient to drive the growth of the crystals.

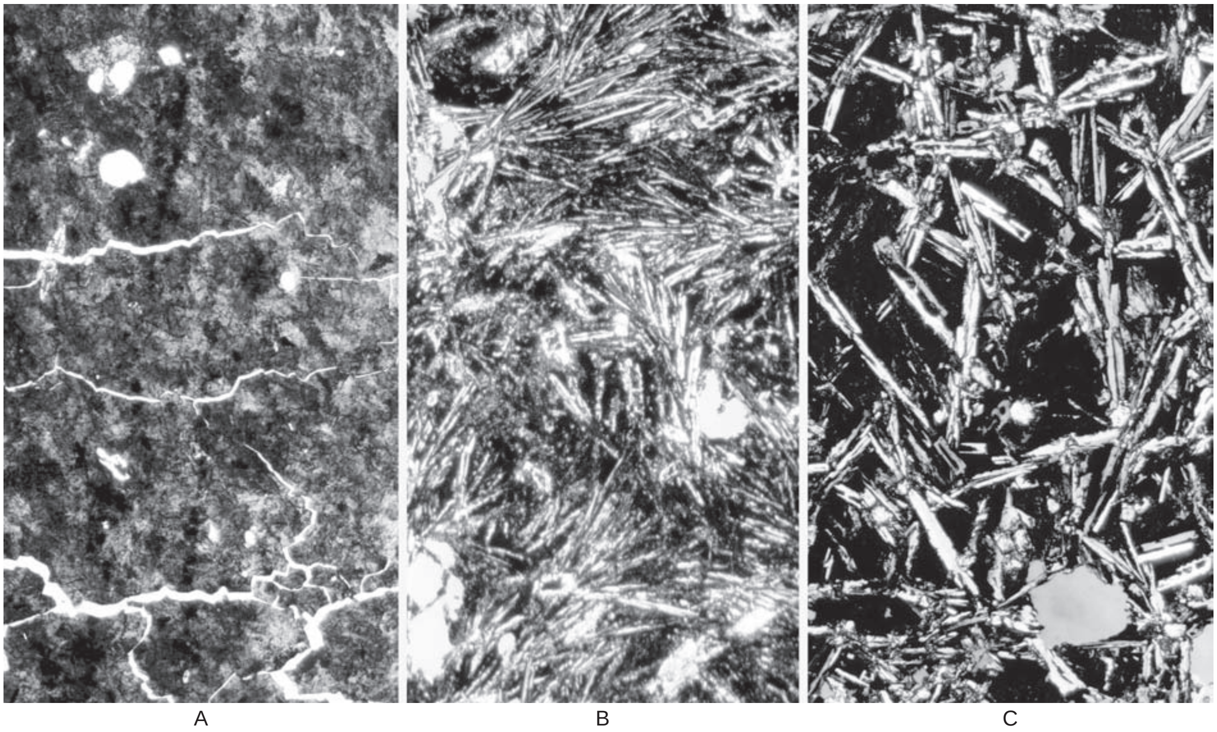
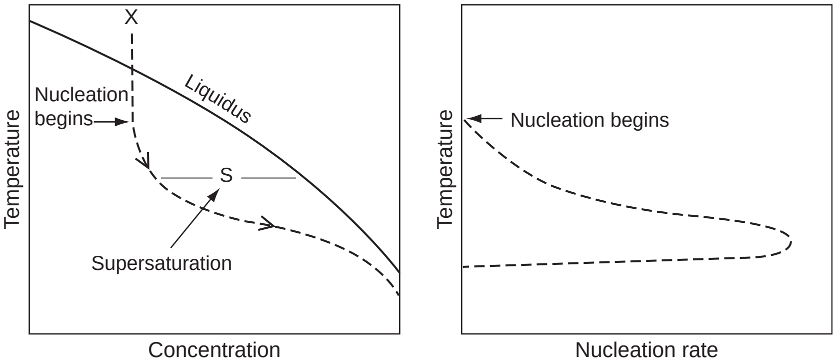


Fig. 12.10 Textural variation from glassy selvage (left) to core (right) of basaltic pillow from Mid-Atlantic Ridge. Near the margin of the pillow (**A**), feldspar crystallized from glass to produce patches of slightly birefringent material. Cracks can be seen entering glassy selvage from left. With slower cooling (**B**), radiating and branching fibrous plagioclase crystals formed. In the center of the pillow where cooling was slowest (**C**), crystals have planar faces, but some skeletal forms are still present. The width of each field is 1.0 mm.

Fig. 12.11 Temperature–composition diagram showing the path followed by a magma of initial composition X that cools below the liquidus of a mineral and becomes progressively more supersaturated (S) until nucleation occurs. Also shown is the rate of nucleation as a function of undercooling.



In relatively slowly cooled melts where nucleation occurs over a small temperature interval, the rate of nucleation, dN/dt, is found to be relatively insensitive to temperature changes and be very nearly proportional to the degree of supersaturation raised to a power m, which is related to the number of molecules forming the critical nucleus (typically, between 2 and 8) (Gray, 1978); that is,

$$\frac{\mathrm{d}N}{\mathrm{d}t} = ks^m \tag{12.13}$$

Thus, as the degree of supersaturation of the magma increases with undercooling, the greater will be the rate of nucleation. But once nuclei form and decrease the supersaturation, the rate of nucleation decreases to zero (Fig. 12.11). The total number of nuclei formed during this interval determines the number of crystals per unit volume of rock that are eventually formed and hence the mean grain size of the rock. The actual number of nuclei formed can be determined by integrating Eq. (12.13).

The variation in the supersaturation must be known as a function of time before Eq. (12.13) can be integrated. Supersaturation depends on the degree of undercooling, the slope of the liquidus (Fig. 12.11), and the amount of material removed from the melt by crystallization, which in turn depends on the crystal growth mechanism. For a flash-injected diabase dike (no feeder flow), Gray (1978) showed that integration of Eq. (12.13) leads to Eq. (12.12). Figure 12.12 shows plots of Eq. (12.12) for three different growth-controlling mechanisms – dislocation-, diffusion-, and phase boundary reaction-controlled growth – when the value of m in Eq. (12.13) is chosen to be 3.2. The slopes of these three lines correspond to values of n in Eq. (12.12) of -0.9, -1.5, and -2.2, respectively. If a different value of m had been chosen, the slopes of the lines in Figure 12.12 would have been different for the three rate-determining processes (different n). Unfortunately, the value of m cannot be determined from field data. However, the slopes in Figure 12.12 are the ones that best fit the data for minerals from diabase dikes.

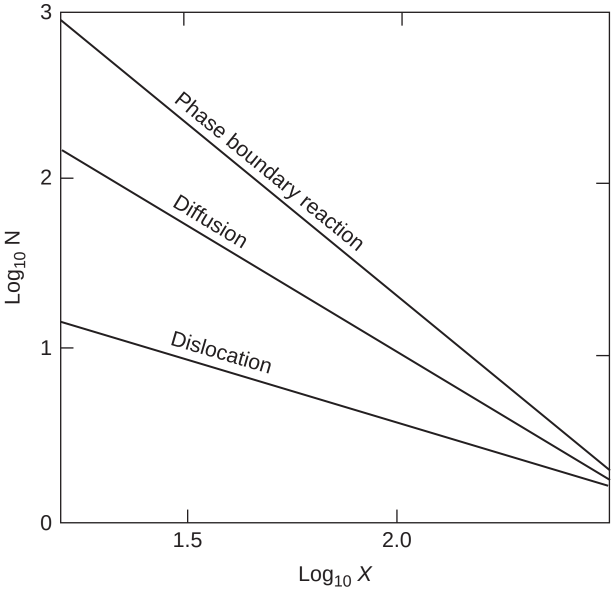


Fig. 12.12 Plot of the logarithm of the number of crystals of a mineral per unit volume of diabase versus the logarithm of the distance into a dike from the contact. Assuming the value of m in Eq. (12.12) is 3.2, the three lines indicate the variations to be expected with the three growth-controlling mechanisms – phase boundary reaction, diffusion, and dislocation. (After Gray, 1978.)

Although the growth-controlling mechanism during nucleation cannot be determined with certainty from the slopes of lines in log N versus log x plots, the fact that such plots give straight lines is evidence for the operation of a single rate-determining process. Moreover, if a line has a break in slope, the two segments of the line must reflect different growth-controlling mechanisms. Similarly, if two minerals have different slopes, they must have had different growth-controlling mechanisms. Clinopyroxene, for example, typically has a significantly more negative value of n than does coexisting plagioclase.

The main textural variation in diabase dikes is determined by the ratio of the number of plagioclase to clinopyroxene crystals per unit volume of rock. If this ratio is less than 1, the texture is *intergranular*; if it is greater than 1, the texture is *ophitic*; and if it equals 1, the texture is *subophitic* (Fig. 12.13). The number of plagioclase crystals in a volume of rock at a given distance from the contact is given by $N_{\mathrm{Plag}}x^{n\mathrm{Plag}}$. Similarly, the number of clinopyroxene crystals is given by $N_{\mathrm{Cpx}}x^{n\mathrm{Cpx}}$. The ratio of the number of plagioclase to clinopyroxene crystals, then, can be expressed as a function of distance from the contact by

$$\frac{N_{\mathrm{Plag}}}{N_{\mathrm{Cpx}}} \propto x^{(n\mathrm{Plag}-n\mathrm{Cpx})} \tag{12.14}$$

Because the value of n for clinopyroxene is typically more negative than that for plagioclase, the value of $N_{\mathrm{Plag}}/\mathrm{N}_{\mathrm{Cpx}}$ will increase into the dike. Thus a dike with an intergranular texture ($N_{\mathrm{Plag}}/\mathrm{N}_{\mathrm{Cpx}} < 1$) near the contact will have an ophitic texture toward the center of the dike as long as the dike is sufficiently wide.

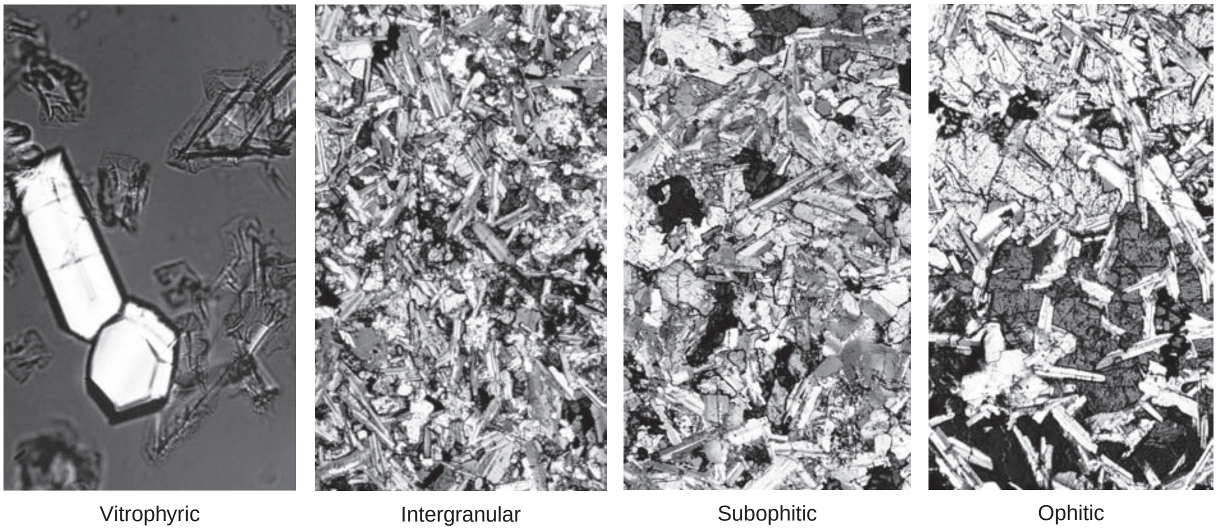

Vitrophyric
Intergranular
Subophitic
Ophitic

Fig. 12.13 Textural variation from margin (left) to middle (right) of Mesozoic diabase dike, Hartford Basin, Connecticut. At the margin, olivine microphenocrysts were quenched in a glass that contains dendritic crystals of plagioclase and augite; the texture here is vitrophyric. At slower cooling rates an intergranular texture developed with numerous pyroxene crystals occurring in each space between the plagioclase laths. At still slower cooling rates, a subophitic texture developed, with numbers of plagioclase and pyroxene crystals being approximately equal. At the slowest cooling rates, the number of pyroxene crystals is less than the number of plagioclase crystals, and the texture is ophitic. Width of field in photomicrograph at dike margin is 0.1 mm; in all others, it is 2.5 mm.

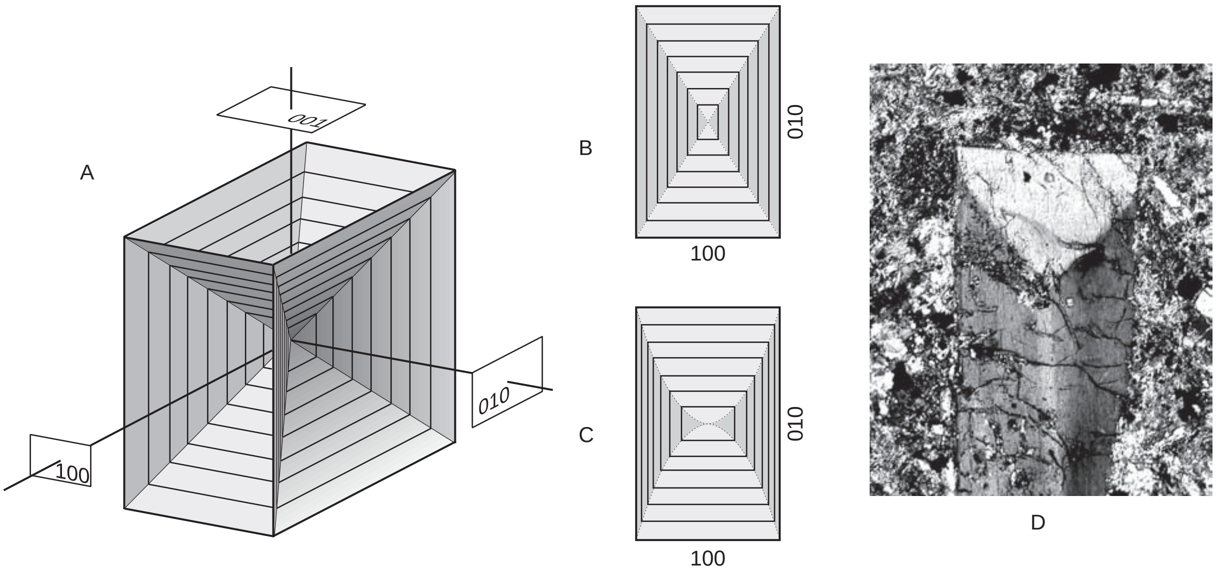


Fig. 12.14 (**A**) Crystals can grow with slightly different compositions on crystallographically different faces, giving rise to sector zoning. The shapes of sectors depend on growth rates of adjoining faces. If these rates remain constant, the sector boundary is straight (**B**); if they change at different rates, the boundary is curved (**C**). (**D**) "Hourglass" sector zoning in titanaugite crystal in alnöite from Isle Cadieux, Quebec.

One special case where a record of the rate-controlling crystal growth mechanism may be preserved in the rock involves sector-zoned crystals (Holister, 1970; Gray, 1971; Brophy *et al*., 1999). This type of zoning, which is common in a number of igneous and metamorphic minerals (titanaugite, plagioclase, staurolite, and chloritoid, for example), is visible in thin section because different sectors of a crystal may have different colors, birefringence, or extinction angle (Fig. 12.14(D)). It is normally produced by different faces on the crystal growing with slightly different compositions. As these faces grow, they produce pyramid-shaped sectors extending out from the center of the crystal (Fig. 12.14(A)). In titanaugite, the (100) sector commonly has higher concentrations of Ti, Al, and Fe^{3+} than has the (010) or (001) sectors, but Brophy *et al*. (1999) found that augite phenocrysts in an Aleutian high-alumina basalt had highest Ti in the (010) sector relative to concentrations in the (100) and (110) sectors, and least in the (111) sector. In staurolite the (010) sector has a higher concentration of Ti than has the (001) or (110) sectors.

The boundary between two sectors marks the locus of points of intersection of two crystal faces at various stages of crystal growth; that is, as one crystal face advances some distance x, the other face advances a distance y, and the sector boundary marks the point of intersection of the two faces at each stage of growth. The boundary can be thought of as a graph of the growth rate of one face plotted against the growth rate of the other face. If, for example, the growth rates on the two faces remain constant, the boundary between the two sectors would be planar and in section would be a straight line (Fig. 12.14(B)). This is not to say that the rates need be the same on both faces; one face could grow faster than the other. If the growth rates changed at different rates during growth, the sector boundary would be curved (Fig. 12.14(C) and (D)). Crystals that exhibit strong zoning, especially of the oscillatory type, preserve this same geometrical information about the shape of the crystal during growth.

Gray (1971) carefully measured the shape of the "*hourglass*" sector zoning in a titanaugite crystal from a 5.5-m-thick mafic alkaline sill near Montreal, Quebec. As the name "hourglass" suggests, the boundaries between sectors are curved. Gray found that the (010)–(100) sector boundary is a parabolic cylinder (Fig. 12.14(C)). This indicates that as the crystal grew, the rate of advance of the (010) face decreased relative to that of the (100) face. Because these changes produce a parabolic sector boundary, they must be systematic. Indeed, if the (010) face grew by a diffusion-controlled mechanism and the (100) face by a phase-boundary reaction mechanism, the resulting sector boundary would be parabolic (Gray, 1971).

If the growth of the (010) face is diffusion-controlled, we can obtain an expression for the distance it advances (x_{010}) in time t by integrating Eq. (12.7) to give $(x_{010})^2/2k_D = t$. The constant of integration will be zero if the crystal at time zero is approximated by a point. Similarly, the distance the (100) face advances (x_{100}) in time t can be obtained by integrating Eq. (12.8) to give $x_{100}/k_R = t$. Any point on the sector boundary must represent a simultaneous solution to these two equations; that is, the time in both equations must be the same. Combining these two equations, we obtain the following expression for the sector boundary:

$$(x_{010})^2 = 2\left(\frac{k_D}{k_R}\right)x_{100} \tag{12.15}$$

This is the equation of a parabola ($y^2 = cx$) and would thus explain the shape of the boundary between the (010) and (100) sectors. Any other combination of rate-controlling mechanisms would have produced a sector boundary of a different shape (Problem 12.2).

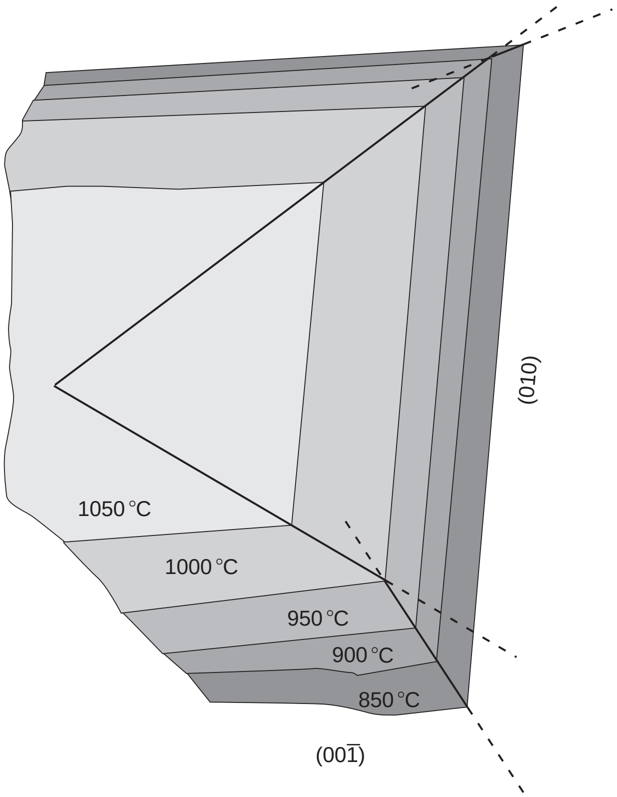


Fig. 12.15 Sector zoning in plagioclase crystal grown in experiments by Lofgren (1974, plate 3(A); or 1980, fig. 9(A)). Changes in the rate-determining mechanisms of growth on the various faces are indicated by the discontinuities in the lines drawn through the corners of the crystal at various stages of growth. See text for discussion.

If the rate-determining process for growth on a face of a sector-zoned crystal changes during the growth of the crystal, a discontinuity forms on the sector boundary. This can be illustrated by considering the shape of the sectors formed by growth zones in a plagioclase crystal grown in experiments by Lofgren (1974, plate 3(A); 1980, fig. 9(A)). A plagioclase melt of An_{50} composition containing 10 wt% H_2O at a pressure of 0.5 GPa was cooled to successively lower temperatures in 50 °C increments. While the temperature remained constant at each step, a normally zoned growth increment formed on each face of the crystal. The shape of the crystal at the end of each growth period is evident from the zoning (Fig. 12.15).

Consider first the sector boundary between (001) and (010). In thin section, this boundary is a straight line over most of its length, indicating that the rate-determining growth step on both faces was the same or had the same rate dependency. During the final growth period, however, rates on the two faces changed differently, as evidenced by the fact that the outer corner of the crystal does not lie on the line passing through the corner at earlier stages of growth. This deflection of the sector boundary could result from the (001) face growing more slowly or the (010) face growing more rapidly. If it was the latter, however, we might expect to see an inflection of the boundary between (010) and $(00\bar{1})$ sectors at this same cooling step, and we do not. It is likely, therefore, that the growth rate on the (001) face slowed relative to that on the (010) face near 900 °C, indicating a change in the rate-determining step for growth on the (001) face.

A change in the rate-determining step for growth on the $(00\bar{1})$ face can similarly be argued to have occurred when the temperature dropped below 1000 °C. On this face, however, the growth rate must have increased relative to that on (010) rather than decreased as it did on (001). Of course, because plagioclase is noncentrosymmetric, the growth-controlling mechanisms on (001) and $(00\bar{1})$ are not required by symmetry to be the same.

Considerable information on the nucleation and growth mechanisms of garnet crystals in metamorphic rocks is preserved in the zoning (Tracy, 1982) and grain-size distribution (Kretz, 1973) of this mineral. Most metamorphic garnets are strongly zoned, a feature not readily evident in thin section because garnet is isotropic and little color change accompanies the zoning. Compositional variation, however, is easily detected with electron microprobe analyses. Most garnets in pelitic rocks (rocks derived from shale), for example, have a manganese-rich core and iron-rich rim.

Of importance to interpreting nucleation and growth mechanisms from zoning is the fact that once a garnet has formed its composition cannot easily be changed, even by subsequent contact metamorphism. Diffusion rates in garnets are so low that steep compositional gradients show no signs of being lowered during most metamorphic events. As a first approximation, therefore, the composition at any point in a garnet crystal can be assumed to have remained unchanged since that part of the crystal grew.

Garnet can form by a number of reactions, the simplest of which is

$$\begin{aligned}&\underset{\text{(chlorite)}}{(\mathrm{Fe,Mn})_9\mathrm{Al}_3(\mathrm{Al}_3\mathrm{Si}_5)\mathrm{O}_{20}(\mathrm{OH}_{16})} + \underset{\text{(quartz)}}{4\mathrm{SiO}_2}\\ &\quad = \underset{\text{(garnet)}}{(\mathrm{Fe,Mn})_9\mathrm{Al}_6\mathrm{Si}_9\mathrm{O}_{36}} + \underset{\text{(water vapor)}}{8(\mathrm{H_2O})}\end{aligned} \tag{12.16}$$

In most rocks, chlorite and garnet would also contain magnesium, but for simplicity, we will restrict the discussion to the two components Fe and Mn. Chlorite and garnet both belong to solid solution series; the reaction, therefore, takes place over a range of temperatures (Fig. 12.16). The cryoscopic equation (Eq. (10.13)) developed in Section 10.9 to express the equilibrium between a solid solution and a liquid solution is equally applicable to the equilibrium between two solid solutions. If the chlorite and garnet are assumed to form ideal solutions, we can write

$$\ln\frac{X^{\mathrm{G}}_{\mathrm{Fe}}}{X^{\mathrm{Ch}}_{\mathrm{Fe}}} = \frac{\Delta H^{\mathrm{FeCh}}}{R}\left(\frac{1}{T^{\mathrm{FeCh}}} - \frac{1}{T}\right) \tag{12.17}$$

and

$$\ln\frac{X^{\mathrm{G}}_{\mathrm{Mn}}}{X^{\mathrm{Ch}}_{\mathrm{Mn}}} = \frac{\Delta H^{\mathrm{MnCh}}}{R}\left(\frac{1}{T^{\mathrm{MnCh}}} - \frac{1}{T}\right) \tag{12.18}$$

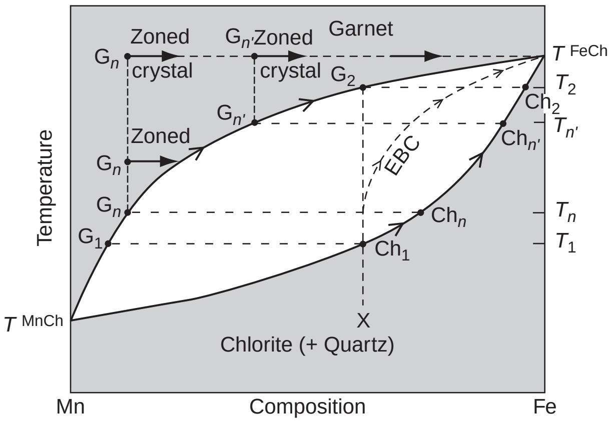


Fig. 12.16 Temperature–composition plot of the reaction of chlorite solid solution with quartz to form garnet solid solution. See text for discussion.

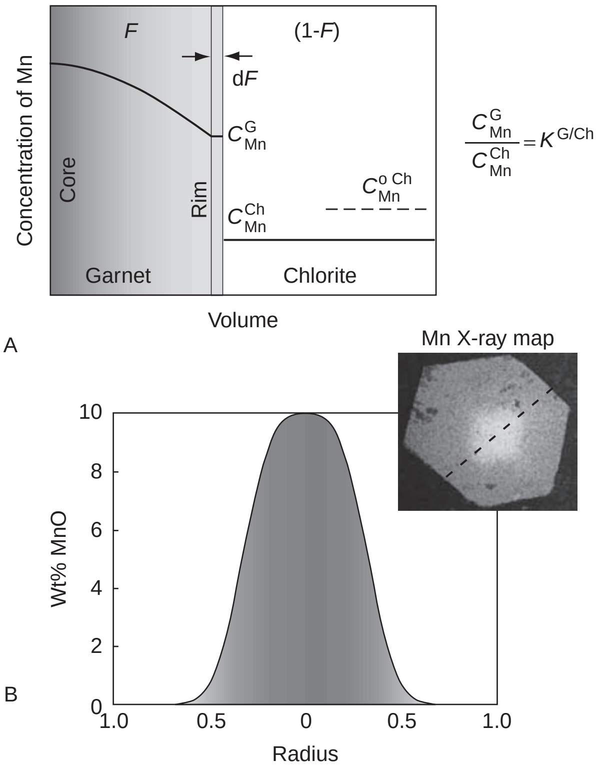


Fig. 12.17 (**A**) Model for the growth of a zoned garnet crystal from chlorite plus quartz by Rayleigh fractionation of Mn. F is the fraction of garnet formed; $(1-F)$ is the fraction of chlorite remaining; $\mathrm{d}F$ is a small growth increment, in which the concentration of Mn is related to the concentration of Mn in the chlorite by the Nernst distribution coefficient (see text for discussion). (**B**) Typical concentration profile of Mn across a garnet crystal that has formed by Rayleigh fractionation. Inset image is an electron microprobe X-ray map of Mn concentration in a garnet from the Barrovian garnet zone, Scotland. Mn content is proportional to gray scale, with highest concentration shown in the lightest shade. Note concentration is highest in the center of the crystal, and drops off toward the edges, consistent with the Rayleigh model. Dashed line = 1.51 mm.

where ΔH^{FeCh} and ΔH^{MnCh} are the enthalpies of reaction to form garnet from pure Fe chlorite and pure Mn chlorite, respectively, and T^{FeCh} and T^{MnCh} are the temperatures of these reactions for the pure Fe and Mn end members, respectively. The resulting phase equilibrium diagram (Fig. 12.16) looks identical to the melting loops for olivine and plagioclase except that only solid solutions are involved in this case.

Under equilibrium conditions, heating of a rock containing quartz and chlorite with Fe/(Fe + Mn) ratio of X in Figure 12.16 results in the formation of relatively Mn-rich garnet (G_1 in Fig. 12.16) at temperature T_1. As the temperature rises, the chlorite and garnet both become progressively enriched in Fe, reaching compositions Ch_2 and G_2, respectively, at temperature T_2 where, according to the lever rule, the last chlorite is consumed. But garnet crystals are zoned, and therefore equilibrium is clearly not achieved. As stated above, garnets, once formed, do not change their composition significantly. Consequently, the reaction from chlorite to garnet will approach complete fractionation. A more likely heating path for rock of composition X in Figure 12.16 would therefore be as follows. The rock must first be superheated to temperature T_n before garnet crystals nucleate. These crystals will again be Mn-rich (G_n) but not quite as rich as those first formed by equilibrium crystallization (G_1). With rising temperature, the garnet becomes more Fe-rich, but the initial Mn-rich cores remain because diffusion is unable to homogenize the crystals. A zoned crystal is therefore produced. Only the outer growing surface of the crystal has the equilibrium composition for a given temperature. This causes the effective bulk composition of the rock (EBC in Fig. 12.16) to become progressively enriched in Fe, and the garnet becomes zoned to rims of the pure Fe end member; chlorite in contact with the garnet also progressively changes its composition toward the Fe end member.

Because diffusion rates in garnet are so low and compositions remain essentially unchanged from the time of growth, the type of zoning that will result can be predicted quite accurately. The change from chlorite to garnet involves elements other than Fe and Mn (Mg and Ca, for example); indeed, the amount of Mn in the rock may be quite small. Let us assume that at a given temperature a garnet of some bulk composition is in contact with chlorite (Fig. 12.17(A)), and that the composition of the surface layer of garnet is in equilibrium with the chlorite. We will further assume that the surface layer of garnet has a concentration of Mn of c^{G}_{Mn}, which is related to the concentration of Mn in the chlorite, $c^{\text{Ch}}_{\text{Mn}}$, by the *Nernst distribution coefficient* (Section 14.12, Eq. (14.32)), $K^{\text{G/Ch}}$, which is defined as

$$K^{\text{G/Ch}} = \frac{c^{\text{G}}_{\text{Mn}}}{c^{\text{Ch}}_{\text{Mn}}} \qquad (12.19)$$

Because of the low diffusion rates, the Mn in each growth increment of garnet will be effectively removed from the system. To determine the zoning pattern that will develop in the garnet, we must be able to calculate how c^{G}_{Mn} will change during growth of the crystal.

If we let the fraction of garnet formed (Eq. (12.16)) at a given time be F, the amount of remaining chlorite will be $(1-F)$ (Fig. 12.17). The total amount of Mn in the chlorite at this

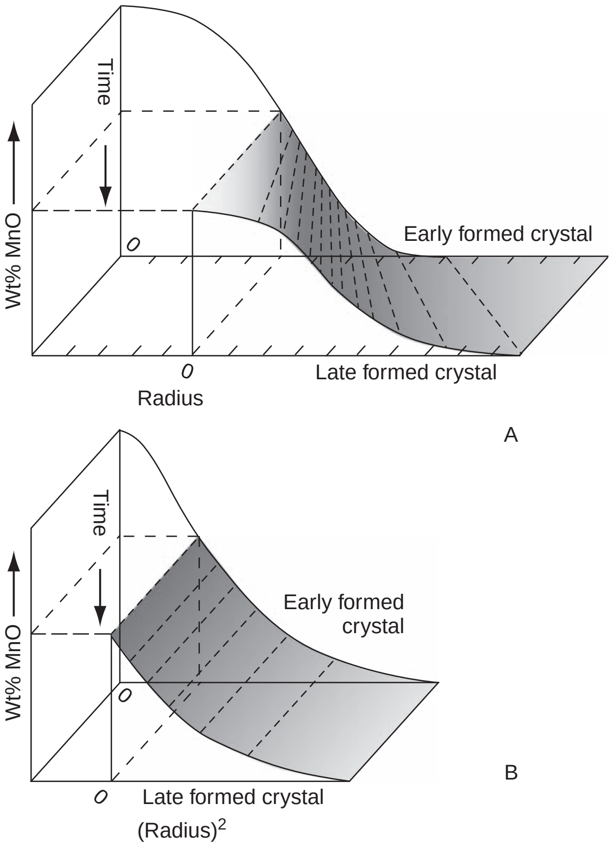


Fig. 12.18 (**A**) Compositional profiles of MnO in garnet crystals that grew early and late in a metamorphic episode. The time of nucleation of the late-formed garnet in terms of the growth of the early-formed crystal is indicated by the lines joining points of equal composition. (**B**) Same profile as in (A) but plotted against the square of the radius of the crystal. See text for discussion.

time will be $c_{\mathrm{Mn}}^{\mathrm{Ch}}(1-F)$, which from Eq. (12.19), can be written as $\left(c_{\mathrm{Mn}}^{\mathrm{G}}/K^{\mathrm{G/Ch}}\right)(1-F)$. Note that $c_{\mathrm{Mn}}^{\mathrm{G}}$ refers to the concentration of Mn in the surface layer of garnet, for it is here only that equilibrium exists – the core of the crystal may have a very different composition. If the garnet grows by a small increment, dF, the amount of Mn that enters that increment must be equal to the amount that the chlorite loses: that is, $-\mathrm{d}\left[\left(c_{\mathrm{Mn}}^{\mathrm{G}}/K^{\mathrm{G/Ch}}\right)(1-F)\right]$. The concentration in the growth increment, $c_{\mathrm{Mn}}^{\mathrm{G}}$, must therefore be

$$c_{\mathrm{Mn}}^{\mathrm{G}}=\frac{-\mathrm{d}\left[\left(c_{\mathrm{Mn}}^{\mathrm{G}}/K^{\mathrm{G/Ch}}\right)(1-F)\right]}{\mathrm{d}F}$$

On rearranging,

$$-K_{\mathrm{Mn}}^{\mathrm{G/Ch}}=\frac{\mathrm{d}c_{\mathrm{Mn}}^{\mathrm{G}}}{\mathrm{d}F}-c_{\mathrm{Mn}}^{\mathrm{G}}-F\left(\frac{\mathrm{d}c_{\mathrm{Mn}}^{\mathrm{G}}}{dF}\right)$$

Multiplying through by dF and rearranging, we obtain

$$\frac{\frac{\mathrm{d}c_{\mathrm{Mn}}^{\mathrm{G}}}{c_{\mathrm{Mn}}^{\mathrm{G}}}}{1-K}=\frac{\mathrm{d}F}{1-F}$$

Integration of this expression gives

$$\frac{\ln c_{\mathrm{Mn}}^{\mathrm{G}}}{1-K}=-\ln(1-F)+\text{constant} \qquad (12.20)$$

When $F=0$, $c_{\mathrm{Mn}}^{\mathrm{G}}=c_{\mathrm{Mn}}^{\mathrm{^oCh}}K^{\mathrm{G/Ch}}$, where $c_{\mathrm{Mn}}^{\mathrm{^oCh}}$ is the initial concentration of Mn in chlorite before garnet started to form. The constant of integration is therefore

$$\text{constant}=\frac{\ln\left(\mathrm{c}_{\mathrm{Mn}}^{\mathrm{^oCh}}K^{\mathrm{G/Ch}}\right)}{1-K}$$

Equation (12.20) then becomes

$$\ln c_{\mathrm{Mn}}^{\mathrm{G}}=\frac{-\ln(1-F)+\left[\ln\left(c_{\mathrm{Mn}}^{\mathrm{^oCh}}K^{\mathrm{G/Ch}}\right)\right]}{1-K}(1-K)$$

which simplifies to

$$c_{\mathrm{Mn}}^{\mathrm{G}}=c_{\mathrm{Mn}}^{\mathrm{^oCh}}K^{\mathrm{G/Ch}}(1-F)^{(K-1)} \qquad (12.21)$$

Equation (12.21) expresses how the Mn concentration in the surface layer of a garnet crystal changes as a function of the fraction of garnet formed, the Mn concentration in the initial chlorite, and the Nernst distribution coefficient for Mn between garnet and chlorite. After converting the fraction of garnet formed, which is a volume term, to crystal radius, we obtain a concentration gradient of the form shown in Figure 12.17(B) (Problem 12.3). Gradients such as this are common in garnet crystals (e.g. Holister, 1966).

This particular growth model, which is known as a *Rayleigh fractionation* process, has certain limitations. For example, it requires that diffusion of Mn to the garnet through the chlorite not be the rate-determining step. If Mn could not diffuse to the garnet rapidly enough, the chlorite near the garnet would become depleted in Mn, and the Mn content of the garnet would drop more rapidly than would otherwise happen. Eventually, when the Mn did diffuse to the garnet, it would produce a Mn-rich rim, a feature that is common to many garnets. Equation (12.21) also treats the Nernst distribution coefficient as a constant, but if the temperature changes during the reaction, as is likely during progressive metamorphism, the distribution coefficient may change. More complicated growth models have been developed (Tracy, 1982), but the Rayleigh fractionation model provides a simple explanation of the zoning process.

From Figure 12.16 it is clear that if during the growth of zoned garnet crystals – say, at temperature $T_{n'}$ – a second batch of crystals nucleates, their core composition would be poorer in Mn (G_n) than the cores of the first-nucleated crystals (G_n). The core composition of the late-formed crystal could be used to determine at what stage in the growth of the first crystal the later one nucleated. This is illustrated in Figure 12.18(A) by the line joining the core composition of the later-formed crystal to the point of identical composition in the zoning profile of the early-formed crystal. Also, because the second generation of crystals would have had less time to grow than the first, they would be recognizable in the rock as the smaller crystals.

If we were to compare the zoning profiles in the first and second generations of garnet crystals, we would expect the zoning in the second crystals to match that in the outer part of the first crystals. Kretz (1973), who measured the compositional profiles across garnets of different size from a single hand specimen, found that despite similar compositions, the gradients were different. He found that early-formed crystals grew proportionately less toward the end of their growth than did later-formed crystals in this same growth period, where the growth periods were defined by identical compositional changes (Fig. 12.18(A)).

This correlation between diminishing growth rates and the time of nucleation can be explained only if diffusion were the rate-controlling crystal growth process (Eq. (12.6)). If a phase boundary reaction had controlled growth, the rate would have been independent of the size of the crystals (Eq. (12.8)), and if surface nucleation had controlled growth, the larger crystals would have grown faster (Eq. (12.9)). With diffusion-controlled growth, the rate of change in the radius of a crystal is, according to Eq. (12.6), $dr/dt = K(D/t)^{1/2}$, where K is a constant and D is the diffusion coefficient. Upon integrating, we obtain

$$r = \text{constant} \times \sqrt{t}$$

which, if squared and differentiated with respect to time, gives

$$\frac{dr^2}{dt} = \text{constant} \qquad (12.22)$$

The squared term in Eq. (12.22) indicates that for diffusion-controlled growth the rate of increase of a crystal's surface area is a constant. If the zoning profiles in early- and late-formed garnet crystals in Figure 12.18(A) are plotted against the square of the radius, the profiles are seen to be the same. By this means, Kretz (1973) demonstrated that diffusion through the surrounding rock had been the rate-determining process for garnet growth in the sample he studied. He also showed, through the range of crystal sizes and core compositions, that nucleation of garnet had occurred not just at one time but throughout much of the growth period of garnet in that rock.

12.5 CRYSTAL SIZE DISTRIBUTION

During crystallization of magma or reaction of a metamorphic rock to form a new mineral assemblage, nuclei are continuously created, with those formed early having the opportunity to grow larger than those formed late. The result is a range of crystal sizes. The porphyritic texture, which characterizes so many igneous rocks, is an obvious example, as is the range in size of garnet crystals in a metamorphic rock, as discussed in the previous section. This range in size, which depends on both nucleation and crystal growth rates, is an important characteristic of the rock's texture.

Although terms such as aphanitic, porphyritic, groundmass, and fine-, medium-, and coarse-grained, etc. (see Section 6.4) convey qualitative information about the overall size and size distribution of grains in an igneous rock, they do not give a quantitative measure of this property. Marsh (1988a) introduced to petrologic studies a quantitative way of doing this, which had already been employed in the chemical engineering community to measure the distribution of crystal sizes (Randolph and Larson, 1971). In the industrial production of crystals, grain size must be carefully controlled if production of the desired size is to be maximized. Typically, a "mother liquor" is fed continuously through a "crystallizer" or "reactor" where nucleation and growth occur before the product appears at the outlet of the reactor. Thus, the time for nucleation and growth are known, as is the temperature, which is normally held constant. The resulting *crystal size distribution* (CSD) can be measured, from which important information about nucleation and crystal growth can be extracted. Although the process is different from the way in which crystals form in nature, the resulting CSDs can be analyzed in a similar manner.

The grain size of a rock depends on the number of crystals formed, which is determined by rates of nucleation and crystal growth. If the nucleation rate is low and the growth rate high, few crystals form, and the rock is coarse grained. If the nucleation rate is high and growth rate low, many crystals form, and the rock is fine grained.

Figure 12.19 illustrates the importance of these two rate processes to determining the grain size of a rock. This sample shows the lower contact of a decimeter-thick horizontal sheet of coarse-grained ferrodiorite from the center of the

Fig. 12.19 Lower contact of decimeter-thick ferrodiorite segregation sheet from the center of the 200-m-thick Holyoke flood-basalt flow, North Branford, Connecticut. Plagioclase is white, augite and pigeonite are gray, magnetite is opaque, and large dark areas in ferrodiorite are altered fayalite. The ferrodiorite separated from the host basalt when it was only one-third crystallized, bringing with it few nuclei; consequently, the ferrodiorite is coarse-grained, in contrast with the host basalt, which had a large number of nuclei. Width of field ~3 cm.

200-m-thick Holyoke flood-basalt flow in North Branford, Connecticut. This flow formed a huge sea of lava that covered much of Connecticut and possibly New Jersey. Based on measured cooling rates in Hawaiian lava lakes, the center of this flow probably took ~50 years to crystallize. The ferrodiorite formed from a liquid that was expelled from the underlying basalt when it was one-third crystallized. The contrast in grain size between the host basalt and the ferrodiorite is striking. In the host basalt, plagioclase, pyroxene, and magnetite grains are all less than a millimeter, whereas in the ferrodiorite, these same minerals have sizes approaching a centimeter. Both rocks formed at about the same temperature and had the same length of time to crystallize. Their grain-size difference is due to the number of nuclei in the host basalt being far greater than in the ferrodiorite. Despite the great thickness of the Holyoke flow, it is fine grained throughout. Large numbers of nuclei must have been dispersed throughout the flow at an early stage so that, despite slow cooling, large numbers of crystals formed and the rock is consequently fine grained. In contrast, very few nuclei were present in the ferrodiorite liquid when it separated from the underlying basaltic crystal mush when it was only one-third crystallized. Nucleation rates in the ferrodiorite were so low that crystals nucleated on preexisting crystals in the host basalt, as evidenced from the way crystals can be seen growing up from the lower contact (Fig. 12.19). The ferrodiorite liquid was also more iron-rich than the initial basalt, which would have reduced its viscosity and increased diffusion rates, which, in turn, would have increased crystal growth rates. Crystal size is clearly not determined simply by slow cooling.

The grain size of a rock depends on both nucleation and crystal growth, and if nucleation is a continuous process during the period of crystal growth, a range of crystal sizes will result. Let us assume that, as magma is crystallizing or new metamorphic minerals are forming, a distribution of grain sizes develops, such as that shown in Figure 12.20(A). The actual nature of this distribution need not concern us for the moment. We will assume that the number of crystals per unit volume, n, of a given size L is a function of L, where L is a measure of crystal size (e.g. the diameter of a garnet crystal or the length of a plagioclase crystal). The distribution in Figure 12.20 shows a typical pattern for crystals in a solidifying lava flow, where the value of n decreases as the grain size increases; that is, small crystals are more abundant than large ones. This increase is commonly found to be exponential, so that a plot of $\ln(n)$ versus crystal size is linear (Fig. 12.20(B)). Because the function $n(L)$ describes the number of crystals per volume as a function of size, it is known as the *population density* and has units of number/volume/length, which in the SI system is m^{-4} but because small volumes and small crystals are normally measured, cm^{-4} or mm^{-4} are more commonly used. The number of crystals in a small size interval from L to $(L + \mathrm{d}L)$ (shaded area in Fig. 12.20(A)) in a system with volume V is then

$$\mathrm{d}N = Vn(L)\mathrm{d}L \tag{12.23}$$

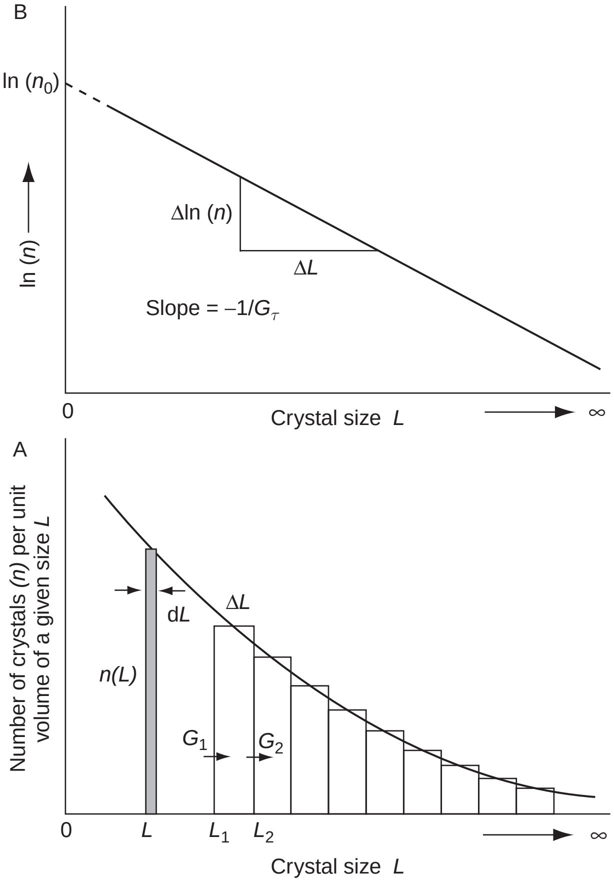


Fig. 12.20 (**A**) Crystal population density, $n(L)$, as a function of crystal size, L. The population density is the number of crystals per volume of a given size interval. G_1 and G_2 are the rates at which crystals grow into and out of the grain-size interval defined by L_1 and L_2 (see text for discussion). (**B**) Because $n(L)$ typically increases exponentially with decreasing crystal size, a plot of $\ln(n)$ versus L is commonly linear, with the slope being the negative inverse of the crystal growth rate × crystallization time, and the intercept at $L = 0$ (n_0) being the nuclei population density.

It is worth checking that the units here cancel to give numbers. The cumulative number of crystals from the smallest ($L \rightarrow 0$) to size L is given by integrating Eq. (12.23), that is,

$$N = \int_0^L n(L)\mathrm{d}L \tag{12.24}$$

and the number of crystals in any size interval L_1 to L_2 is

$$N(L_1, L_2) = \int_{L_1}^{L_2} n(L)\mathrm{d}L \tag{12.25}$$

Equation (12.25) is important because, in practical terms, this is the number we can measure. For example, sedimentologists use sieve sizes (L_1 and L_2) to determine the quantity of sediment falling in a particular grain-size interval. Similarly, in thin section, we can determine (though not easily) the number of crystals that fall in a particular grain-size window.

As magma crystallizes or a metamorphic reaction progresses, crystals will grow and new ones will nucleate, and consequently the population density will continuously change. Imagine that you can monitor this process with a scope that allows you to see the population density in any given size window ΔL, say from L_1 to L_2 in Figure 12.20. The population density in this window will increase as crystals, growing at a rate G_1, enter from the next smaller window. At the same time, the population density in the window will decrease as crystals grow out of it at a rate G_2 into the next higher window. The change in the population density in the window due to growth will be the balance between the rates of crystals growing into and out of the window; that is, for a time interval of Δt, in a system of fixed volume,

$$\begin{array}{ccccc} \text{growth accumulation} = & \text{growth input} & - & \text{growth output} \\ \Delta n \times V \times \Delta L & = V \times n_1 \times G_1 \times \Delta t & - & V \times n_2 \times G_2 \times \Delta t \end{array}$$

Again, check that the units cancel to give numbers.

The population density can also be changed by the flux through the system physically removing crystals from our window. If the influx is Q_i and the outflux Q_o, the change in population due to flux is

$$\begin{array}{ccccc} \text{flux accumulation} = & \text{flux input} & - & \text{flux output} \\ \Delta n \times V \times \Delta L & = Q_\mathrm{i} \times n_\mathrm{i} \times \Delta L \times \Delta t & - & Q_\mathrm{o} \times n_\mathrm{o} \times \Delta L \times \Delta t \end{array}$$

The total change in the population density in the time interval Δt is therefore the cumulative effect of both growth and flux; that is,

$$\Delta n V \Delta L = \overset{\text{growth}}{(V n_1 G_1 \Delta t - V n_2 G_2 \Delta t)} + \overset{\text{flux}}{(Q_i n_i \Delta L \Delta t - Q_o n_o \Delta L \Delta t)}$$

On rearranging, this equation becomes

$$\frac{\Delta n}{\Delta t} = \frac{-(n_2 G_2 - n_1 G_1)}{\Delta L} + \frac{Q_\mathrm{i} n_\mathrm{i}}{V} - \frac{Q_\mathrm{o} n_\mathrm{o}}{V} \tag{12.26}$$

In the limit when ΔL and Δt approach zero, Eq. (12.26) becomes

$$\frac{\partial n}{\partial t} = -\frac{\partial (nG)}{\partial L} + \frac{Q_\mathrm{i} n_\mathrm{i}}{V} - \frac{Q_\mathrm{o} n_\mathrm{o}}{V} \tag{12.27}$$

Equation (12.27) describes the population balance and how it changes with time.

If the population balance achieves a steady state, then $\partial n / \partial \mathrm{t} = 0$, and Eq. (12.27) becomes

$$\frac{\partial (nG)}{\partial L} = \frac{Q_\mathrm{i} n_\mathrm{i}}{V} - \frac{Q_\mathrm{o} n_\mathrm{o}}{V} \tag{12.28}$$

For simplicity, let us consider the case where the mother liquor entering a crystallizer or magma entering a chamber contains no crystals; then, $n_\mathrm{i} = 0$. In addition, we can define the total length of time the liquor spends in a crystallizer (τ) as being the volume of the system divided by the flux through the system, that is, $\tau \equiv V/Q$. In addition, if the growth rate, G, does not vary with crystal size, G can be taken out of the derivative. (Note that this would not be valid for the garnet crystals or the (010) sectors of titanaugite crystals discussed in the previous section, because their growth was diffusion controlled, and hence the growth rate decreased as the crystals grew larger.) Making these substitutions into Eq. (12.28) gives

$$G \frac{\partial n}{\partial L} = -\frac{n}{\tau} \tag{12.29}$$

which, upon integration gives

$$n = n_0 \exp\left(-\frac{L}{G\tau}\right) \tag{12.30}$$

Taking natural logarithms of both sides of this equation, we obtain

$$\ln n = \ln n_0 - \left(\frac{1}{G\tau}\right) L \tag{12.31}$$

which is the equation of a straight line, such as that shown in Figure 12.20(B). If the conditions we have imposed are met, then n_0 is the population density as L goes to zero, that is, the population density of nuclei, and the slope of the line is the negative inverse of the growth rate × the crystallization time. If either the growth rate or the crystallization time is known independently, then the other can be determined from this slope. The growth rate here is defined as $G \equiv \mathrm{d}L/\mathrm{d}t$, and if L is the length or diameter of a crystal, this growth rate would have to be halved to compare with growth rates discussed earlier of a crystal rate which was defined in terms of $\mathrm{d}r/\mathrm{d}t$ where r is the radius of an equant crystal (e.g. Eq. (12.9)). From Eq. (12.23) we know that

$$n_0 = \frac{1}{V} \frac{\mathrm{d}N}{\mathrm{d}L}\bigg|_{L=0}$$

But the nucleation rate (J) in a volume V is

$$J = \frac{1}{V} \frac{\mathrm{d}N}{\mathrm{d}t}\bigg|_{L=0}$$

Using the chain rule, we can rewrite this as

$$J = \frac{1}{V} \cdot \frac{\mathrm{d}N}{\mathrm{d}L} \cdot \frac{\mathrm{d}L}{\mathrm{d}t}\bigg|_{L=0} = n_0 G \tag{12.32}$$

Thus, the nucleation rate is the product of the nucleation density and the growth rate when the crystals are very small, which will be the same as for larger crystals if the growth rate remains constant. We can also substitute the functional form of $n(L)$ from Eq. (12.30) into Eq. (12.24) and, upon integrating from $L = 0$ to ∞, obtain the total number of crystals in a rock; that is,

$$N_\mathrm{T} = \int_0^\infty n(L) \mathrm{d}L = n_0 G\tau \tag{12.33}$$

It can also be shown that the total length of crystals is given by

$$L_\mathrm{T} = \int_0^\infty L n(L) \mathrm{d}L = n_0 (G\tau)^2$$

from which it follows that the mean length of the crystals will be given by

$$\bar{L} = \frac{L_{\mathrm{T}}}{N_{\mathrm{T}}} = G\tau \tag{12.34}$$

Thus, instead of using qualitative textural terms, such as medium-grained, the *mean grain length* of a rock can be given a quantitative value by using the inverse of the slope in the CSD plot.

Equation (12.31) was derived for an industrial crystallizer where fluid passes through with a given flux, spending a given time at temperature in the reactor. The crystallization that results is analogous to that which would occur during the passage of magma through a feeder dike, but it does not fit the conditions for a stagnant body of magma cooling and crystallizing. If power to an industrial crystallizer were cut off, flow would cease and the crystallizer would cool down. These conditions would be closer to those encountered in a lava lake or a closed magma chamber. Chemical engineers have treated this situation, which they refer to as *batch crystallization*. Although more complex, the equation derived for the population balance in this case is almost identical to Eq. (12.31).

For batch crystallization, the volume of liquid does not remain constant but is continuously decreasing as crystallization progresses. The volume is therefore a function of time and crystal size, which we can account for by dividing Eq. (12.27) through by V and placing V within the derivative. In addition, because there is no influx or outflux of liquid, the last two terms in the equation disappear, giving

$$\frac{\partial(nV)}{\partial t} = -\frac{\partial(nGV)}{\partial L} \tag{12.35}$$

Marsh (1988a) gives the solution to this equation for constant growth rate as

$$n = n_0 \exp\left(-\frac{L}{Gh} + \frac{t}{h}\right) \times \left(\frac{V_0}{V(t)}\right) \tag{12.36}$$

where h is a constant with units of time. It is similar to τ in Eq. (12.29) and represents the total crystallization time. Taking natural logarithms of both sides of Eq. (12.36) gives

$$\ln n = \ln n_0 - \left(\frac{1}{Gh}L\right) + \frac{t}{h} + \ln\left(\frac{V_0}{V(t)}\right) \tag{12.37}$$

Equation (12.37) is similar in form to Eq. (12.31). At early stages of crystallization, the last two terms in Eq. (12.37) are very small and, consequently, a plot of ln(n) versus L is linear, with the slope being the negative inverse of the growth rate times the crystallization time. Because of this similarity, most petrologic analyses have used the simpler Eq. 12.31 to interpret CSDs.

The fact that CSDs for many igneous and metamorphic rocks plot as straight lines in ln(n) versus L plots (Figs. 12.21 and 12.22) has led petrologists to conclude that nucleation and crystal growth in rocks behave similarly to that in industrial crystallizers, although usually on a much longer timescale, and that valuable information can be extracted from the CSDs. Since Marsh's (1988a) introduction of CSDs to the

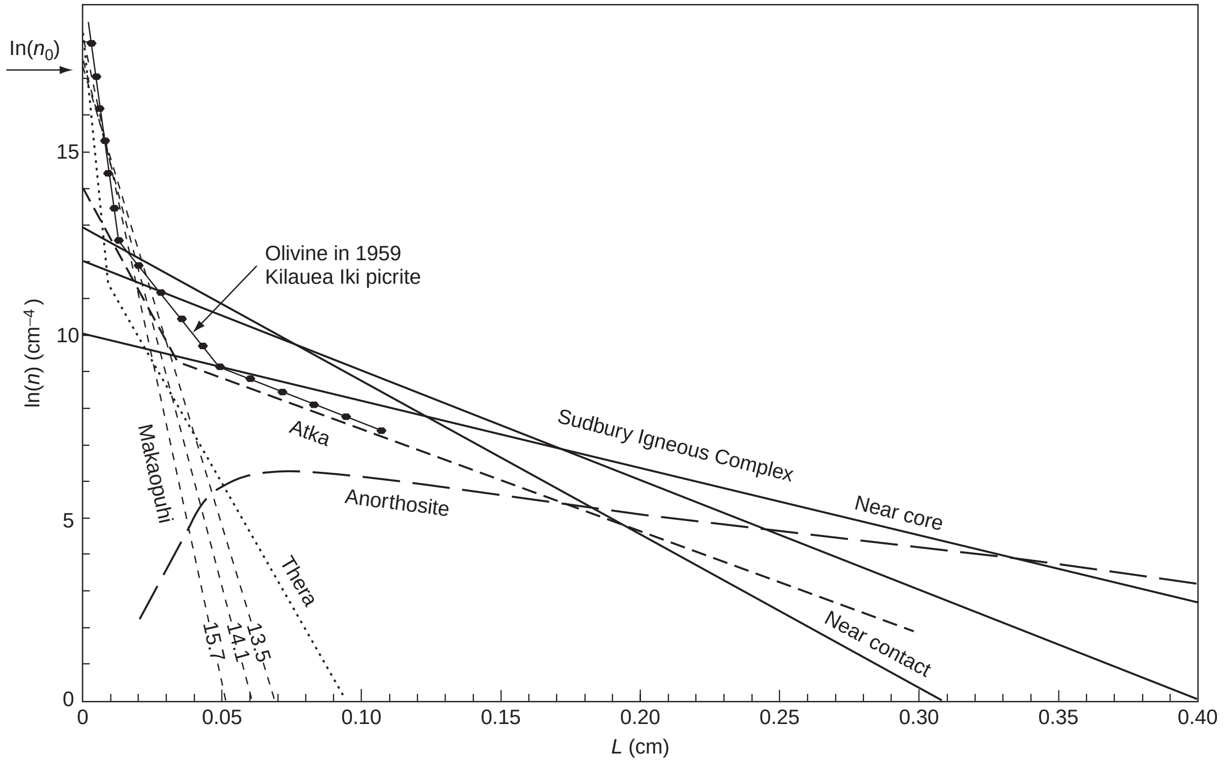


Fig. 12.21 Crystal size distributions of some representative igneous rocks plotted in terms of the natural logarithm of the population density, ln(n), versus the crystal size, L. (Data from Cashman and Marsh, 1988; Marsh, 1998; Zieg and Marsh, 2002; Higgins, 1996, 1998.)

petrologic community, there has been an explosion of petrologic research dealing with CSDs (Higgins, 2006).

The earliest of these studies was by Cashman and Marsh (1988), where they studied CSDs in samples of drill core from Makaopuhi lava lake, Hawaii. This lava lake formed in 1965, and as it cooled and solidified, members of the USGS periodically drilled through its thickening crust. Its solidification was therefore well documented, and the samples provided an ideal opportunity to test the CSD model because crystallization temperatures and times were known. Samples were studied from a 1968 drill core that penetrated the crust, which by then had thickened to ~16 m; below this, the lava lake was still largely liquid. The CSDs were obtained on samples at depths ranging from 13.5 to 16.5 m, where the temperature increased from 1011 to 1082 °C and the amount of crystallization decreased from 82.8% to 20.1%. The CSDs for plagioclase were all straight lines in the $\ln(n)$ versus L plot and showed a systematic decrease in slope with increasing amount of crystallization and, hence, time of crystallization (short dashed lines in Fig. 12.21). If the growth rate (G) remained relatively constant, the longer crystallization time (τ) in the shallower samples would make τG larger and, hence, the slope shallower. From the slopes and intercepts and using Eq. (12.30) and (12.32), they showed that the growth rate and nucleation rate of plagioclase was 5.4 to 9.9×10^{-11} cm s^{-1} (0.02 mm a^{-1}) and 0.034 to 0.016 cm^{-3} s^{-1} (5 to 10×10^4 $cm^{-3}a^{-1}$), respectively. These values agree well with those determined previously by Kirkpatrick (1977) for this lava lake. The growth rates are very much smaller than the experimentally determined values discussed in Section 12.3 (~10^{-4} cm s^{-1}), but these were maximum growth rates measured at 100 to 200 °C of undercooling. Cashman and Marsh conclude that plagioclase in the Makaopuhi lava lake grew with <1 °C of undercooling.

Plutonic igneous rocks can also show linear CSDs. Zieg and Marsh (2002), for example, show that plagioclase crystals in the norite of the Sudbury Igneous Complex in Ontario, which was formed by a large meteorite impact (see Section 15.9), have linear CSD plots that fan, with those closest to the contact having the steeper slopes and higher n_0 than those in the core of the body (solid lines in Fig. 12.22). If plagioclase growth rates remained constant throughout the body, the shallower slopes of the CSDs in the interior of the body are consistent with a longer crystallization time (Problem 12.4).

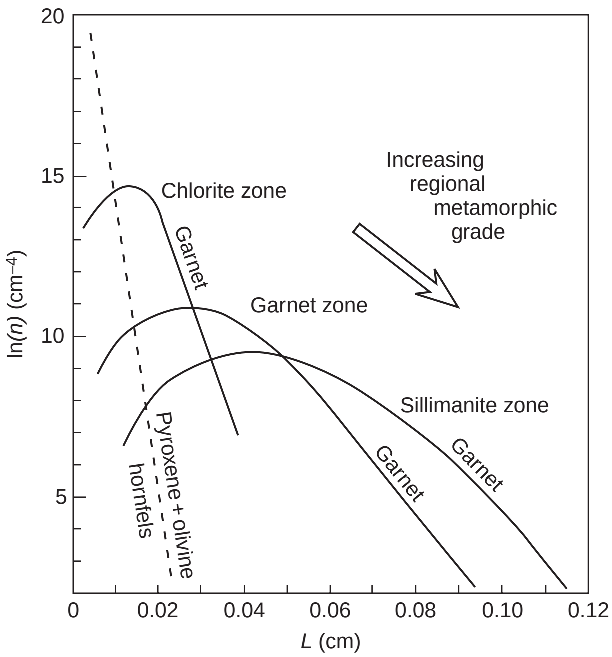


Fig. 12.22 Crystal size distributions of pyroxene and olivine crystals in a contact metamorphic hornfels from the Isle of Skye and of garnet crystals in regional metamorphic rocks of the Waterville Formation of Maine, USA. (Drawn from data presented by Cashman and Ferry, 1988.)

Not all CSDs plot as straight lines. For example, the high-alumina basalt from the Atka volcano in the Aleutian Islands, Alaska, shows a kinked plot (medium dashed line in Fig. 12.21), which Marsh (1998) interprets as the product of two distinct nucleation environments, with the coarser crystals having nucleated more slowly in a deep magma chamber and the finer crystals having nucleated more rapidly during migration to the surface. Similar kinked plots can result from the addition of mature crystals to a crystallizing magma, as might happen by crystal settling. The picrite that formed the 1959 Kilauea Iki lava lake (Fig. 2.2) contains an excess of olivine crystals over what would be expected in normal basalt. The addition of earlier-formed crystals gives the CSD plot a distinct concave upward pattern (Fig. 12.21) because of the larger number of large crystals (Marsh, 1998). Concave upward CSD plots can also result from magma mixing (Higgins, 1996) as illustrated by the plagioclase crystals in the dacite from Thera, Greece (dotted line in Fig. 12.21). CSDs can also show concave downward plots such as that of the plagioclase crystals in the Precambrian anorthosite from Lake St. John, Quebec (long dashed line in Fig. 12.21). This plot shows a distinct lack of small crystals, which Higgins (1998) attributes to *Ostwald ripening*, a process by which large crystals grow at the expense of small ones.

CSDs in metamorphic rocks show a similar range of patterns to those in igneous rocks (Fig. 12.22). There is a striking difference, however, between the CSDs in contact and regional metamorphic rocks (Cashman and Ferry, 1988). The CSDs of minerals that form in contact metamorphic zones tend to be linear and resemble those found in many igneous rocks. They have high values of n_0 and steep slopes. This similarity is not surprising considering that nucleation and crystal growth in a contact metamorphic aureole must operate on a similar time scale to the crystallization of the igneous body causing the metamorphism. In contrast, the CSDs in regional metamorphic rocks show a pronounced concavity downwards, due to a decrease in the number of small crystals. This decrease becomes more pronounced as the grade of metamorphism increases, as illustrated by the CSDs of garnet crystals in the chlorite, garnet, and sillimanite zones of the metamorphosed pelitic Waterville formation of Maine (see Section 16.3 for discussion of metamorphic zones). Cashman and Ferry interpret the low number of small crystals to be due to Ostwald ripening; that is, after long periods of heating, large crystals grow at the expense of small ones. Because this process requires diffusion through the matrix, it would be more rapid at higher temperatures, thus explaining the more pronounced curvature in the CSDs with increasing metamorphic grade. We must keep in mind, however, that a linear CSD was based on a constant growth rate. In the last section, we saw that compositional zoning patterns in large and small garnet crystals indicated that diffusion had controlled their growth, in which case growth rates should decrease with size of crystal. The decreased growth rate of large crystals would create a CSD that was concave downward.

In our discussion, we have assumed that data on crystal sizes and population density are readily obtainable. In industrial production, crystals can be removed from the slurry that exits a crystallizer and their sizes measured easily. In petrology, however, the situation is far more difficult, because we are usually restricted to examining rocks in two dimensions, for example in thin sections, polished sections, or outcrop surfaces. It is therefore necessary to determine three-dimensional information, such as the size and number of crystals per volume, from 2-D sections. The extraction of 3-D information from 2-D data is the field of stereology.

If the crystals we wish to measure were spheres of all one diameter, the mean diameter that would be obtained from random sections can be shown to be equal to 0.785 × true diameter. If the spheres have a range of sizes, the problem of extracting the true diameters from 2-D data becomes very much more difficult, because now the size variation can be due to the sectioning effect or the size range of the spheres. In addition, the probability of intersecting small spheres is less than that of intersecting large spheres. Although garnet crystals could be approximated by spheres, most crystals are parallelepipeds that can have highly variable dimensions. The shape plays an important role in determining how closely a measured crystal size comes to the true crystal size. If the crystals were long needles, for example, very few are likely to lie in the plane of section, and the mean size is likely to be far below the maximum length.

It is possible, however, to determine the true size and population density from serial sections (Jerram and Higgins, 2007). Multiple sections can be cut through a rock or its surface can be ground down incrementally, and the outline of crystals from one level connected with their outline in other sections. This is easier to do with coarser samples, but it is very time consuming. The 3-D distribution of crystals in a rock can also be obtained by X-ray computed tomography (CT) scan (see for example, Philpotts *et al*., 1999 and Fig. 14.11). This is identical to the medical technique used to obtain a 3-D image of parts of the human body. It creates a series of closely spaced 2-D images that are produced by the attenuation of X-rays passing through the object. These sections are then combined by computer to create a 3-D image. Minerals must have different X-ray attenuation coefficients to be distinguishable by this technique. Unfortunately, juxtaposed crystals of the same mineral show as one object in a CT scan, and therefore the technique is not ideal for measuring CSDs. Another drawback is the expense; one CT image of a sample can cost as much as $1000.

Despite these new techniques, routine measurements of CSDs are likely to continue to be made from 2-D sections. In an attempt to extract true crystal sizes and numbers from 2-D sections, Higgins (2000, 2006) created a computer program, *CSDCorrections*, which attempts to deal with all of the stereological problems. The program and other useful information on image analysis can be downloaded free from his web site at www.dsa.uqac.uquebec.ca/%7Emhiggins/home.html. The program takes measured widths or lengths of crystals from a measured area, and asks for estimated ratios of the long, intermediate, and short axes of the crystals and degree of crystal rounding and whether the sample is foliated or lineated. It then returns a plot ln(n) versus L for the corrected data. The results cannot be considered absolute, because they depend on the shape of the crystal chosen, which is not known. However, reasonable estimates of the shape can usually be made so that the corrected CSDs closely approach the actual distributions.

12.6 EQUILIBRIUM SHAPE OF CRYSTALS

The nucleation and growth of crystals is a nonequilibrium process that is controlled by kinetic factors. We have seen that these processes can be different on different faces of the same crystal. The shapes of crystals in rapidly cooled magma are a direct expression of these growth kinetics. However, given sufficient time and appropriate environment, crystals may develop forms that tend to minimize their surface free energy. Such shapes are independent of kinetics and are described as *equilibrium forms*. This process, first proposed by Gibbs (1875), can be expressed as

$$a_1\gamma_1 + a_2\gamma_2 + \ldots + a_n\gamma_n \text{ tends to a minimum} \tag{12.38}$$

where a_1 is the area of face number 1 and γ_1 is the surface free energy per unit area of that face. The surface free energy is defined as the change in free energy accompanying the creation of new surface area; that is,

$$\gamma_i \equiv \left(\frac{\partial G}{\partial a_i}\right)_{T,P,n} \tag{12.39}$$

Each face of a crystal has a surface free energy, which is determined by the structure of the mineral on that face and the material with which that face comes in contact. Strongly anisotropic minerals, such as mica, have very different surface energies on different faces. By adjusting the areas of these faces, a crystal can minimize its total surface free energy.

Pierre Curie (1885) showed that for a crystal that had minimized its surface free energy, the perpendicular distance from a crystal face to the crystal's center (d_i) divided by the surface free energy of that face is a constant for that crystal; that is,

$$\frac{d_1}{\gamma_1} = \frac{d_2}{\gamma_2} = \cdots = \frac{d_n}{\gamma_n} = \text{constant} \tag{12.40}$$

This expression later became known as *Wulff's theorem*, named after one of its chief proponents. It shows that the higher the surface free energy of a crystal face, the farther that face is from the center of the crystal, which in turn, because of geometrical considerations, means that this face is smaller than the other faces that have lower surface energies (Fig. 12.23(A)).

Consider, for example, the mineral staurolite, which in micaceous schists typically forms long prisms that have cross sections similar to the one shown in Figure 12.23(A). The cross section shown in Figure 12.23(B) is also possible but is certainly not common. The difference between these two crystals, which have identical volumes (same areas in cross section), is in the relative sizes of the (110) and (010) faces. If the (110) face has a lower surface free energy than the (010) face, the crystal in Figure 12.23(A) would clearly have a lower total surface free energy than the one in Figure 12.23 (B). According to Wulff's theorem, $d_{110}/\gamma_{110} = d_{010}/\gamma_{010}$ (Fig. 12.23), or $\gamma_{010}/\gamma_{110} = d_{010}/d_{110}$; that is, the surface free energy on the (010) face relative to that on the (110) face is greater by a factor of d_{010}/d_{110}, which in this case is 1.7. The independent values of the surface free energies, however, cannot be determined from this relation.

In comparing the total surface free energies of the crystals in Figure 12.23, one might ask: could the surface free energy be lowered still more by completely eliminating the (010) face? However, note that as the (010) face is made smaller, the (110) face has to be extended, and the total surface area of the crystal actually increases. The more equant a crystal is the smaller is its surface area for a given volume. The crystal in Figure 12.23(B), for example, has a smaller surface area (perimeter) than the one in Figure 12.23(A). To achieve the equilibrium shape, therefore, a crystal must strike a balance between eliminating high-energy faces and keeping the total surface area of the crystal small. Note that for a truly isotropic mineral, the equilibrium shape would be a sphere.

Little is known about the actual values of surface free energy on common minerals, but there is little doubt that they play an important role in determining the shape of many metamorphic minerals and the ways in which these minerals are intergrown. In a rock, equilibrium requires that the surface free energy of the rock as a whole tends toward a minimum. Wulff's theorem is, therefore, not directly applicable to most minerals because it considers a single crystal in isolation. In a rock, however, some mineral grains are bounded by faces that would not be expected if that grain were considered in isolation. It is unlikely, for geometrical reasons, that all minerals in a rock can be bounded by crystal faces. In striving for an equilibrium texture, some minerals develop prominent crystal faces while others tend to fill in around the euhedral crystals. The effect of this is to decrease the surface area of the euhedral ones and increase that of the anhedral ones. Clearly, euhedral minerals must have greater surface free energies and thus need to be bounded by the faces that minimize their contribution to the total surface free energy of the rock. At the same time, however, the development of euhedral faces on one mineral means that another mineral with low surface energy is probably increasing its surface area and increasing the surface area of the grains in the rock as a whole. Equilibrium, then, must strike a balance between the development of crystal faces on some minerals and increasing the total grain surface area.

Based on petrographic experience, metamorphic minerals can be arranged in what is known as the *crystalloblastic series* (Table 12.1). A mineral can develop euhedral faces against any mineral below it in the series. The series must therefore reflect decreasing surface free energies from top to bottom.

Because surface free energies depend on the way in which crystal structures come together across grain boundaries, the total surface free energy of a rock can be lowered if grains are able to modify their relative orientations. Mica crystals, for example, when oriented with their basal planes parallel, as occurs in schists, create a lower surface free energy than when the crystals are randomly oriented. This factor must play some role in the development of schistosity, but the direction

Table 12.1 *Crystalloblastic series*

Magnetite, Rutile, Sphene, Pyrite
Sillimanite, Kyanite, Garnet, Staurolite, Tourmaline
Andalusite, Epidote, Zoisite, Forsterite, Lawsonite
Amphibole, Pyroxene, Wollastonite
Mica, Chlorite, Talc, Prehnite, Stilpnomelane
Calcite, Dolomite, Vesuvianite
Cordierite, Feldspar, Scapolite
Quartz

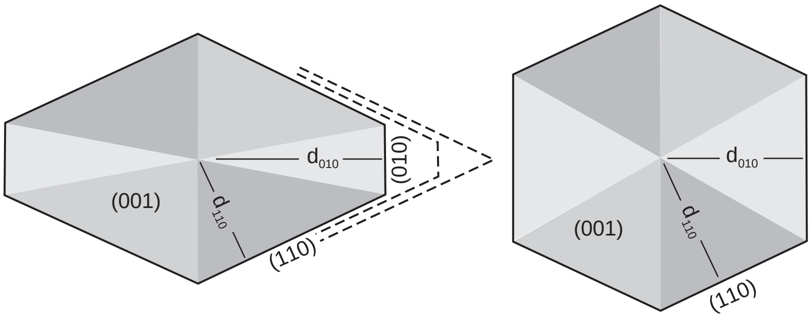


Fig. 12.23 Cross sections of staurolite crystals showing common (**A**) and uncommon (**B**) relative sizes of crystal faces. The smaller size of the (010) face in common crystals indicates that this face has a higher surface free energy than does the (110) face. According to Wulff's theorem, the ratio of the surface free energies on (010) and (110) is given by the ratio of the lengths of the lines d_{010} and d_{110}

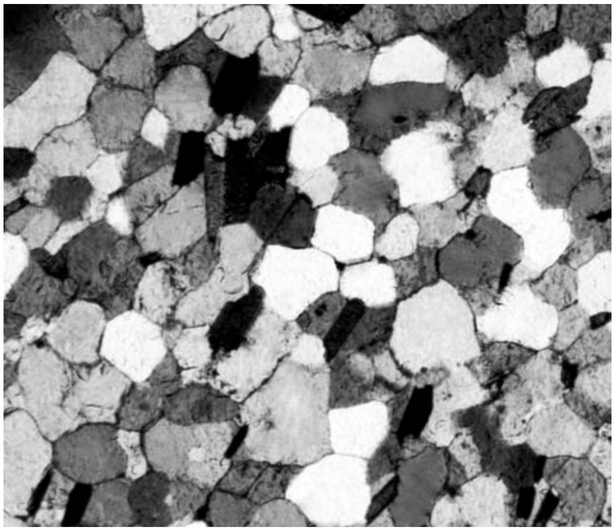


Fig. 12.24 Quartzite exhibiting granoblastic-polygonal texture. Most grain boundary intersections meet at angles of approximately 120°, and most grains are about the same size, indicating a close approach to an equilibrium texture. Width of field is 2 mm.

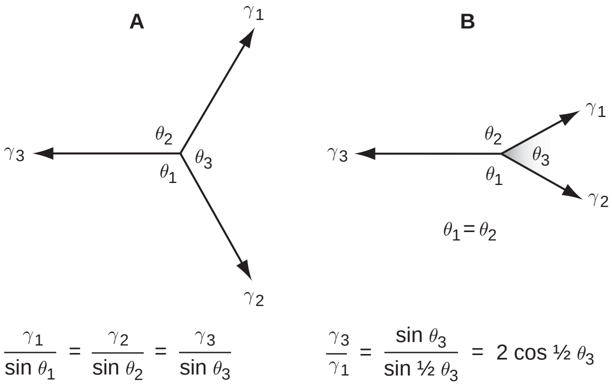


Fig. 12.25 (**A**) The angles between grain boundaries meeting at a triple junction are related through the law of sines to the surface free energies between the grains. These energies can be expressed as tensions acting on the triple point. At equilibrium, these forces must balance. If the surface free energies on all three boundaries are the same, the angles between the boundaries will be equal, that is, 120°. (**B**) At a triple junction involving two different minerals, the ratio of the surface free energies between the like grains (γ_3) and the unlike grains ($\gamma_1$and γ_2) is given by twice the cosine of half the dihedral angle subtended by the odd mineral.

of preferred orientation in a schist is, of course, determined by the stress field extant during crystallization. Some minerals actually nucleate on others in special crystallographic orientations that minimize the surface free energy between them. Such growth is described as *epitaxial*. A special case of epitaxy involves the growth of exsolution lamellae in such minerals as pyroxene, amphibole, and feldspar (Robinson *et al*., 1971).

Many grain boundaries in rocks are not rational crystallographic planes. For example, the essentially monomineralic metamorphic rocks quartzite and marble typically consist of anhedral polygonal grains. Such rocks commonly have a striking texture known as *granoblastic-polygonal*, where grain boundaries meet at angles of approximately 120° to form triple junctions (Fig. 12.24). This texture, which is also common in annealed metals (Fig. 12.4), can be explained in terms of surface free energies.

Commonly, surface free energies are discussed in terms of *surface tension*, which is the force necessary to stretch a surface. Such usage seems natural when discussing the close packing of soap bubbles, but its application to rigid crystals, while valid, is not inherently obvious. Nonetheless, this means of treating surface free energies provides a simple means of analyzing the angles at which grains come together.

Consider three grains of quartz, each with different crystallographic orientations, meeting at a triple junction. The surface free energies on the three boundaries are γ_1, γ_2, and γ_3, and the angles between grains opposite these boundaries are θ_1, θ_2, and θ_3, as shown in Fig. 12.25(A). If we treat these surface energies as tensions pulling at the triple junction, their magnitudes and directions can be represented by vectors. Then, according to the *law of sines*,

$$\frac{\gamma_1}{\sin\theta_1} = \frac{\gamma_2}{\sin\theta_2} = \frac{\gamma_3}{\sin\theta_3} \tag{12.41}$$

If the three surface tensions are equal, so must be the three sines, in which case $\theta_1 = \theta_2 = \theta_3 = 360°/3 = 120°$. Of course, surface tensions cannot literally pull the grain boundaries into this equilibrium shape. However, if the quartz is able to diffuse along grain boundaries, especially where aided by a fluid phase, solution and redeposition can result in the triple junction adjusting its shape to the equilibrium form.

The angles between grain boundaries at triple junctions will be 120° only when the surface free energies on these boundaries are all the same. Based on the frequency of occurrence of such junctions in quartzite, marble, and some amphibolites, surface free energies on quartz, calcite, and amphibole cannot be strongly dependent on crystallographic orientation. Some minerals, such as mica, however, have very different surface energies on different faces, and these minerals may not develop 120° triple junctions, nor do minerals that are surrounded by grains with a very different surface free energy, such as magnetite in quartz.

Consider, for example, the situation illustrated in Figure 12.25(B) where the mineral subtending the angle θ_3 at the triple junction is different from the mineral that forms the other two grains. The angles θ_1 and θ_2 are equal, because they involve the same mineral boundaries and hence the same surface energies. Consequently, $\theta_1 = {}^1/_2(360° - \theta_3)$. According to the sine law,

$$\frac{\gamma_3}{\sin\theta_3} = \frac{\gamma_1}{\sin\theta_1} = \frac{\gamma_1}{\sin\left(180 - \frac{1}{2}\theta_3\right)} = \frac{\gamma_1}{\sin\frac{1}{2}\theta_3}$$

from which it follows that

$$\frac{\gamma_3}{\gamma_1} = \frac{\sin\theta_3}{\sin\frac{1}{2}\theta_3} = 2\cos\frac{1}{2}\theta_3 \tag{12.42}$$

This equation, then, relates the *dihedral angle*, θ_3, to the ratio of the surface energies on the two types of grain boundary.

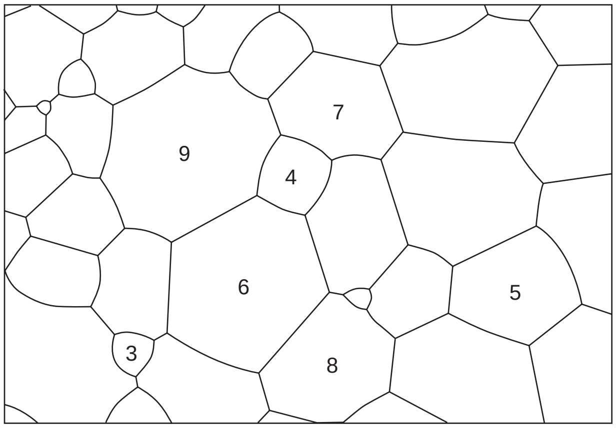


Fig. 12.26 All grain boundaries in this illustration intersect at angles of 120°. Six-sided polygons tend to be surrounded by planar faces. Polygons with more than six sides are surrounded by concave outward faces; those with fewer than six sides have convex outward faces. With time, small grains with convex outward faces are consumed by larger grains with concave outward faces.

Kretz (1966) found that the dihedral angle formed by pyroxene grains contacting pairs of scapolite grains in a high-grade metamorphic rock is 128°. According to Eq. (12.42), the surface free energy on the scapolite–pyroxene boundaries is therefore 1.14 times that on the scapolite–scapolite boundaries.

We have seen from Eq. (12.41) that if surface free energies on all grain boundaries in a rock are equal, the angles between grain boundaries at triple junctions will tend toward 120° at equilibrium. If all grains have hexagonal shapes in cross section, this angular requirement can easily be satisfied. But with polygonal grains having more or less than six sides in cross section, 120° triple junctions can exist only if grain boundaries are curved (Fig. 12.26). Grains with fewer than six sides will be bounded by faces that are concave inward, and those with more than six sides will be bounded by faces that are concave outward. In examining thin sections, we are presented with a slice through grains and the angles we see bounding crystals may be only apparent angles. For a discussion of the identification of equilibrium dihedral angles from random sections, see Elliott *et al*. (1997).

An important consequence of curved grain boundaries is that pressure differences develop across them, which provide a driving force to eliminate small grains and make large ones grow. This is the *Ostwald ripening* process referred to in the last section. Consider for a moment the pressure difference between the inside (P_i) and outside (P_o) of a small gas bubble in a lava. This difference determines the force that must be applied if the bubble is to be inflated. The work done in changing the volume at this pressure difference, $(P_i - P_o)dV$, must equal the surface free energy times the increase in area of the bubble; that is,

$$(P_\mathrm{i} - P_\mathrm{o})\mathrm{d}V = \gamma \mathrm{d}A \tag{12.43}$$

We can differentiate the expressions for the volume ($V = {}^4/_3\pi r^3$) and area ($A = 4\quad{}^2$) of a sphere with respect to radius to obtain $\mathrm{d}V = 4\pi r^2\mathrm{d}r$ and $\mathrm{d}A = 8\pi r\ \mathrm{d}r$. Substituting these into Eq. (12.43) gives

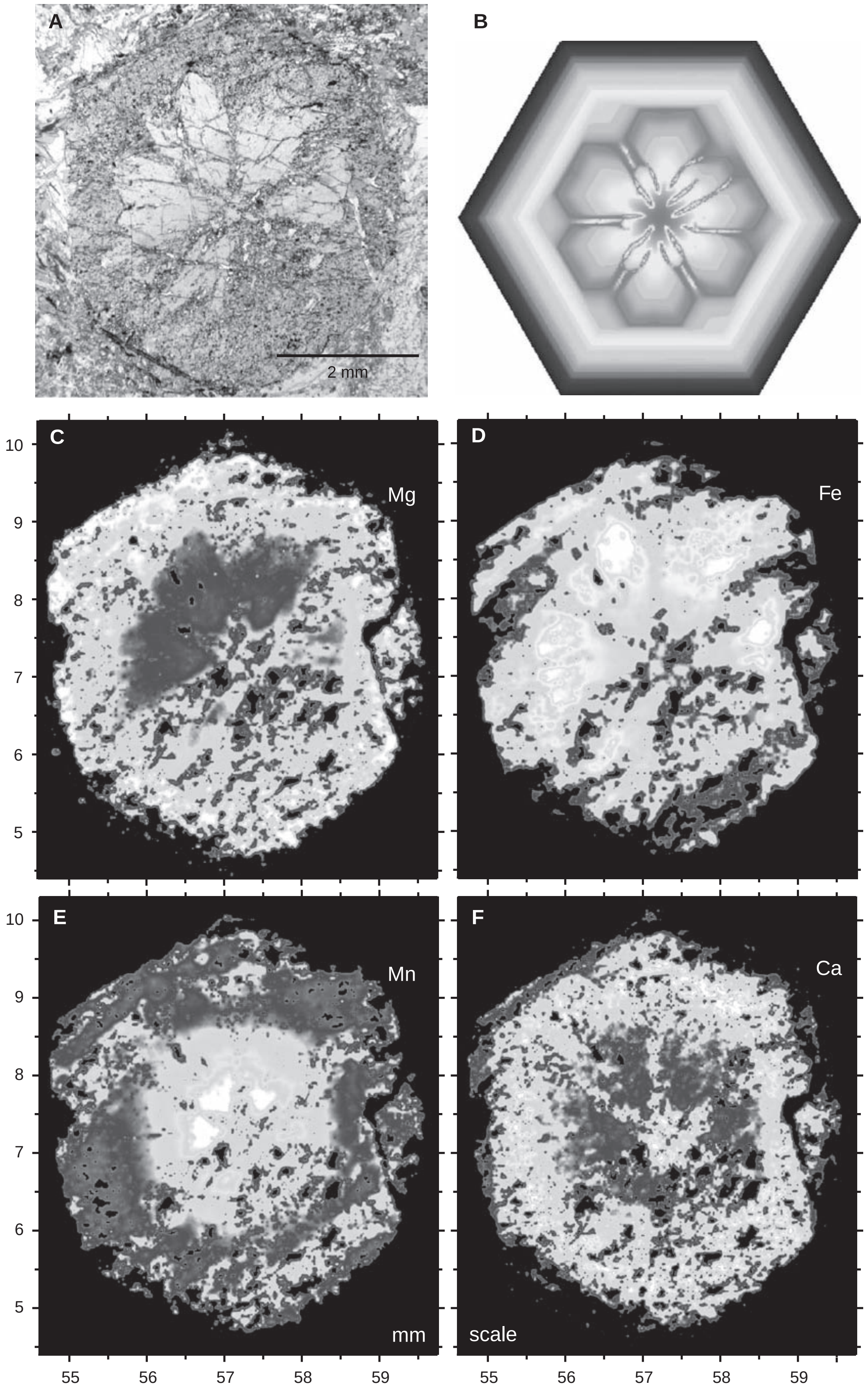
$$P_\mathrm{i} - P_\mathrm{o} = \frac{2\gamma}{r} \tag{12.44}$$

This relation indicates that the pressure difference between the inside and outside of a bubble is inversely proportional to the radius of curvature of the bubble. Thus, when a bubble is very small, the pressure in the bubble is very much greater than that outside. This explains why bubbles have difficulty nucleating (Problem 12.6).

Pressure differences across curved boundaries between crystalline phases are also given by Eq. (12.44). Because the radius of curvature of small grains is less than that of larger ones, the pressure in small grains is greater than that in large ones. From Eq. (7.32) we know that increased pressure raises the free energy of a phase $[(\partial G/\partial P)_\mathrm{T} = \bar{V}]$. Thus, small grains are metastable with respect to larger grains and will, as equilibrium is approached, be eliminated in favor of the larger ones. This process accounts for the general coarsening of metamorphic rocks and the development of concave downward CSDs during recrystallization (previous section).

In some rocks, a directed pressure (stress) can modify the shapes of grains by causing solution or diffusion of material from points of high pressure to ones of low pressure. This is known as *Riecke's* principle. Its effects are most clearly seen in deformed clastic sedimentary rocks. Quartz grains, for example, on being forced together dissolve at points of contact; the dissolved silica is transported through pore fluids and precipitated in crystallographic continuity on the sides of the grains where the pressure is less (Fig. 16.3). Transfer of material from the sides of grains where the compressive stress is at a maximum to the sides where it is a minimum causes grains to become flattened, and the rock develops a prominent foliation normal to the direction of maximum compressive stress (see Section 17.2).

Some garnet crystals in regional and contact metamorphic rocks have cores with a distinctive "branched" morphology, often surrounded by more euhedral rims (Fig. 12.27). This branching, known as textural sector-zoning (Rice and Mitchell, 1991), is common in the chiastolite variety of andalusite, and has been found in other minerals as well including corundum, beryl, and diamond. The cause of this phenomenon is still actively debated. One interpretation is that the branched forms represent growth far from chemical equilibrium, and are akin to the dendrites discussed previously in Section 12.3. In this type of growth, crystal faces propagate more rapidly than edges, or vice versa, leading to shapes that differ markedly from those expected for growth near chemical equilibrium (Ben-Jacob and Garik, 1990). Chemical maps of the garnets show that the core and rim areas are chemically distinct and thus grew under different conditions (Fig. 12.27(C)–(F)). Note that our earlier dendrite examples involved crystal growth from a melt, but for metamorphism the crystals grow from a matrix of solid mineral grains, with a fluid (normally water-rich) in the pore spaces between grains.


A
2 mm
B
C
Mg
D
Fe
E
Mn
mm
F
Ca
scale
10
9
8
7
6
5
10
9
8
7
6
5
55
56
57
58
59
55
56
57
58
59

For the crystal in Figure 12.27, it is possible that the equilibrium pressure-temperature-composition conditions for garnet growth were significantly overstepped as a result of, for example, activation energy barriers to garnet nucleation and growth, or sluggish dissolution of reactants. When the degree of overstepping was sufficiently large, growth began relatively far from equilibrium, producing disequilibrium crystal shapes. A later stage of growth occurred closer to equilibrium, yielding the euhedral rim. A computer simulation of crystallization accounting for rates of nutrient transport to the growing crystal and bonding energetics is able to qualitatively reproduce the garnet morphology if the initial crystal growth takes place during overstepped conditions, and the later growth is closer to equilibrium (Fig. 12.27(B)) (Wilbur and Ague, 2006, 2007). In this example, overstepping took place during metamorphic heating, but it can also occur as reactions are encountered on the cooling path. Growth occurring as a result of significant overstepping has considerable implications for metamorphism and metamorphic fluid flow, as we will see in Chapter 21.

12.7 SURFACE FREE ENERGY AND WETTING OF CRYSTALS BY MAGMA

The way in which water beads on a freshly waxed and polished car but spreads out evenly on an unwaxed surface is a familiar example of the "wetting" ability of a liquid. Wetting causes water to rise in a glass capillary and mercury to sink. The way in which a liquid wets a surface is determined by the tendency of the system to minimize its surface free energy.

The Si-rich liquid mesostasis of many tholeiitic basalts contains droplets of immiscible Fe-rich liquid (Philpotts, 1982) that commonly nucleate and grow as hemispheres on the surface of plagioclase crystals (Fig. 12.28(A)). The fact that these droplets do not wet the surface of the plagioclase indicates that such spreading would increase the surface free energy of the rock as a whole. We can conclude, therefore, that the surface free energy between plagioclase and Si-rich liquid must be less than that between Fe-rich liquid and plagioclase. But can we tell from the shape of the droplets how different these surface energies are?

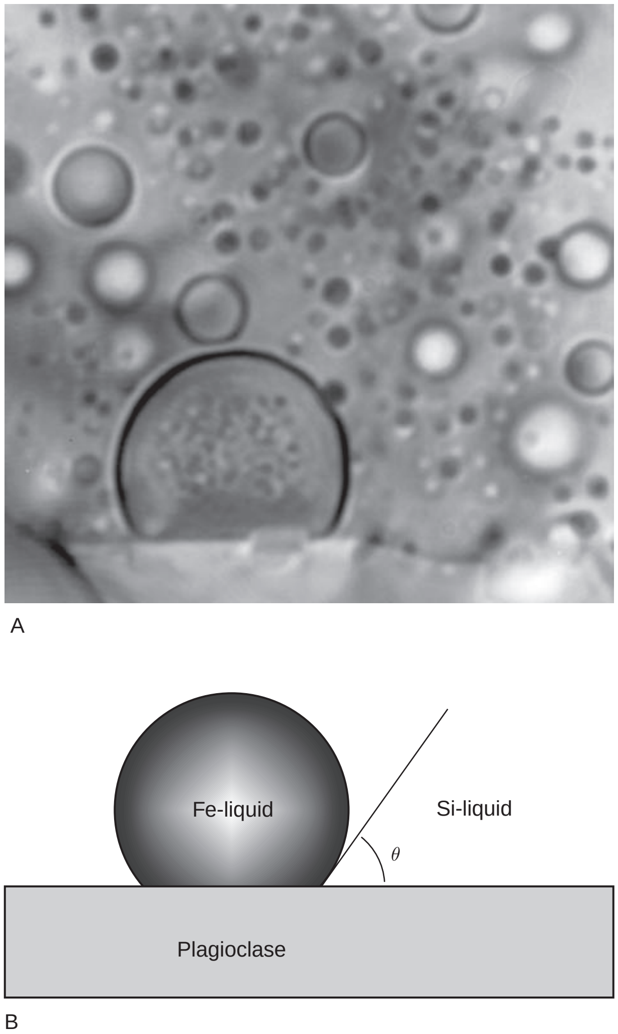


Fig. 12.28 (**A**) Droplet (20 μm diameter) of Fe-rich immiscible liquid on the surface of a plagioclase crystal in contact with Si-rich liquid in a tholeiitic basalt. (**B**) Schematic representation of the droplet in (A) showing the angle θ formed between the immiscible liquid boundary and the surface of the plagioclase crystal.

Figure 12.28(B) is a schematic representation of the point of contact of the boundary between the immiscible liquids and the plagioclase crystal. The surface free energies on the three different boundaries are $\gamma_{\mathrm{Fe-Si}}$ on the immiscible liquids boundary, $\gamma_{\mathrm{Fe-Pl}}$ on the Fe-rich liquid–plagioclase boundary, and $\gamma_{\mathrm{Si-Pl}}$ on the Si-rich liquid–plagioclase boundary. Each of these can be represented by a vector acting on the triple junction, which is assumed to be able to move to achieve equilibrium. Unlike the case considered in Figure 12.25, however, the face of the plagioclase crystal can be considered a rigid plane, which the deformable liquid boundary is able to move along. If at equilibrium the immiscible liquid boundary forms an angle of θ degrees with the crystal face, then the surface tension on the two-liquid boundary produces a component of force of $\gamma_{\mathrm{Fe-Si}} \cos\theta$ pulling in the same direction as the surface tension between plagioclase and Si-rich liquid. At equilibrium, the sum of these two forces must balance the surface tension between the Fe-rich liquid and the plagioclase; that is,

---

Caption for Fig. 12.27 Regional metamorphic garnet illustrating core with "branched" morphology, and more euhedral rim. Sample from south-central Connecticut, USA. (**A**) Plane-polarized light photomicrograph. (**B**) Computer simulation of garnet growth pattern (see text for discussion). (**C**)–(**F**) Mg, Fe, Mn, and Ca chemical element maps of garnet shown in part (A) obtained using the electron microprobe. Bright areas correspond to highest concentrations. The branched core is chemically distinct from the rim, indicating that the core and rim grew under different conditions. Note Mn-rich core in (E), consistent with Raleigh fractionation (compare with Fig. 12.17). (Modified from Wilbur and Ague, 2006, 2007; published by permission of the Geological Society of America.)

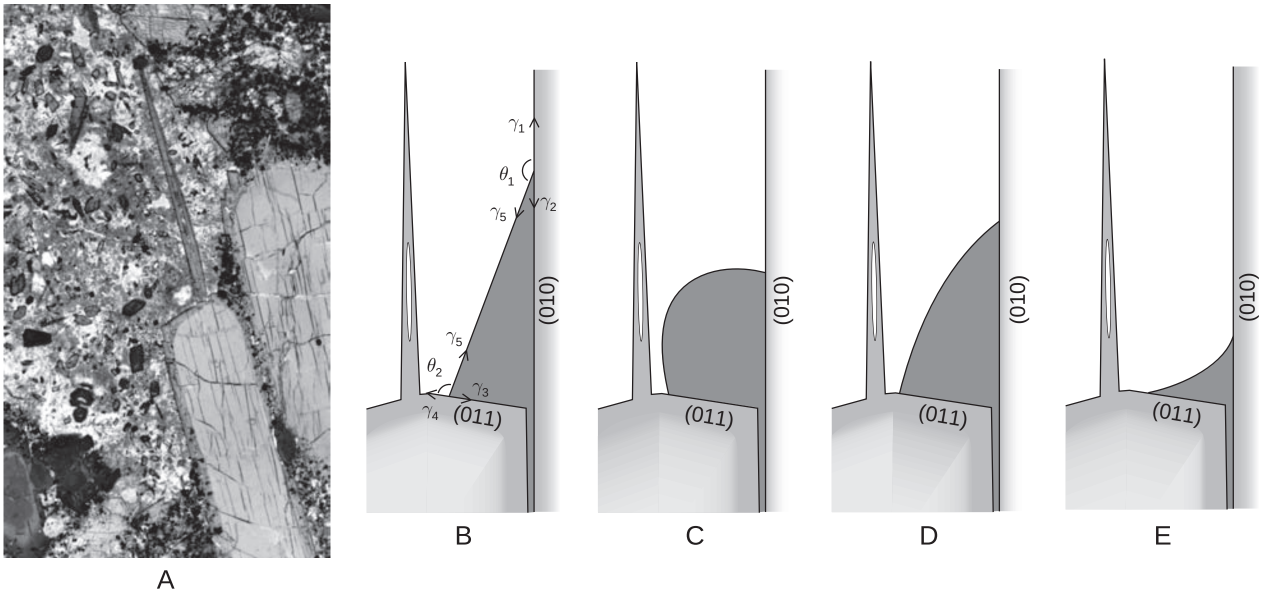


Fig. 12.29 (**A**) Large, zoned hornblende phenocrysts protruding across the boundary of a felsic ocelli in a lamprophyric dike, Montreal, Quebec. The fine-grained mafic groundmass of the lamprophyre and the felsic material are interpreted to have formed from immiscible liquids. Note that the boundary of the ocelli, as marked by the presence of magnetite crystals in the mafic fraction, can be traced through the marginal zone of the hornblende phenocrysts. The width of field is 4 mm. (**B**) Schematic representation of the patch of "mafic liquid" trapped between the two large hornblende phenocrysts in (A). (**C** – **E**) illustrate other conceivable configurations that this patch of liquid could have taken had the wetting properties of the liquids with the hornblende been different. See text for discussion.

$$\gamma_{\text{Fe-Si}} \cos\theta + \gamma_{\text{Si-Pl}} = \gamma_{\text{Fe-Pl}}$$

or

$$\cos\theta = \frac{\gamma_{\text{Fe-Pl}} - \gamma_{\text{Si-Pl}}}{\gamma_{\text{Fe-Si}}} \tag{12.45}$$

If $\gamma_{\text{Fe-Pl}} = \gamma_{\text{Si-Pl}}$, then $\theta = 90°$. If $\gamma_{\text{Fe-Pl}} < \gamma_{\text{Si-Pl}}$, then $\cos\theta < 0$, and $\theta > 90°$. In this case, the Fe-rich liquid would tend to wet the plagioclase. If $\gamma_{\text{Fe-Pl}} > \gamma_{\text{Si-Pl}}$, then $\cos\theta > 0$, and $\theta < 90°$. In this case, the Si-rich liquid tends to wet the plagioclase. Because the angle θ on the Fe-rich droplet in Figure 12.28 is less than 90°, we can conclude that $\gamma_{\text{Fe-Pl}} > \gamma_{\text{Si-Pl}}$ (Problem 12.7).

The result that the surface free energy on the Si-rich liquid–plagiocalse boundary is less than that on the Fe-rich liquid–plagioclase boundary is not surprising. The Si-rich liquid is highly polymerized and has a structure that is similar to that of the plagioclase. The amount of mismatch between the structures is therefore small and the surface free energy would also be small. The Fe-rich liquid is less polymerized, and therefore its structure does not match as closely the framework structure of the plagioclase and a higher surface free energy results.

Some lamprophyres contain *ocelli* (small globules of felsic rock), which are probably formed as immiscible droplets of felsic liquid in mafic melt (Philpotts, 1972). Where these droplets contact phenocrysts of amphibole or pyroxene, contact angles indicate that the mafic melt has a greater tendency to wet the surface of the ferromagnesian minerals than does the felsic melt (Fig. 12.29(A)). The surface free energy on the mafic liquid–crystal boundary is therefore less than on the felsic liquid–crystal boundary, which also would be expected from the structures of the phases involved.

A small pocket of mafic liquid trapped between the two hornblende phenocrysts shown in Figure 12.29(A) gives some indication of the relative surface free energies on different faces of the hornblende crystals. This pocket of liquid, which is shown schematically in Figure 12.29(B), wets the (010) face of one crystal and the 011 face of the other. From the angle the droplet's surface makes with both crystal faces, we can conclude that the mafic melt preferentially wet the hornblende crystal on both faces more than did the felsic melt. If it had not, the droplet would have had a shape such as that shown in Figure 12.29(C). If the hornblende showed only a slight preference to be wet by the mafic melt, a convex outward meniscus would still be possible (Fig. 12.29(D)). The fact that the meniscus is almost straight indicates that on both crystal faces the hornblende has a strong preference to be wet by the mafic melt rather than by the felsic one. Even so, the meniscus can still have several different configurations. If the meniscus took the form shown in Figure 12.29(E), we would conclude that the (011) face of the hornblende was wet more easily by the mafic melt than was the (010) face, and that the repulsion of the felsic liquid by the (010) face was less than it was by the (011) face. The actual shape of the meniscus (Fig. 12.29(A) and (B)) indicates that the reverse relations must be true. Using the nomenclature for the surface free energies on the different boundaries shown in Figure 12.29(B), we can conclude that $\gamma_2 = \gamma_1 + \gamma_5 \cos\theta_1$, and $\gamma_3 = \gamma_4 + \gamma_5 \cos\theta_2$, from which it follows that

$$\frac{\gamma_1 - \gamma_2}{\gamma_4 - \gamma_3} = \frac{\cos\theta_1}{\cos\theta_2} \tag{12.46}$$

Because $\theta_1 > \theta_2$, the ratio of the cosines is greater than 1. Therefore,

$$(\gamma_1 - \gamma_2) > (\gamma_4 - \gamma_3) \tag{12.47}$$

The individual values of surface free energy on the different boundaries cannot be determined from Eq. (12.47). But because the meniscus is not convex outward, the surface free energies between mafic liquid and hornblende (γ_2 and γ_3) must be significantly less than those between felsic liquid and hornblende (γ_1 and γ_4). We can conclude from Eq. (12.47) that γ_1 is probably greater than γ_4. This would indicate that the structure of a highly polymerized felsic melt is able to attach more easily to the ends of double chains of silica tetrahedra exposed on the (011) face of the hornblende crystal than it is to the lengths of such chains exposed on the (010) prism faces.

The differences in wetting ability of mafic and felsic melts on different faces of hornblende (and pyroxene) crystals may explain the striking difference in morphology of these crystals in mafic and felsic rocks. In gabbros, hornblende and pyroxene commonly form stubby crystals, whereas in felsic rocks (granites, syenites) they tend to be much more acicular. Indeed, this difference is seen in the lamprophyre illustrated in Figure 12.29(A), where the hornblende crystals in the mafic part of the rock have an *aspect ratio* (length/breadth) of approximately 3:1, whereas those in the ocelli have an aspect ratio of more than 10:1. One extremely long needle of hornblende extends in optical continuity from one of the stubby phenocrysts of hornblende into the ocelli.

If the mechanism controlling growth on the (010) and (011) faces on these crystals is a phase-boundary reaction, the greater ease with which the highly polymerized felsic melt attaches to the (011) face relative to its attachment on the (010) face would cause more rapid growth parallel to the *c*-axis. In a mafic melt where the ratio of bridging to non-bridging oxygens requires that the liquid be largely polymerized into silica chains, structural units of melt may be able to attach themselves almost as easily on faces paralleling the *c*-axis as they can on ones transverse to the *c*-axis. Differences in growth rate on these faces would therefore not be so great, and the aspect ratio would be smaller.

Finally, wetting plays an important role in determining the permeability of partially molten rock. Permeability, in turn, determines how easily melt can segregate from zones of melting in the mantle or lower crust to form bodies of magma or how easily melt can percolate through piles of crystal mush in magma chambers in response to compaction of the mush or convection in the melt (Cheadle *et al.*, 2004). Fusion of rock occurs first where grains of different composition come together (Fig. 10.5), for here is where the eutectic or peritectic mixtures of minerals occur. If the boundaries between this melt and crystalline phases are at equilibrium, the melt can be distributed at grain boundaries in three different ways depending on the interfacial energies (Fig. 12.30). From Eq. (12.42), the dihedral angle produced by the liquid (θ) is expressed in terms of the surface free energy between the liquid and solids (γ_{LS}) and between the solids themselves (γ_{SS}) by

$$\cos\frac{1}{2}\theta = \frac{\gamma_{SS}}{2\gamma_{LS}} \tag{12.48}$$

If γ_{LS} is greater than γ_{SS}, the dihedral angle will be greater than 120°, and the liquid forms isolated pockets at four-grain corners (Fig. 12.30(A)). If γ_{LS} is less than γ_{SS} but greater than $\gamma_{SS}/\sqrt{3}$, the dihedral angle is between 120° and 60°, and the liquid partly penetrates along grain edges (Fig. 12.30(B)). If γ_{LS} is less than $\gamma_{SS}/\sqrt{3}$, the dihedral angle is less than 60°, and the melt is able to penetrate along the entire length of grain edges (Fig. 12.30(C)).

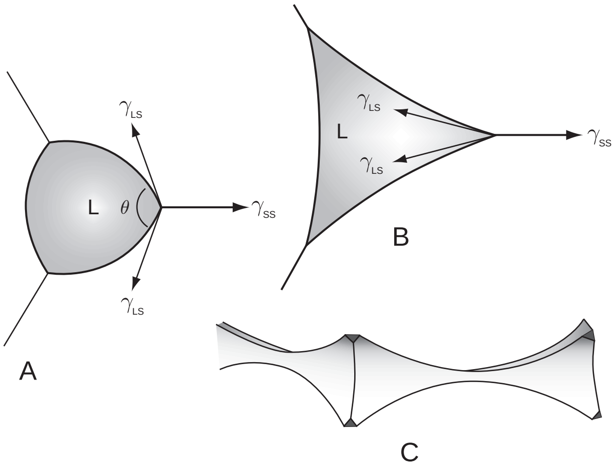


Fig. 12.30 Partial fusion of rock first generates pockets of liquid at grain junctions where minerals of different composition come together. The shape of the pockets depends on the wetting of the minerals by the magma, which is indicated by the dihedral angle, θ. If this angle is greater than 120° (**A**), liquid does not penetrate along grain boundaries; if the angle is between 120° and 60°, partial penetration occurs (**B**); and if the angle is less than 60°, the liquid penetrates along the entire length of grain intersections (**C**).

If partial melting gives rise to a liquid that does not wet the surface of grains ($\gamma_{LS} > \gamma_{SS}$), isolated pockets of melt formed at grain intersections (Fig 12.30(A)) will not connect with each other until a large fraction of the rock has melted. Only then will the melt be able to segregate and form bodies of magma large enough to rise. On the other hand, if the melt easily wets the solids ($\gamma_{SS} > \gamma_{LS}\sqrt{3}$), any small amount of melting will form a continuous liquid phase along grain boundaries. In this case, magma may be extracted from the partly fused rock at very small degrees of melting. In this way the wetting ability of magma plays an important role in determining the composition of bodies of magma (Section 23.4).

Holness (2006) measured the dihedral angle between liquid and clumps of phenocrysts of common minerals. With anisotropic minerals, different crystal faces can have different surface free energies, and hence the dihedral angle may vary depending on crystal orientation. The standard deviation (SD) of Holness' measurements is a reflection of this anisotropy. The median value (M) of the dihedral angle and its standard deviation both decrease from a maximum in clinopyroxene with M = 38° (SD 14) through amphibole with M = 31 (SD 14), olivine with M = 29° (SD 13), plagioclase with M = 25° (SD 11), leucite with M = 20° (SD 11), to quartz with M = 18° (SD 9). These measurements show that the dihedral angle is far less than 60°, and consequently these liquids will permeate crystal boundaries at exceedingly low melt fractions as long as equilibrium shapes exist. As might be expected, the more anisotropic minerals show a larger standard deviation than do the less anisotropic ones, indicating that surface energies in those crystals are more dependent on crystallographic orientation than in the less anisotropic ones. Although the melts investigated by Holness covered a wide compositional range, dihedral angles showed no apparent dependency on composition.

The liquid fraction in a partially crystallized magma can move through the crystal mush in response to compaction or convection, but to do so the mush must remain permeable. As we have seen, this will occur down to extremely low melt fractions (0.5%) if the liquid–crystal interfaces are at equilibrium. This is not the case, however, if crystal–liquid interfaces are governed by kinetic crystal growth processes. In this case, the volume of patches of liquid that are formed by crystals impinging on one another may have to exceed ~10% before they connect and cause the mush to be permeable (Cheadle *et al*., 2004).

All gradations between textures governed by crystal growth and those governed by equilibration of surface energies can be found (Holness *et al*., 2005). In those governed by growth, pore spaces (i.e. patches of minerals formed from residual liquid) are bounded by crystal faces, with the bounding crystals creating dihedral angles that have a median of ~60° but with a large standard deviation (25°). This indicates that the patches are roughly triangular. In equilibrated textures, crystals are commonly bounded by curved surfaces, and dihedral angles are ~28° with a smaller standard deviation (14°). Rapid cooling favors formation of growth-controlled textures and slower cooling favors equilibration textures.

12.8 PROBLEMS

12.1 Plot a graph of the free energy of growth of a vapor bubble in a basaltic magma as a function of radius if the ΔG_{vol} for the formation of the bubble is -2.4×10^3 kJ m^{-3}, and the surface free energy of the bubble is 0.3 J m^{-2}. Note that ΔG_{vol} is given in kJ m^{-3} rather than per mole so there is no need to divide by the molar volume. By differentiating Eq. (12.2), show that the critical radius is given by $r_{\text{c}} = -2\Delta G_{\text{intf}}\bar{V}/\Delta\bar{G}_{vol}$, and determine the value of r_{c}. Does your answer agree with your graph?

12.2 If growth on the (010) face of a crystal is controlled by a phase boundary reaction, and growth on the (100) face, which is 90° from (010), is controlled by surface nucleation, derive a mathematical expression for the sector boundary formed by the growth of these two faces. In integrating the growth equations, assume some finite size for the nucleus on which these faces grow. Draw a section through the crystal showing the sectors, and compare the result with the parabolic hourglass structure in Figure 12.14(C). Would you be able to distinguish these two patterns in thin section? Would "martini glass" rather than hourglass be a better description of this zoning (see, for example, Bryan, 1972)?

12.3 Assuming a Rayleigh fractionation model for garnet growing from chlorite (+quartz), calculate the type of zoning that would result if the initial concentrations of MnO and FeO were 0.5 and 1.0 wt%, respectively, and the distribution coefficient between garnet and chlorite is 20 for MnO and 0.5 for FeO. Also, calculate the profile for an element whose concentration is initially 1.0 wt% and has a distribution coefficient (G/Ch) of 5.0.

12.4 If we assume that the growth rate of plagioclase in the Sudbury Igneous Complex was similar to that in the Makaopuhi lava lake, say 6×10^{-11} cm s^{-1}, from the slopes of the CSD plots in Figure 12.21, calculate the length of time for the margin and core samples to crystallize.

12.5 Kretz (1966) found that the dihedral angle formed by grains of scapolite in contact with pairs of pyroxene grains in a metamorphic rock is 109°. What is the ratio of the surface free energies on the pyroxene–scapolite and pyroxene–pyroxene boundaries?

12.6 **(a)** What is the excess pressure in the bubble of Problem 12.1 when it has the critical radius (0.25 μm)? The surface free energy is 0.3 J m^{-2}.

(b) If the molar volume of the vapor is 4×10^{-3} m^3, what increase in the free energy of the vapor is caused by the excess pressure?

(c) If the entropy of the vapor phase is 280 J mol^{-1} K^{-1}, what degree of supercooling would be required to eliminate the increase in free energy due to the excess pressure in the bubble?

12.7 If the angle the surface of the Fe-rich droplet of immiscible liquid makes with the surface of the plagioclase crystal shown in Figure 12.28 is 64° and the surface free energy on the immiscible liquid boundary is 0.02 J m^{-2}, what is the difference between the surface free energies for the Fe-rich liquid–plagioclase boundary and the Si-rich liquid–plagioclase boundary?

13 Isotope geochemistry related to petrology

13.1 INTRODUCTION

Some of the major petrologic advances in recent years can be attributed to isotope geochemistry. The first contributions were in the field of absolute age determinations. Not only was the long-debated question of the age of the Earth settled, but absolute dating provided a means of unraveling the chronology of that large fraction (90%) of Earth history that had previously been designated "Precambrian" because of its lack of an adequate paleontological record with which to subdivide it. With increased refinements and new techniques, absolute dating now tackles problems ranging from the cosmological, such as determining the age of heavy elements in our solar system, to the very specific, such as the time of last movement on faults at a potential nuclear power plant site.

Perhaps of still greater importance to petrology is the contribution that isotope geochemistry has made to determining the provenance of magmas. Isotopic analyses can distinguish between magmas of mantle and crustal origin. In addition, degrees of contamination of mantle-derived magmas by crustal rocks can be measured through isotopic changes that could not be detected using major element chemistry. Isotopes can also indicate how much meteoric water is circulated through the rock cycle. In addition, isotopic studies reveal important information about the mantle itself, such as how its composition can change with time as a result of removal of partial melts or introduction of metasomatizing fluids.

Isotope geochemistry is such a large subject that only some of its petrologically most important aspects can be dealt with in this chapter. What is an isotope, and how does isotope geochemistry differ from what might be referred to as normal geochemistry? Atoms of an element that have different numbers of neutrons are known as *isotopes*. They consequently have different atomic masses, but their numbers of protons and electrons are the same (for a particular element), and thus normal chemical processes do not distinguish between most isotopes. Indeed, isotopic analyses cannot be done by normal chemical means but must be done with a mass spectrometer. At first, the inability of chemical processes to fractionate most isotopes suggests that isotopic analyses of rocks might serve little use. This, however, is not the case. The lack of fractionation of most isotopes by chemical processes means that once a rock has developed a particular isotopic ratio, this ratio will remain unchanged except in special cases, regardless of the chemical or physical processes that take place; thus the isotopic ratio serves as an indelible fingerprint with which to identify the rock. Elements themselves cannot possibly be used in this way, because every process induces some chemical change, which can be explained only if the process is fully understood (this is usually not the case).

Isotopic ratios can change through *radioactive decay* and *mass fractionation*. Some isotopes are radioactive and decay, at known rates, into isotopes of other elements. For example, ^{87}Rb, which substitutes for K in minerals, decays to ^{87}Sr, which in turn substitutes for Ca in minerals. Rubidium and strontium are chemically so different that they will be fractionated if exposed to processes such as partial melting. If left undisturbed, however, this pair provides the basis of one of the most useful absolute age determination methods. Isotopes with an atomic mass of less than about 20 can be fractionated to a measurable extent by some physical and chemical processes in which the mass of the isotopes makes a difference. For example, light isotopes can escape more easily than heavy ones during evaporation; consequently, ratios of the isotopes of H, He, and O in the surface layer of the oceans are depleted in the light isotopes relative to their ratios in the deep ocean. A list of all isotopes and their properties, including decay products and rates of decay if they are radioactive, can be found at http://atom.kaeri.re.kr/ton/index.html or http://t2.lanl.gov/data/map.html.

In general, most isotopic studies of petrologic interest fall into one of three main categories: (1) absolute dating using radioactive isotopes, (2) evolution of isotopic reservoirs in the mantle and crust, and (3) stable isotopes as indicators of ancient environments. These may overlap considerably, but the divisions provide a convenient basis for a brief survey of the subject, which is all that can be done in this chapter. For more extensive coverage, the books by Jager and Hunziker (1979), Faure (1986), Dickin (1995), Valley (2001), Faure and Mensing (2004) and Sharp (2006), and the articles by Hart and Allègre (1980), O'Nions *et al*. (1980), DePaolo (1981a), and O'Nions (1984) are recommended.

13.2 RADIOACTIVE DECAY SCHEMES

Some isotopes are inherently unstable and will, with time, change or "decay" to stable isotopes. The rate of decay is determined only by the instability of the radioactive nucleus and cannot be changed by external forces. This immutability is the basis of isotopic age determinations involving such element pairs as Rb–Sr, U–Pb, and K–Ar. The first element of each of these pairs has an isotope (or isotopes) that decays to an isotope of the second. Rubidium has two isotopes ^{85}Rb and ^{87}Rb; the first is stable, but ^{87}Rb decays to ^{87}Sr, which is just one of four stable isotopes of strontium, the others being ^{88}Sr, ^{86}Sr, and ^{84}Sr. Uranium has several isotopes, two of which, ^{238}U and ^{235}U, decay to ^{206}Pb and ^{207}Pb, respectively. Potassium has three isotopes, ^{39}K, ^{40}K, and ^{41}K, but only ^{40}K is unstable. It, however, can decay in two ways, either to ^{40}Ca or to ^{40}Ar. Each of these reactions involves a change in the nucleus of the parent isotope, but in each case, the way in which this takes place is different.

A nucleus can change by radioactive decay in four different ways. One is by emitting a particle consisting of two neutrons and two protons from the nucleus. This group of four nuclear particles is equivalent to a helium nucleus and is known as an *alpha particle* (α). It is the source of much of the helium in the Earth. The alpha particle has an atomic mass of 4 and a charge of +2. Another mode of decay involves the emission of an electron from the nucleus. This electron, which is known as a *beta particle* (β), is formed when a neutron in the nucleus changes into a proton. β particles have negligible mass. A third type of decay involves the capture of an electron by the nucleus where it combines with a proton to form a neutron. This decay process is known as *electron capture* (ec). The fourth mode affects only uranium and thorium, and involves the *spontaneous fission* of the nucleus into two nuclei of approximately equal mass.

Let us consider what changes must take place in the nucleus of $^{87}_{37}$Rb for it to change to $^{87}_{38}$Sr. The superscripts here refer to the atomic mass of the *isotope*, and the subscripts refer to the atomic number of the *element*. The atomic number is the number of protons in the nucleus, which defines that element. Because the atomic masses of ^{87}Rb and ^{87}Sr are the same, the decay process cannot involve α particles. The change, instead, involves an increase by one in the number of protons. This can be achieved only by a neutron changing into a proton, which requires that a β particle be emitted from the nucleus. The reaction can be written as

$$^{87}_{37}\mathrm{Rb} \rightarrow {}^{87}_{38}\mathrm{Sr}+\beta$$

The decay of uranium isotopes to those of lead is more complicated than that of rubidium to strontium. Both ^{238}U and ^{235}U decay in a long chain of reactions involving intermediate radioactive isotopes of the elements Po, Th, Ra, Pa, Bi, At, Ac, Rn, Fr, and Tl. Some of these isotopes require thousands of years to decay, whereas others last only fractions of a second. Despite the complexity of the chain reactions, the overall nuclear reactions can be written easily. The decay of $^{238}_{92}$U to $^{206}_{82}$Pb involves a mass loss of 32, which indicates that 8 α particles must be emitted. Because each α particle has a charge of +2, the loss of 8 α particles would cause the charge on the nucleus to drop to 76, but it only decreases to 82, the atomic number of lead. Consequently, the loss of 8 α particles must be accompanied by the emission of 6 β particles. The overall reaction can then be written as

$$^{238}_{92}\mathrm{U} \rightarrow (\text{chain}) \rightarrow {}^{206}_{82}\mathrm{Pb} + 8\alpha + 6\beta$$

In a similar way the decay of $^{235}_{92}$U can be written as

$$^{235}_{92}\mathrm{U} \rightarrow (\text{chain}) \rightarrow {}^{207}_{82}\mathrm{Pb} + 7\alpha + 4\beta$$

Although the overall decay scheme from U to Pb is used for dating rocks, the abundance and decay of some of the intermediate members of the series, such as thorium and radium, are used to study relatively short-lived processes in magma chambers. This topic is discussed in Section 15.3.

The decay of $^{40}_{19}$K to $^{40}_{18}$Ar and $^{40}_{20}$Ca involves no change in mass and so cannot involve α particles. The charge on the nucleus decreases by one when $^{40}_{19}$K changes to $^{40}_{18}$Ar so an electron must be captured by the nucleus, where it combines with a proton to form a neutron. The reaction can be written as

$$^{40}_{19}\mathrm{K} + \mathrm{ec} \rightarrow {}^{40}_{18}\mathrm{Ar}$$

The decay of $^{40}_{19}$K to $^{40}_{20}$Ca, on the other hand, involves an increase in the nuclear charge and must therefore involve emission of a β particle. The reaction is

$$^{40}_{19}\mathrm{K} \rightarrow {}^{40}_{20}\mathrm{Ca}+\beta$$

Of these three decay processes, only the emission of α particles has any significant effect on the material surrounding the decaying atom. Where this material is pleochroic, the bombardment by α particles affects the way in which light is absorbed. The result is a darkened region around the radioactive source that is known as a *pleochroic halo* (Fig. 13.1 (A)). With time, pleochroic haloes grow more intense. Unfortunately, there are too many variables to make halo intensity a reliable means of determining absolute ages.

Having far greater effect on the surroundings than α particles are particles given off by the spontaneous fission of ^{238}U and to a very much smaller extent of ^{235}U and of ^{232}Th. These heavy nuclei split into two nuclei of approximately equal mass and liberate ~200 MeV of energy. Following fission, the two nuclei recoil in opposite directions from each other, and as they pass through the surrounding mineral, they damage its structure, leaving a *fission track* (Fig. 13.1(B)). Although these damaged zones are so small that they are visible only in the electron microscope, etching of polished surfaces can highlight and enlarge them to the point that they are readily visible in a normal optical microscope. The older a mineral is, the more fission tracks it will contain for a given content of ^{238}U, and on this is based the fission track method of absolute dating.

13.3 RATE OF RADIOACTIVE DECAY

All radioactive decay, whether involving α particles, β particles, electron capture, or fission is a statistical process; the more atoms present of the radioactive nuclide, the more

Table 13.1 *Radioactive decay schemes and constants*

Radioactive isotope	Daughter isotope	Type of decay	Decay constant (a^{-1})	Half-life $t_{1/2}$ (a)
$^{14}_{6}\mathrm{C}$	$^{14}_{7}\mathrm{N}$	β	1.21×10^{-4}	5730
$^{40}_{19}\mathrm{K}$	$^{40}_{20}\mathrm{Ca}, ^{40}_{18}\mathrm{Ar}$	β, ec	5.543×10^{-10}	1.25×10^{9}
$^{87}_{37}\mathrm{Rb}$	$^{87}_{38}Sr$	β	1.42×10^{-11}	48.8×10^{9}
$^{147}_{62}Sm$	$^{143}_{60}Nd$	α	6.54×10^{-12}	106.0×10^{9}
$^{176}_{71}Lu$	$^{176}_{72}Hf$	β	1.94×10^{-11}	35.7×10^{9}
$^{182}_{72}Hf$	$^{182}_{74}W$	2β	7.7×10^{-8}	9×10^{6}
$^{187}_{75}Re$	$^{187}_{76}Os$	β	1.666×10^{-11}	41.6×10^{9}
$^{232}_{90}Th$	$^{208}_{82}Pb$	$6\alpha, 4\beta$	4.9475×10^{-11}	13.9×10^{9}
$^{235}_{92}U$	$^{207}_{82}Pb$	$7\alpha, 4\beta$	9.8485×10^{-10}	0.704×10^{9}
$^{238}_{92}U$	$^{206}_{82}Pb$	$8\alpha, 6\beta$	1.55125×10^{-10}	4.47×10^{9}

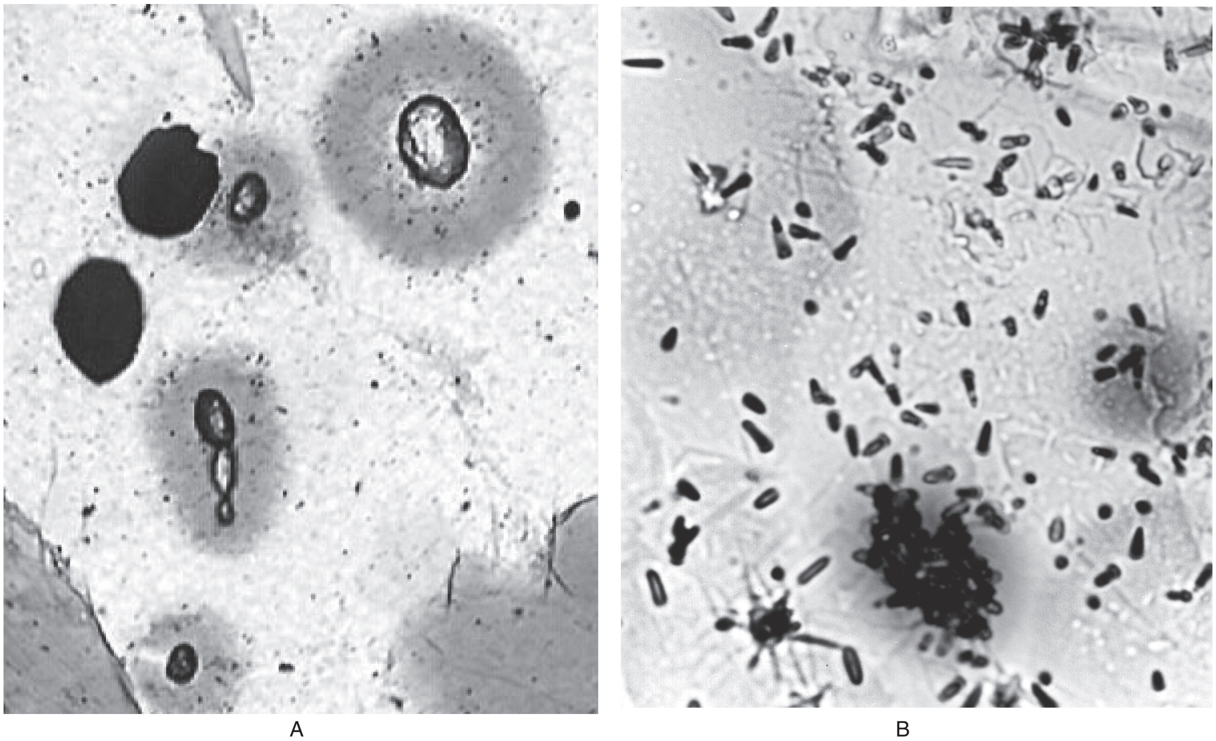


Fig. 13.1 (**A**) Yellow pleochroic halos surrounding zircon crystals in cordierite. Zircon typically contains small amounts of U and Th, which, on undergoing radioactive decay, affect the structure of the surrounding mineral. Plane light; width of field is 0.3 mm. (**B**) Fission tracks in biotite, revealed by etching for 20 s with concentrated hydrofluoric acid. Individual tracks are formed by single uranium atoms that have spontaneously split into two atoms of approximately equal mass. Clusters of tracks near the bottom of the photograph are associated with a pleochroic halo, at the center of which is a small crystal of zircon. The pleochroic halo near the right of the photograph is beneath the surface, so tracks within it have not been etched. Plane light; width of field is 0.2 mm.

chance a decay has of occurring. This can be expressed mathematically as

$$\text{rate of decay} \propto N \tag{13.1}$$

where N is the number of atoms present of the radioactive nuclide. According to the nomenclature of chemical reaction rates, such a process is referred to as a *first-order* reaction because the power to which N is raised in the rate expression is 1. We can write the rate of this process as the change in the number of atoms with time, inserting a minus sign to indicate that the change is negative, that is, decay:

$$\frac{\mathrm{d}N}{\mathrm{d}t} \propto -N \quad \text{or} \quad \frac{\mathrm{d}N}{\mathrm{d}t} = -\lambda N \tag{13.2}$$

where the constant of proportionality, λ, is known as the *decay constant*. Typical values of λ range from $10^{-4}\ a^{-1}$ for the rapid decay of ^{14}C to ^{14}N, to $10^{-12}\ a^{-1}$ for the very slow decay of ^{147}Sm to ^{143}Nd (Table 13.1).

Equation (13.2) gives the instantaneous change in the number of atoms with time. If we wish to know the total change in N over some interval of time, the equation must be integrated:

$$\int_{N=N_o}^{N=N_p} \frac{dN}{N} = -\lambda \int_0^t dt$$

Here N_p and N_0 are, respectively, the present and original numbers of atoms of the parent nuclide. Integrating yields

$$\ln \frac{N_p}{N_0} = -\lambda t$$

or

$$N_p = N_0 e^{-\lambda t} \tag{13.3}$$

This equation describes how the number of atoms of the parent nuclide decays with time.

For purposes of comparison and illustration, the concept of *half-life* is introduced; that is, the length of time necessary for the number of atoms of a nuclide to decay to one-half the initial number. In this case, $\frac{1}{2}N_0$ is substituted for N_p in Eq. (13.3), and then the half-life, $t_{1/2}$, is given by

$$t_{1/2} = \frac{0.693}{\lambda} \tag{13.4}$$

Thus the half-life for the decay of ^{14}C is 5730 a, and for ^{147}Sm is 1.06×10^{11} a. It is instructive to compare the half-lives given in Table 13.1, keeping in mind that the age of the Earth is 4.55×10^9 a. Since the formation of the Earth, approximately half the original ^{238}U has decayed, but very little ^{147}Sm has decayed. Isotopes that have a half-life considerably greater than the age of the Earth do not make accurate radiometric clocks, except for very old rocks, because of the analytical uncertainty associated with measuring extremely small amounts of daughter isotope. Conversely, isotopes with very short half-lives can be used for dating only young rocks, because of difficulties in accurately measuring extremely small amounts of parent isotope in old rocks.

Equation (13.3) is of no practical use because we cannot measure N_0, but we do know that each time a parent nuclide decays, it produces a daughter nuclide. Therefore, $N_0 = N_p + N_d$, where N_d is the present number of atoms of daughter nuclide. Substituting this into Eq. (13.3) and eliminating N_0, we obtain

$$\frac{N_p}{N_p + N_d} = e^{-\lambda t}$$

which, on rearranging, gives

$$N_d = N_p\left(e^{+\lambda t} - 1\right) \tag{13.5}$$

For an approximate solution to this equation, the term $e^{\lambda t}$ can be expanded in a power series as $1 + \lambda t + (\lambda t)^2/2! + (\lambda t)^3/3! + \cdots + (\lambda t)^n/n!$. Because λ is so small (about 10^{-10} a^{-1}), only the first two terms of the expansion are significant. Substituting these into Eq. (13.5) gives the approximate relation

$$N_d \approx N_p \lambda t \tag{13.6}$$

Before using Eq. (13.5) to determine the absolute age of a rock, we must consider the possibility that although radioactive decay has produced daughter nuclides from the parent, some daughter nuclide may have been present in the environment prior to the beginning of the decay of the system under consideration. In such a case, the present number of atoms of the daughter nuclide would be the sum of the initial number present, N_d^i, and those produced by decay from the parent in the system. We can rewrite Eq. (13.5) as

$$N_d = N_d^i + N_p\left(e^{\lambda t} - 1\right) \tag{13.7}$$

This equation, then, allows us to determine the age of a rock if the numbers of atoms of parent and daughter nuclides can be determined. Although this can be done, in practice it is more convenient and more accurate, with a mass spectrometer, to measure isotopic ratios. For example, this can be done if the daughter element has another isotope that is stable and not affected by radioactive decay. We can then divide both sides of Eq. (13.7) by the number of atoms of the stable isotope, N_s, and obtain

$$\frac{N_d}{N_s} = \frac{N_d^i}{N_s} + \frac{N_p}{N_s}\left(e^{\lambda t} - 1\right) \tag{13.8}$$

This, then, is the working equation for determining absolute ages from measured isotopic ratios involving parent, daughter, and stable nuclides.

Equation (13.8) contains two unknowns, the initial ratio, N_d^i/N_s, and the age, t. To solve for either of these unknowns requires two equations. These could be obtained by analyzing two different minerals that are known to have crystallized at the same time or analyzing two different rocks that had a common source and time of origin. There are special cases where the initial ratio is known, in which case an age can be obtained from a single analysis. We will examine examples of both of these cases.

Equation (13.8) is also the equation of a straight line (Fig. 13.2). It has an intercept on the ordinate of N_d^i/N_s, the initial ratio, and a slope of $(e^{\lambda t}-1)$, which is proportional to the age. If a number of different minerals crystallize together, in a magma for example, each contains the same initial isotopic ratio, that of the magma, because there can be no fractionation of isotopes between phases at high temperature, especially of the heavy isotopes. Each mineral starts life, then, with the same isotopic ratio of N_d^i/N_s; that is, they lie on the horizontal dashed line in Figure 13.2. Each mineral, however, contains different quantities of the radioactive parent (points along the dashed line). With time, various amounts of daughter product are generated, depending on the amount of parent, so the initial ratio is added to by an amount $(N_p/N_s)(e^{\lambda t}-1)$. At any instant in time, the isotopic ratio in each of the minerals lies on a sloping straight line such as that shown in Figure 13.2. With increasing time, the slope of the line becomes steeper, but its intercept on the vertical axis remains the same.

In practice, the initial isotopic ratio is unknown. If, however, geological evidence indicates that a group of minerals was formed together at the same time, their isotopic ratios, when plotted in a diagram such as that of Figure 13.2, would

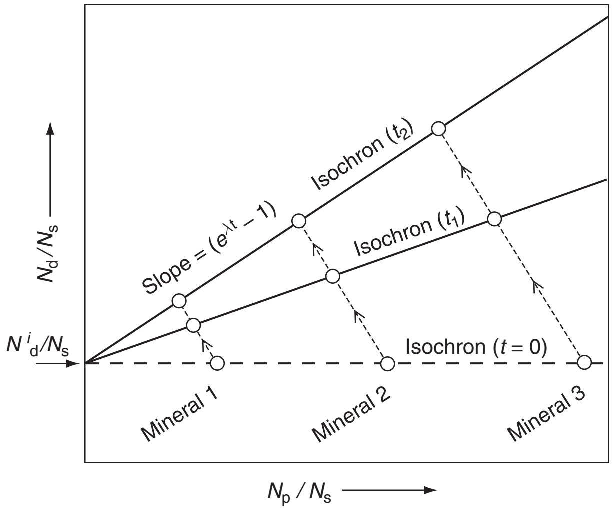


Fig. 13.2 Plot of Eq. (13.8) where N_p is the number of atoms of the parent nuclide, N_d the number of atoms of the daughter element, and N_s the number of atoms of a nonradiogenic isotope of this same element. Three minerals formed in a rock at the same time have different concentrations of the parent nuclide but all have the same initial ratio of N^i_d/N_s. With time, the parent nuclide decays, and the amount of daughter product formed is proportional to the amount of parent present. At any time (*t*), the minerals plot on a straight line known as an isochron; the age can be determined from the slope of this line.

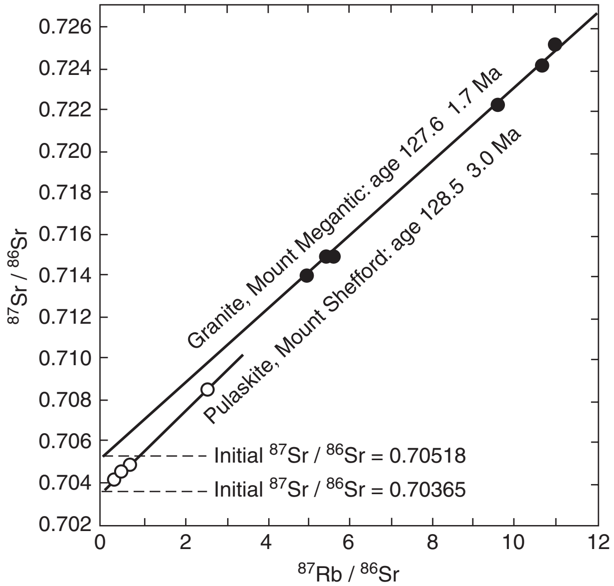


Fig. 13.3 Rb–Sr isochrons plotted through data for rocks from two Monteregian intrusions, Quebec. The slopes of the two lines are similar; consequently, the intrusions have similar ages. Granitic rocks of Mount Megantic, however, have a higher initial $^{87}Sr/^{86}Sr$ ratio than do nepheline syenites (pulaskite) of Mount Shefford, indicating that Megantic rocks assimilated larger amounts of crustal rocks, which normally have a high $^{87}Sr/^{86}Sr$ ratio. (After Eby, 1984.)

be expected to lie on a straight line, from which could be determined both the initial isotopic ratio and the absolute age. Such a line is known as an *isochron*, because all points on it have the same age. Isochrons can also be constructed using whole-rock isotopic analyses of rocks that have a range of compositions and that field evidence indicates were formed from a common source at approximately the same time – as, for example, a suite of differentiated rocks within a layered intrusion.

The *rubidium–strontium method* of absolute dating is the most common technique making use of this type of analysis of the data. ^{87}Sr is formed from the decay of ^{87}Rb, but in any geological environment there is already some ^{87}Sr present, formed from earlier decay of ^{87}Rb or inherited from the formation of the solar system. Isotopic ratios of ^{87}Sr and ^{87}Rb are measured against the amount of the stable isotope ^{86}Sr, and an isochron fitted to the following equation:

$$\frac{^{87}\mathrm{Sr}}{^{86}\mathrm{Sr}} = \frac{^{87}\mathrm{Sr}^{\mathrm{i}}}{^{86}\mathrm{Sr}} + \frac{^{87}\mathrm{Rb}}{^{86}\mathrm{Sr}}\left(e^{\lambda t} - 1\right) \tag{13.9}$$

where $^{87}\mathrm{Sr}^{\mathrm{i}}/^{86}\mathrm{Sr}$ is the initial ratio of these isotopes and λ is 1.42×10^{-11} a^{-1}. Figure 13.3 shows two typical isochrons fitted to data from two Monteregian intrusions in southern Quebec (Eby, 1984). The slopes of the two isochrons are similar, and therefore the two intrusions have approximately the same age. The difference between the two isochrons is in their initial ^{87}Sr/^{86}Sr ratios, that for the granite from Mount Megantic having a value of 0.70518 and that for the pulaskite from Mount Shefford having a value of 0.70365. As will be shown in the next section, these differences in initial isotopic ratio can be used to distinguish mantle and crustal sources of magma, and, in the case of mantle-derived magmas, the amount of crustal contamination.

In addition to providing information on the initial ^{87}Sr/^{86}Sr ratios, the Rb–Sr method of dating is particularly useful because Rb and Sr enter common rock-forming minerals, such as micas, feldspars, and amphiboles, substituting for K and Ca, respectively. Recrystallization of a rock is likely to cause radiogenic strontium to be expelled from these minerals because this strontium would have originally been present as ^{87}Rb substituting for K. These minerals, then, will have their radioactive clocks reset by recrystallization. The expelled ^{87}Sr, however, will enter nearby calcium-bearing minerals, such as apatite and plagioclase, so that the whole-rock Rb–Sr data will be unchanged. In this way, Rb–Sr dating, especially of large whole-rock samples, is able to see through metamorphic events (Fig. 13.4). Care must be taken, however, because whole-rock data can be affected by fluxes of fluids. In addition, dating of young material is unreliable unless samples contain high concentrations of Rb, because of the long half-life ($t_{1/2}$=4.89 × 10^{10} a) of the rubidium decay.

The *uranium–lead method* of absolute dating is useful because the two decay schemes, ^{238}U to ^{206}Pb and ^{235}U to ^{207}Pb, which have different decay constants, provide independent measures of the age of a sample. Thus, if uranium or lead has been removed from or added to a sample, the ages determined from the two methods will not agree; that is, the ages will be *discordant*. If, on the other hand, the sample has remained closed, the two ages will be *concordant*

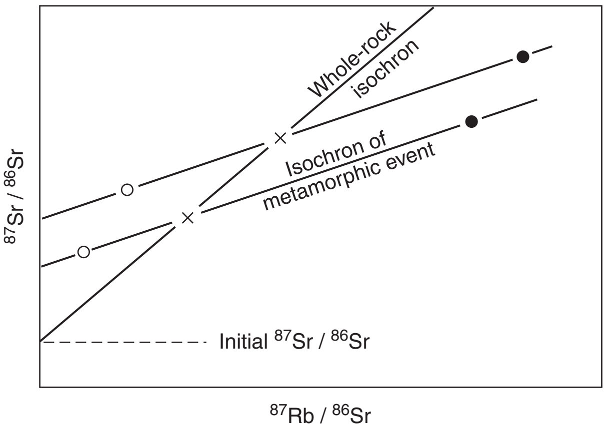


Fig. 13.4 Large-sample, whole-rock analyses (crosses) for Rb and Sr isotopes can give reliable isochrons, despite metamorphic overprinting, if Rb and Sr have not moved out of the volume sampled. Local redistribution of Rb and Sr within this volume between minerals of high Rb (solid circles) and low Rb (open circles) allows isochrons to be constructed that give the age of the metamorphism.

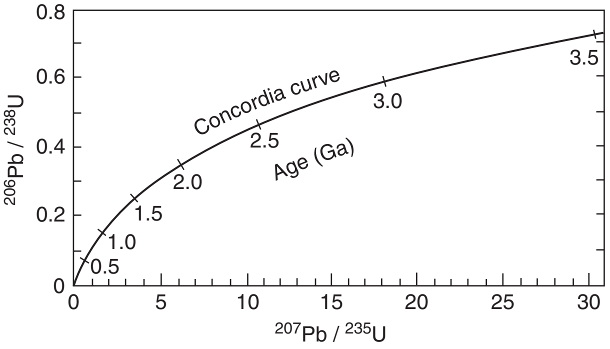


Fig. 13.5 Plot of Eq. (13.12) showing the variation in the $^{206}Pb/^{238}U$ and $^{207}Pb/^{235}U$ ratios in a system that has remained closed to U and Pb. Ages in Ga are marked on the concordia curve.

Using Eq. (13.5), we can write for the two decay schemes

$$^{206}\mathrm{Pb} = {}^{238}\mathrm{U}\left(e^{\lambda_{238}t} - 1\right) \tag{13.10}$$

and

$$^{207}\mathrm{Pb} = {}^{235}\mathrm{U}\left(e^{\lambda_{235}t} - 1\right) \tag{13.11}$$

where $\lambda_{238}=1.55125 \times 10^{-10}\ \mathrm{a}^{-1}$ and $\lambda_{235}=9.8485 \times 10^{-10}\ \mathrm{a}^{-1}$. The slower decay rate of ^{238}U relative to that of ^{235}U means that the ratio of $^{206}Pb/^{238}U$ grows more slowly than that of $^{207}Pb/^{235}U$ during the early stages of decay. The relative changes in these isotopic ratios can be found simply by equating Eqs. (13.10) and (13.11) through a common time. This gives

$$\frac{^{206}\mathrm{Pb}}{^{238}\mathrm{U}} = \exp\left[\frac{\lambda_{238}}{\lambda_{235}}\ln\left(1 + \frac{^{207}\mathrm{Pb}}{^{235}\mathrm{U}}\right)\right] - 1 \tag{13.12}$$

Equation (13.12), which is plotted in Figure 13.5, shows how the isotopic ratios of $^{206}Pb/^{238}U$ and $^{207}Pb/^{235}U$ change with time in a completely closed system. This line is known as the *concordia* curve, because ages determined by either Eq. (13.10) or (13.11) on a closed system give the same (concordant) results and thus plot on the same line. The plot in Figure 13.5 is known as a concordia diagram, and is the most commonly used graph for displaying U–Pb age data.

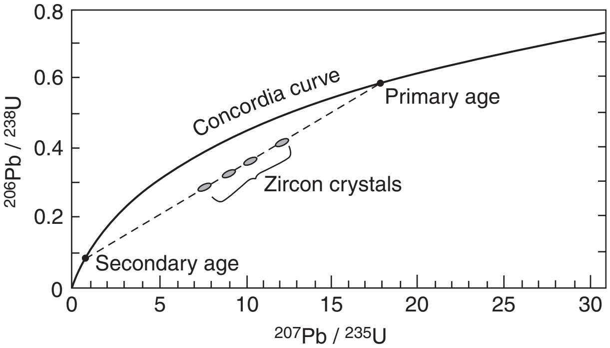


Fig. 13.6 Many rocks and minerals, such as the zircon crystals shown in this diagram, do not plot on the concordia curve; that is, the ratios of $^{206}Pb/^{238}U$ and $^{207}Pb/^{235}U$ are not concordant and indicate that the system has been disturbed. One interpretation of the linear array exhibited by the zircons in this diagram is that they formed 3.0 Ga ago but were later disturbed by an event 0.6 Ga ago.

Many geological systems have not remained closed since their formation, and because of the two different decay rates, any disturbance of the system results in different ages being determined from the two decay schemes. Because the ages do not match, they are said to be *discordant*. In the concordia diagram, these discordant samples do not lie on the concordia curve, but instead typically plot to the right of it (Fig. 13.6). Discordant samples that are genetically related commonly form a linear array in the concordia diagram (Fig. 13.6). Such a plot can contain considerable information about the history of a rock, but the interpretation is not always unambiguous and must be supported with sound geological field evidence.

The simplest and most common explanation for a linear array of discordant ages is that some initial age, given by the older intercept of the linear array with the concordia line, was disturbed by an event that occurred at the time given by the younger intercept of the linear array with the concordia line. For example, in Figure 13.6, the discordant ages shown are typical of what is obtained from a population of zircon crystals separated from a single large sample. Zircon is a particularly good mineral for absolute dating because it takes uranium into its structure but excludes lead. In addition, it has low diffusion rates and commonly preserves complex zoning, which can be analyzed using modern microanalytical tools, such as the ion-beam probe. The intercept of the array of zircon crystals in Figure 13.6 with the concordia line at 3.0 Ga could be the time of crystallization of these zircon crystals from a magma, whereas the secondary age of 0.6 Ga could be a metamorphic event which partially but not completely reset the U–Pb clocks. Numerous other, more complicated interpretations can be advanced, if supported with appropriate geological evidence. The ability of the U–Pb method to record the timing of successive events has made it a particularly powerful tool in unraveling complicated geologic histories.

The *potassium–argon method* of dating differs from other common methods by involving a decay product that is an inert gas. Even at moderately low temperatures (see discussion below), this gas is a fugitive component and is typically not incorporated in minerals. Thus a newly formed mineral contains no argon to begin with, but with time, ^{40}K decays slowly to ^{40}Ar; this argon remains in place as long as the system is not disturbed. The method, in principle, then, is not affected by initial isotopic ratios, as is the Rb–Sr method.

The K–Ar method is complicated slightly by the fact that ^{40}K decays both to ^{40}Ar and ^{40}Ca. The decay constant for ^{40}K (λ_K) of 5.543×10^{-10} a^{-1} is actually the sum of the two decay constants, one from the production of ^{40}Ar ($\lambda_{Ar}=0.581 \times 10^{-10}$ a^{-1}) and the other from the production of ^{40}Ca ($\lambda_{Ca}= 4.962 \times 10^{-10}$ a^{-1}). According to Eq. (13.5), the amount of decay product produced from ^{40}K is $^{40}\mathrm{K}\left(e^{\lambda_K t}-1\right)$. The decay products consist of both argon and calcium, which are present in amounts that are proportional to their respective decay rates. Thus the fraction of the total decay products consisting of ^{40}Ar at time t is

$$^{40}\mathrm{Ar} = \frac{\lambda_{Ar}}{\lambda_K}\,^{40}\mathrm{K}\left(e^{\lambda_K t}-1\right) \tag{13.13}$$

which on rearranging gives

$$t = \frac{1}{\lambda_K}\ln\left(\frac{\lambda_K\,^{40}\mathrm{Ar}}{\lambda_{Ar}\,^{40}\mathrm{K}}+1\right) \tag{13.14}$$

This equation expresses the age of a sample in terms of an easily measured isotopic ratio, ^{40}Ar/^{40}K.

Absolute dates by the K–Ar method can be obtained on both minerals and whole rocks. These may indicate the time since the formation of a sample, but because most igneous and metamorphic rocks form at moderate to high temperatures, most dates indicate the length of time since the material dropped below some critical temperature at which diffusion of argon out of the sample became negligible. This temperature is known as the *blocking temperature*, and varies with the minerals involved, but is approximately 350 °C. For large intrusive bodies or deeply buried metamorphic rocks slow cooling may produce a considerable discrepancy between the true age of a rock and its cooling age obtained from the K–Ar method. Such discrepancies can be found between Rb–Sr and K–Ar ages on the same samples. Blocking temperatures for Rb–Sr are a couple of hundred degrees higher than those for K–Ar.

For an age determination by the K–Ar method to be accurate, the assumption that no radiogenic argon was present to begin with must be valid; also, no radiogenic argon that is produced in the mineral or rock can have escaped, nor can any radiogenic argon from an external source have been absorbed. Each of these can, in particular situations, be a source of serious error. Some environments definitely have an ambient fugacity of ^{40}Ar, which can give a newly formed rock a significant initial content of radiogenic argon. For example, Mesozoic diabase dikes in Liberia, West Africa, which are associated with the early opening of the Atlantic Ocean, give K–Ar ages that are Mesozoic where the dikes cut Paleozoic rocks, but the ages are very much older where they cut Precambrian rocks, which, because of their age, contain high contents of ^{40}Ar (Dalrymple *et al.*, 1975). Some minerals leak ^{40}Ar more easily than others do by diffusion through their structure or along prominent cleavages, and thus they give erroneously young ages. The common K-bearing rock-forming minerals can be arranged according to retentiveness of ^{40}Ar, starting with the least retentive, microcline, and passing up through biotite, sanidine, pyroxene, muscovite, and ending with amphibole, the most retentive. Coarse-grained rocks tend to be less retentive than fine-grained ones. Some minerals, such as beryl, biotite, chlorite, clinopyroxene, and partly kaolinized feldspar, absorb ^{40}Ar and thus give erroneously old ages.

Typically, the outer parts of mineral grains are affected by gain or loss of argon. In the standard K–Ar dating method, there is no way of knowing precisely from where in the grains the analyzed potassium and argon come – the analysis is simply of a bulk sample. In a modified technique that involves measuring the ^{40}Ar/^{39}Ar ratio, loosely bound argon on grain boundaries can be distinguished from that within the core of grains, and some of the problems of excess or lost argon can be eliminated. In the ^{40}Ar/^{39}Ar method, ^{40}K is not measured directly. Instead, it is first converted to ^{39}Ar by placing the sample in a flux of high-energy neutrons in a nuclear reactor. Measurement of ^{39}Ar, then, gives a measure of the ^{40}K content prior to bombardment. The sample is heated in a stepwise manner, and the ^{40}Ar/^{39}Ar released at each stage is determined. The daughter nuclide, ^{40}Ar, and the parent nuclide, ^{40}K, which is now in the form of ^{39}Ar, are released together as argon gas. The first argon released is the most weakly bound, probably coming from grain boundaries. At higher temperatures, more tightly bound argon from the core of grains is emitted. An age can be calculated from the ^{40}Ar/^{39}Ar ratio obtained at each step. If argon has been gained or lost, or the amount of potassium near grain boundaries has changed, variable ^{40}Ar/^{39}Ar ratios and ages would be obtained for the first few steps. With continued heating, however, the ^{40}Ar/^{39}Ar ratio and age should stabilize to produce a plateau on a graph of ^{40}Ar/^{39}Ar (or age) versus heating step (Fig. 13.7). Only when a significant plateau is obtained can any certainty be placed on the age determination.

Dating by the *samarium–neodymium method* has been made possible in recent years by increased precision in mass-spectrometric techniques. ^{147}Sm decays to ^{143}Nd so slowly ($t_{1/2}=1.06 \times 10^{11}$ a) that age determinations depend on measurements of very small changes in the amount of ^{143}Nd. The method, which is only suitable for ancient rocks, is similar to that for Rb–Sr. Isotopic ratios of ^{147}Sm/^{144}Nd and ^{143}Nd/^{144}Nd are determined in a number of samples, which are used to construct an isochron having the form

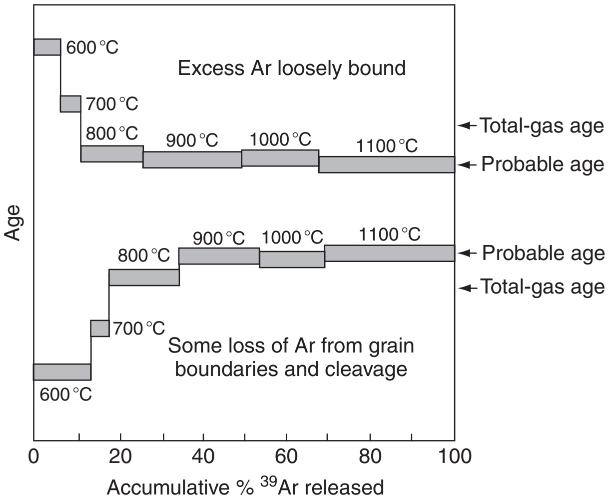


Fig. 13.7 ^{40}Ar/^{39}Ar incremental-heating-release spectra for two samples, one containing excess Ar and the other a deficiency of Ar. An age is calculated from the ^{40}Ar/^{39}Ar ratio in the gas released at each heating step. Only when the age given at each successive step remains constant (plateau) is the age likely to be reliable. Excess Ar, which might be absorbed on grain boundaries, is usually driven off at low temperature, giving erroneously old ages for the first steps. The age calculated from the total gas released (equivalent to the standard K–Ar method) is older than the plateau age if excess Ar is present. Loss of Ar from the sample gives erroneously young ages at the first heating steps, and the total-gas age is younger than the plateau age.

$$\frac{^{143}\mathrm{Nd}}{^{144}\mathrm{Nd}} = \frac{^{143}\mathrm{Nd^i}}{^{144}\mathrm{Nd}} + \frac{^{147}\mathrm{Sm}}{^{144}\mathrm{Nd}}\left(\mathrm{e}^{\lambda t} - 1\right) \qquad (13.15)$$

where ^{143}Nd$^{\mathrm{i}}$/^{144}Nd is the initial ratio of these isotopes. ^{144}Nd, the isotope against which the others are compared, is radioactive, but with a decay constant of 2.89×10^{-16} a^{-1} it is effectively a stable isotope. In ancient samples (>3.0 Ga), where the method has been used successfully, the initial ^{143}Nd$^{\mathrm{i}}$/^{144}Nd ratio is very close to that for chondritic meteorites, which are considered to be the most representative samples we have of the primordial solar nebular material from which the planet Earth was formed. These ancient rocks, then, formed from a source that had changed little since the beginning of the Earth (O'Nions *et al*., 1980). One distinct advantage of the Sm–Nd method is that both Sm and Nd, like the rest of the rare earth elements, are relatively immobile and remain *in situ* during weathering and even metamorphism. Rubidium and strontium, by contrast, are moved easily by solutions. Some of Earth's oldest rocks from Isua, Greenland, which are 3.8 Ga old, are metamorphosed and yet yield an excellent Sm–Nd isochron.

Dating of ancient rocks can also be done by the *rhenium–osmium method* (Shirey and Walker, 1998). Rhenium has two isotopes, ^{185}Re and ^{187}Re, the latter decaying by β emission to ^{187}Os with a half-life of 41.6 Ga. Recent improvements in analytical techniques have allowed the low concentrations of these isotopes to be detected with sufficient accuracy to be usable for dating purposes. The concentrations of ^{187}Re and ^{187}Os are measured relative to the concentration of the stable isotope ^{188}Os and the age is determined from an isochron fitted to the following equation:

$$\frac{^{187}\mathrm{Os}}{^{188}\mathrm{Os}} = \frac{^{187}\mathrm{Os^i}}{^{188}\mathrm{Os}} + \frac{^{187}\mathrm{Re}}{^{188}\mathrm{Os}}\left(\mathrm{e}^{\lambda t} - 1\right) \qquad (13.16)$$

where ^{187}Os$^{\mathrm{i}}$/^{188}Os is the initial ratio of these isotopes. The concentration of rhenium and osmium in most rocks is so low and with the low decay rate ($\lambda = 1.66 \times 10^{-11}$ a^{-1}) accurate ages are difficult to obtain on any but the oldest rocks. Both rhenium and osmium are extremely siderophilic elements; that is, they strongly prefer to be in a metal or sulfide phase. The Re–Os method is consequently one of the few ways of determining the age of platinum deposits. The siderophilic nature of these elements means that early in Earth's history they would have preferentially entered the core. As we will see below, the Re–Os dating method allows limits to be set on when Earth's core formed.

The *hafnium–tungsten method* of dating is also relevant to determining the time of core formation (Jacobsen, 2005). ^{182}Hf decays by β emission to ^{182}W with a half-life of only 9 Ma. With such a high decay rate ($\lambda = 7.7 \times 10^{-8}$ a^{-1}) no ^{182}Hf remains; that is, the decay process has run its course and is now extinct. However, during early Earth history, the process was still active. We can write a standard decay equation for ^{182}Hf to ^{182}W with respect to the abundance of the stable isotope ^{183}W as follows:

$$\frac{^{182}\mathrm{W}}{^{183}\mathrm{W}} = \frac{^{182}\mathrm{W^i}}{^{183}\mathrm{W}} + \frac{^{182}\mathrm{Hf}}{^{183}\mathrm{W}}\left(\mathrm{e}^{\lambda t} - 1\right) \qquad (13.17)$$

However, because of the high decay rate, ^{182}Hf/^{183}W is essentially zero, and $\mathrm{e}^{\lambda t}$ is approaching ∞. A more useful form of this equation can be written by making use of Eq. (13.3) from which we have

$$^{182}\mathrm{Hf} = {}^{182}\mathrm{Hf_i}\,\mathrm{e}^{-\lambda t}$$

where ^{182}Hf is the abundance after decaying for a time t from an initial abundance of ^{182}Hf$_{\mathrm{i}}$. If we divide both sides of this equation by the abundance of the stable isotope ^{180}Hf, we obtain

$$\frac{^{182}\mathrm{Hf}}{^{180}\mathrm{Hf}} = \frac{^{182}\mathrm{Hf_i}}{^{180}\mathrm{Hf}}\mathrm{e}^{-\lambda t} \quad \text{or} \quad {}^{182}\mathrm{Hf} = {}^{180}\mathrm{Hf}\frac{^{182}\mathrm{Hf_i}}{^{180}\mathrm{Hf}}\mathrm{e}^{-\lambda t}$$

This value of ^{182}Hf can be substituted into Eq. (13.17) to give

$$\frac{^{182}\mathrm{W}}{^{183}\mathrm{W}} = \frac{^{182}\mathrm{W^i}}{^{183}\mathrm{W}} + \frac{^{180}\mathrm{Hf}}{^{183}\mathrm{W}}\frac{^{182}\mathrm{Hf_i}}{^{180}\mathrm{Hf}}\mathrm{e}^{-\lambda t}\left(\mathrm{e}^{\lambda t} - 1\right)$$

which simplifies to

$$\frac{^{182}\mathrm{W}}{^{183}\mathrm{W}} = \frac{^{182}\mathrm{W^i}}{^{183}\mathrm{W}} + \frac{^{180}\mathrm{Hf}}{^{183}\mathrm{W}}\frac{^{182}\mathrm{Hf_i}}{^{180}\mathrm{Hf}}\left(1 - \mathrm{e}^{-\lambda t}\right) \qquad (13.18)$$

Because the system is now extinct, $\mathrm{e}^{-\lambda t} \approx 0$ and Eq. (13.18) becomes

$$\frac{^{182}\mathrm{W}}{^{183}\mathrm{W}} = \frac{^{182}\mathrm{W^i}}{^{183}\mathrm{W}} + \frac{^{180}\mathrm{Hf}}{^{183}\mathrm{W}}\frac{^{182}\mathrm{Hf_i}}{^{180}\mathrm{Hf}} \qquad (13.19)$$

In a plot of $^{182}W/^{183}W$ versus $^{180}Hf/^{183}W$, samples with a range of compositions can define a fossil isochron, where the slope is the initial ratio of $^{182}Hf_i/^{180}Hf$. Note that ^{180}Hf is a stable isotope and has not changed. We will have more to say about the hafnium–tungsten system in the next section.

Lutetium–hafnium provide still another pair of isotopes for dating old rocks (Hawkesworth and Kemp, 2006). ^{176}Lu decays by β emission to ^{176}Hf with a half-life of 36.77 Ga. Lutetium is commonly found in significant amounts in zircon crystals, which have long been recognized as ideal phases in which to preserve geological history because of the low diffusion rates of both parent and daughter isotopes. Recent advances in microanalytical techniques, using ion beam probes, laser ablation, and even microdrilling, have allowed detailed isotopic analyses to be done of zoned zircon crystals using the Lu–Hf and U–Pb methods to unravel complex histories preserved within single grains (Davidson *et al*., 2007).

Dating of zoned crystals of monazite ($(Ce,La,Th)PO_4$) has been obtained without isotopic analysis by using electron probe microanalysis (EPMA). Although EPMA cannot distinguish between isotopes, the electron beam is capable of rendering much higher resolution images than the other microanalytical techniques. Obtaining an age through EPMA involves a number of assumptions (Williams *et al*., 2007). First, the monazite is assumed to have contained no significant amount of lead initially. All lead that it now contains came from the radioactive decay of either ^{232}Th, ^{238}U, or ^{235}U. For example, the radioactive decay of ^{232}Th will generate ^{208}Pb. According to Eq. (13.5), the contribution of the ^{208}Pb to the total lead in the EPMA, which is typically expressed in parts per million by weight, is given by

$$\frac{^{208}\mathrm{Pb}}{208} = \frac{^{232}\mathrm{Th}}{232}\left(e^{\lambda^{232}t} - 1\right)$$

where λ^{232} is the decay constant for ^{232}Th (see Table 13.1). The ^{208}Pb and ^{232}Th are both divided by their respective atomic weights to give number of atoms. The EPMA gives only the total amount of uranium and not the amounts of ^{238}U and ^{235}U. The total uranium is consequently partitioned between these two isotopes in the proportion in which they occur in nature; that is, $^{238}U=99.28\%$ and $^{235}U=0.72\%$. The total lead in the sample will be the sum of the contributions from the ^{232}Th, ^{238}U, and ^{235}U, which can be written as

$$\begin{aligned} ^{\mathrm{Total}}\mathrm{Pb} = &\left[\frac{^{232}\mathrm{Th}}{232}\left(e^{\lambda^{232}t} - 1\right)\right]208 \\ &+ \left[\frac{^{\mathrm{Total}}\mathrm{U}}{238}0.9928\left(e^{\lambda^{238}t} - 1\right)\right]206 \\ &+ \left[\frac{^{\mathrm{Total}}\mathrm{U}}{235}0.0072\left(e^{\lambda^{235}t} - 1\right)\right]207 \end{aligned} \tag{13.20}$$

This equation allows the age (t) of the sample to be calculated from the analyses of the Pb, Th, and U.

Monazite crystals exhibit complex zoning patterns, which can be related to growth at different times and in different environments, or to compositional differences developed at the same time in different sectors. The preservation of zoning that commonly exhibits steep concentration gradients testifies to the low diffusion rates in monazite, even at moderately elevated temperatures. The closure temperature (temperature below which no significant diffusion occurs) is thought to be ~800 °C. Monazite, like zircon, is therefore an ideal phase in which to preserve age data.

Absolute dating by the *fission track method* differs from the previous ones in that it does not involve mass-spectrometric or conventional chemical analysis. Its principle is simple (Fleischer and Price, 1964; Fleischer *et al*., 1975; Gallagher *et al*., 1998). ^{238}U undergoes spontaneous fission at a rate of about $10^{-16}\ a^{-1}$. The massive charged fission products damage the structure of the host mineral along a short track (5 to 20 μm). These tracks can be made visible by etching with appropriate chemicals and their number counted under an ordinary optical microscope. For a given content of ^{238}U, the age is proportional to the number of tracks per unit area. ^{235}U and ^{232}Th also undergo fission, but this happens so rarely that they do not contribute significantly to the development of fission tracks. Even the fission rate of ^{238}U is very slow, but fission tracks record every fission event, and thus their measurement is not limited by analytical precision as in mass-spectrometric analyses or EPMA.

The procedure for determining an age by the fission track method first involves finding a mineral with an appropriate ^{238}U content to provide a statistically significant number of tracks. An appropriate uranium content depends on the age of the mineral. Young samples require high contents, but this same amount in older samples could produce too high a track density. The uranium should also preferably be distributed evenly through the sample. Many of the common rock-forming and accessory minerals have been used successfully. Mica, apatite, hornblende, sphene, zircon, and natural glasses are particularly good.

Once the number of fission tracks has been counted, the mineral's uranium content must be determined. This is done by first heating the sample to anneal out all of the fission tracks. Then the sample is placed, along with a piece of standard glass containing a known amount of uranium, in a nuclear reactor where both receive a dose of neutrons. These neutrons induce fission of ^{235}U, which produces new tracks that can be counted after appropriate etching.

The age of the mineral can be calculated from the three measured fission track densities: that produced by spontaneous fission in the original sample, ρ_s; that induced in the sample by the neutron flux in the reactor, ρ_i; and that produced in the dosimeter standard glass, ρ_D. The age is given by

$$\text{fission track age} = \frac{C\rho_s\rho_D}{\rho_i} \tag{13.21}$$

The constant, C, includes terms for the fraction of atoms that are ^{238}U, the total number of atoms in sampled volume, length of fission tracks, fraction of tracks that are efficiently etched, and the fission decay constant.

Reliable ages up to 1 Ga have been obtained by this method. Fission tracks, however, can disappear with time, especially if the sample is heated. The age determined by the method is therefore the time since the sample was last heated above a temperature at which annealing takes place. Different minerals anneal at different temperatures. For example, annealing removes fission tracks from apatite at ~120 °C, whereas this occurs in zircon at ~350 °C. Fission track dates on several minerals from the same rock can reveal the cooling history and erosional uplift rate of the sample. Fission track dating has been particularly useful when used in this way in conjunction with some other method of absolute dating that has a higher blocking temperature and can thus more closely record the true age of the rock.

13.4 EVOLUTION OF ISOTOPIC RESERVOIRS IN THE EARTH

The Earth has been affected throughout geologic time by processes that have differentiated its chemical constituents into the major units of the Earth, the core, mantle, and crust. As discussed in Section 6.1, Earth is believed to have formed with a chondritic bulk composition. Early in its history, however, sinking of metallic iron and other siderophile elements to form the core left a mantle composed predominantly of the silicate fraction of the primordial chondrites. The mantle's composition has subsequently been modified by extraction of small amounts of partial melt to form Earth's crust. The crust's present composition must be largely determined in some complex way by the time-integrated effect of this partial melting process. Phase diagrams and solid–liquid partition coefficients indicate that partial melts can have very different compositions from the solid being melted. Partial melting can therefore be an extremely efficient means of effecting chemical differentiation. Isotopes, at high temperature, however, are not fractionated during partial melting. But most radioactive isotopes have very different solid–liquid partition coefficients from their stable daughter isotopes. Consequently, partial melting does fractionate radioactive nuclides from their decay products. In this way, the isotopic nature of various parts of the Earth has changed throughout geologic time. For example, magmas extracted from the mantle today have different isotopic signatures from those extracted in the past. If we understand the evolution of the isotopic reservoirs, it may be possible to identify magmas derived from that source at any given time by their isotopic character.

Formation of the Earth's core must have occurred very early. Earth's oldest rocks show that a strong magnetic field was already present ~4 Ga ago, and this would have required a molten iron core. Can we determine when core formation occurred?

It will be recalled from Eq. (13.17) that ^{182}Hf decays rather rapidly to ^{182}W, with a half-life of only 9 Ma. Although these two isotopes have the same mass, their chemical behavior is quite different. Hafnium is a lithophile element; that is, it prefers to be in silicates. In contrast, tungsten is moderately siderophile and will preferentially enter metallic iron. Therefore, if the core formed after the ^{182}Hf had decayed to ^{182}W, this tungsten would have sunk along with the iron to form the core, leaving a mantle that would contain less ^{182}W than it would have had it maintained its chondritic composition. On the other hand, if the core formed before the ^{182}Hf had decayed to ^{182}W, this tungsten could not have been removed by the core-forming process. Consequently, the mantle would have a tungsten isotopic composition similar to that of chondrites. This in fact is what is found. Modeling of the core formation using the ^{182}Hf to ^{182}W decay scheme requires that Earth accreted rapidly in ~10 Ma and that core formation must have been complete by ~30 Ma (Jacobsen, 2005).

Rhenium and osmium are both strongly siderophilic, so they would have accompanied the iron into Earth's core when it formed. Iron meteorites, which represent the cores of planetary bodies from the asteroid belt, confirm that rhenium and osmium did enter the cores of these bodies, and their ages, based on rhenium–osmium dating (Eq. 13.16), indicate formation within the first 15 Ma of solar system history at 4.558 to 4.526 Ga (Shirey and Walker, 1998). Earth's core probably formed in this same interval.

Core formation early in Earth history would have sequestered most of the rhenium and osmium into the core. The mantle would then have been strongly depleted in these elements relative to the initial chondritic composition. Iron meteorites indicate that as molten iron solidifies rhenium and osmium both partition into the solid and consequently their concentration in the remaining liquid decreases. This effect is more pronounced for osmium, and consequently, the rhenium/osmium ratio increases in the liquid as crystallization progresses. Through the radioactive decay of ^{187}Re, the ^{187}Os/^{188}Os value would be much higher in the outer liquid core than in chondritic compositions. If there were equilibrium between the outer core and mantle, the ^{187}Os/^{188}Os values in the mantle might be expected to show elevated values relative to chondritic values. However, rocks derived from the mantle indicate values close to chondritic, which has led to the conclusion that this isotopic value was imposed on the mantle by a veneer of chondritic material that was accreted after core formation. Some rocks derived from mantle plumes do show enriched ^{187}Os/^{188}Os, which may reflect some input from the liquid outer core.

The Moon is generally accepted to have formed by a catastrophic impact of a Mars-sized object with Earth (Canup and Asphaug, 2001). Because the Moon has no substantial core, the impact is thought to have occurred after core formation on both Earth and the impactor. The Moon is consequently composed of mantle material from the Earth or the impactor. Jacobsen (2005) argues that the ^{182}W/^{183}W data best fit the impact occurring at ~32 Ma after formation of the Earth.

The decay of ^{146}Sm to ^{142}Nd by alpha emission is another short-lived decay system ($t_{1/2}$=103 Ma) that can provide precise age constraints on the early history of the Earth (Caro *et al.*, 2003). In this method, the radiogenic ^{142}Nd is measured with respect to the stable ^{144}Nd isotope. The early

Earth would be expected to have a chondritic value of ^{142}Nd/^{144}Nd. Analyses of 3.75-Ga-old metasediments from Isua, Greenland, indicate a slightly elevated ^{142}Nd/^{144}Nd. As discussed in more detail below, the extraction of partial melts from the mantle increases the mantle's Sm/Nd ratio. This, in turn, leads to higher ^{142}Nd/^{144}Nd through the decay of ^{146}Sm. The Isua analyses therefore indicate that the mantle had already begun to differentiate by 3.75 Ga.

Although the Isua rocks have slightly elevated ^{142}Nd/^{144}Nd, their initial ratio of ^{143}Nd/^{144}Nd (^{143}Nd is formed by the very slow decay of ^{147}Sm, Eq. (13.15)) is indistinguishable from that of chondritic meteorites. Other slightly younger Archean rocks from South Africa and Western Australia also have initial ratios that, for their particular ages, match the ratio found in chondritic meteorites. The Earth and chondritic meteorites must therefore have had not only the same initial ^{143}Nd^{i}/^{144}Nd ratio but also the same Sm/Nd ratio. Had the Earth had a higher Sm/Nd ratio, for example, the ^{143}Nd/^{144}Nd ratio would have grown at a faster rate than that in the meteorites. Chondritic meteorites and the Earth are therefore believed to have formed from similar primordial solar nebular material, which 4.55 Ga ago had a ^{143}Nd/^{144}Nd ratio of 0.50682 and a Sm/Nd weight ratio of 0.309 (present Sm/Nd=0.308).

Given the Earth's initial ^{143}Nd/^{144}Nd and Sm/Nd ratios, the evolution of the bulk Earth's ^{143}Nd/^{144}Nd ratio through geologic time can be calculated from Eq. (13.8) (see Problem 13.5). Figure 13.8 shows how the value of this ratio has evolved from 0.50682 at the time of the Earth's accretion to 0.51265 at the present. It is important to stress that this growth line is for the bulk Earth; actual Earth material will deviate from this depending on the fractionation of the isotopes during the Earth's differentiation. The sinking of iron and nickel to form the core early in the Earth's history would not have affected the Nd and Sm isotopic ratios because these rare earth elements are strongly lithophile and would simply have been excluded from the core and concentrated in overlying mantle. Differentiation of the remaining material into the crust and mantle would, however, have fractionated these elements, because some of the silicate phases involved in the fractionation incorporate different amounts of the various rare earth elements into their structures.

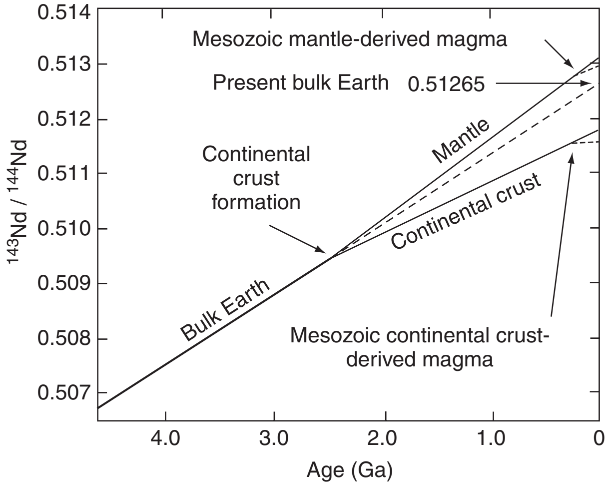


Fig. 13.8 Evolution of the Earth's ^{143}Nd/^{144}Nd ratio throughout geologic time based on an initial value of 0.50682 and an initial Sm/Nd ratio of 0.309. Formation of a continental crust 2.5 Ga ago would have fractionated Nd into the crust slightly more than it would have Sm. Following crust formation, the ^{143}Nd/^{144}Nd ratio would therefore have grown more rapidly in the mantle than in the crust. Magmas extracted from the mantle since the time of crust formation have higher ^{143}Nd/^{144}Nd ratios than crustal rocks (short dashed line). See text for discussion.

All of the rare earth elements (REE) have large ionic radii and preferentially enter the rather open structures of the silicates that form the crust of the Earth, rather than entering the denser structures of the minerals in the mantle. Such elements are referred to as *large ion lithophile* elements (LIL) or elements that are incompatible in the mantle (see Section 14.12). REE that exist in the mantle preferentially enter the more open structure of a melt, if present. Partial fusion and ascent of magma is one way, then, in which the REE can be fractionated into the crust. The REE, however, are not all fractionated to the same degree. With increasing atomic number, the ionic radii of the REE decrease – the so-called *lanthanide contraction*. Consequently, Nd has a slightly larger ionic radius than Sm has, which makes Nd slightly more incompatible in the mantle than is Sm. Thus, although both elements are strongly fractionated into the crust, Nd is fractionated more than Sm. At present, Nd is estimated to be about 25 times more abundant in the continental crust than in the mantle, whereas Sm is only about 16 times more abundant.

Although Sm and Nd are both fractionated into the crust, the stronger fractionation of Nd means that the Sm/Nd ratio is higher in the mantle than in the crust. Because ^{143}Nd is formed from the decay of ^{147}Sm, the higher Sm/Nd ratio in the mantle causes the ^{143}Nd/^{144}Nd ratio to grow more rapidly in the mantle than in the crust. Thus, once the Earth became differentiated into a mantle and crust, the growth line of the ^{143}Nd/^{144}Nd ratio split into two, one for the mantle at a level higher than that of the bulk Earth, and one for the continental crust at a level lower than that of the bulk Earth (Fig. 13.8).

Our next problem in understanding the evolution of the ^{143}Nd/^{144}Nd ratio in the Earth is to determine how and when the continental crust came into existence. Did it, for example, develop gradually throughout geologic time, or were there a number of distinct crust-building episodes, or possibly only one?

The discovery of a detrital zircon grain with an age of 4.4 Ga in a metaconglomerate in the Jack Hills of Western Australia has been used as evidence that the continental crust may have already started forming in the Hadean (Wilde *et al.*, 2001). These data were obtained using the U–Pb method with an ion microprobe, which allowed dates to be obtained on different parts of the crystal. In addition, oxygen isotopic data (to be discussed in Section 13.5) were obtained from this crystal, which suggest that Earth's surface was already cool and probably had oceans at this time. The origin of these crystals and the interpretation of their isotopic composition are hotly debated. The recent discovery of small inclusions of diamonds in the Jack Hills zircons, which have a range of ages from 3.058 to 4.252 Ga, adds to the problem of interpreting these crystals (Menneken *et al*., 2007). Diamond requires extremely high pressure to form, which can result from deep burial (>200 km) or meteorite impact. The fact that zircons containing these diamond inclusions have such a wide range of ages points to some process that was repeated over an extended period during early Earth history.

Although these zircon crystals may indicate that continental crust was already forming by 4.4 Ga ago, the amount of crust formed must have been small. Had there been any significant amount of continental crust formed prior to the development of the rocks at Isua, Greenland (3.75 Ga), the initial ^{143}Nd/^{144}Nd for these rocks would not match the primordial values found in meteorites. Indeed, the fact that all rocks older than ~3.0 Ga have initial ^{143}Nd/^{144}Nd ratios close to those of chondritic meteorites indicates that large amounts of continental crust could not have formed prior to that time. By the end of the Archean (2.5 Ga), however, large areas of continental crust had formed. Since then, more continental crust has undoubtedly formed, but much of it may be reworked Archean crust. For the purpose of our discussion, we will assume, as a first approximation, that the continental crust formed in a single episode 2.5 Ga ago. The fact that it may have had a more prolonged development does not greatly affect our conclusions on the isotopic evolution of the crust and mantle.

Until the formation of the continental crust, the ^{143}Nd/^{144}Nd ratio evolved along the line for the bulk Earth, reaching a value of 0.5094 just prior to the first major period of crust formation (Fig. 13.8). With formation of the crust about 2.5 Ga ago, the REE were fractionated into the crust, the lighter ones more so than the heavy ones. From then on, the ^{143}Nd/^{144}Nd ratio in the mantle evolved at a faster rate than that in the crust because of the higher Sm/Nd ratio in the mantle. Just prior to crust formation, the bulk Earth ^{147}Sm/^{144}Nd ratio had evolved to a value of 0.1900. Following crust formation, the ^{147}Sm/^{144}Nd ratio in that part of the mantle affected by differentiation probably became about 0.2245 and that of the crust became about 0.1435. With these ratios, ^{143}Nd/^{144}Nd ratios in the upper mantle and crust evolved along the separate lines shown in Figure 13.8 (see Problem 13.6).

It would be interesting to know what fraction of the mantle must have been involved in forming the continental crust, and whether any pristine mantle still exists. Today the largest volumes of mantle-derived magma are erupted as basalts along the mid-ocean ridges, the so-called MORBs. These come from a source that is strongly depleted in incompatible elements, and they have high Sm/Nd and consequently high ^{143}Nd/^{144}Nd ratios. Because of the enormous volumes of MORB erupted, this depleted source must constitute a significant volume of the mantle. Because it is the most strongly depleted source recognized, we can take it as possibly representing the mantle formed by the extraction of the crust. Knowing the volume of the continental crust and its isotopic composition, we can calculate the volume of mantle having a MORB-like isotopic composition that would be necessary to provide the bulk Earth composition. Such calculations indicate that the maximum mass of depleted mantle is ~40% of the total mass of the mantle. This indicates that as little as about 600 km of the upper mantle need have been involved in the formation of the continental crust. If the upper mantle is not everywhere as strongly depleted as the source of MORB, a proportionately greater depth of mantle would be involved, possibly as much as 90% (Allègre *et al*., 1983). Considering the thickness of the mantle, it appears likely that an upper part has been depleted in incompatible elements by the process of crust formation and a separate (decoupled) lower part has maintained its primordial composition to a large extent. Indeed, most isotopic models for the whole Earth require, in addition to the continental crust and a MORB-depleted mantle, a large primitive or enriched reservoir (presumably the lower mantle) to store many of the lithophile elements (Tolstikhin *et al*., 2006). This conclusion is supported by other geochemical arguments and geophysical evidence (see Section 23.2).

Arguments similar to those used for ^{143}Nd/^{144}Nd can be used to trace the evolution of the ^{87}Sr/^{86}Sr ratio, which is tied to the Rb/Sr ratio through the decay of ^{87}Rb to ^{87}Sr. Unfortunately, initial values in ancient rocks do not provide as accurate a measure of primordial values as do those of Nd, because of the mobility of Sr and to a lesser extent Rb during metamorphism. Values obtained from meteorites can be taken as representative of primordial material similar to that which formed the Earth. It is possible to deduce the bulk Earth value for ^{87}Sr/^{86}Sr from the Nd data, as will be explained below.

Both Rb and Sr are LIL elements, and thus they enter melts in the mantle and are transported into the crust. Rb, however, is far more incompatible in mantle solid phases than is Sr, and therefore Rb is more strongly fractionated into the continental crust. But ^{87}Rb is the isotope that decays to ^{87}Sr. Thus the ^{87}Sr/^{86}Sr ratio, in contrast to that of ^{143}Nd/^{144}Nd, is more radiogenic in the crust than in the mantle. In Figure 13.9, a continental crust, developed 2.5 Ga ago, can be seen to have the ^{87}Sr/^{86}Sr ratio evolve to much higher values than that of the bulk Earth or the remaining mantle, which is depleted by the formation of the crust (see Problem 13.8).

Figures 13.8 or 13.9 can be used to show how a mantle-derived rock differs isotopically from one formed from crustal material. Consider, for example, a melt derived from a depleted mantle at the beginning of the Mesozoic era (0.245 Ga). Its ^{143}Nd/^{144}Nd ratio would evolve from an initial value of 0.5127 to a present value of about 0.5129, the precise value depending on the Sm/Nd ratio developed in the melt during partial fusion (Fig. 13.8). Rocks derived from crustal material, by contrast, would be expected to have values less than 0.5120. In terms of the ^{87}Sr/^{86}Sr ratio (Fig. 13.9), this mantle-derived melt would evolve from an initial value of 0.7026 to a present value of possibly 0.7036, again a very different value from that of a melt derived from an ancient continental crust, which would be expected to have a value near 0.7190. If instead of being derived from a depleted mantle, a melt was formed from mantle that was brought from greater depth (possibly by a mantle plume) and thus

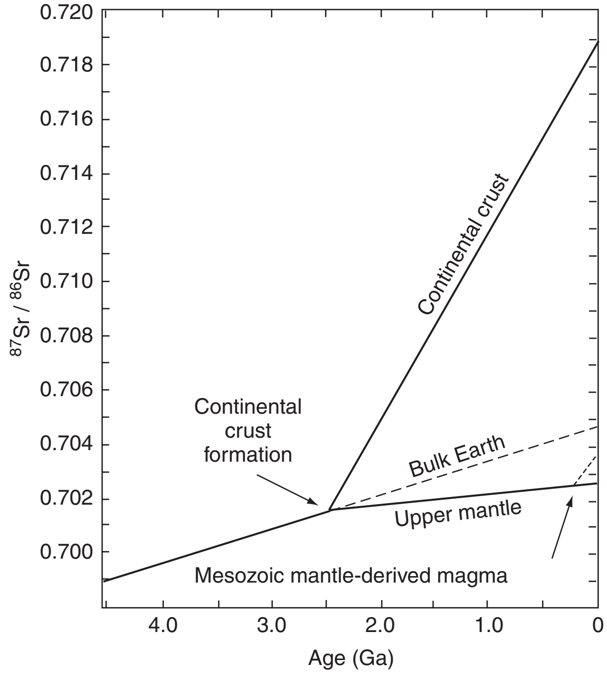


Fig. 13.9 Evolution of the Earth's $^{87}Sr/^{86}Sr$ ratio throughout geologic time. The stronger fractionation of Rb relative to Sr into magmas rising from the mantle to form the crust results in the $^{87}Sr/^{86}Sr$ ratio growing more rapidly in continental crust than in the bulk Earth or mantle. The continental crust is shown here as forming mainly at one time 2.5 Ga ago. Magmas derived from the mantle after crust formation evolve to lower $^{87}Sr/^{86}Sr$ ratios than do continental crustal rocks.

possibly isotopically closer to the composition of the bulk Earth, its isotopic signature would still be quite distinct from that of crustal-derived rocks. For magmas derived from the mantle at much earlier times, the task of distinguishing them from crustal ones obviously becomes more difficult and, in fact, impossible as their age approaches 2.5 Ga. This is why it is difficult to say precisely when a significant amount of continental crust was first formed.

Contamination of a mantle-derived magma by ancient continental crust is also clearly indicated by changes in the Nd and Sr isotopic ratios. Contamination is the simplest explanation for the difference between the initial $^{87}Sr/^{86}Sr$ ratios in the two Monteregian intrusive bodies plotted in Figure 13.3. Both intrusions have the same Cretaceous age; that is, their isochrons have identical slopes (about 128 Ma). The pulaskite (syenite with minor nepheline) from Mount Shefford has an initial $^{87}Sr/^{86}Sr$ ratio of 0.70365, which indicates a mantle source (see Fig. 13.9). The granite from Mount Megantic, on the other hand, has an initial value of 0.70518, which is well above the bulk Earth ratio for that age. The Megantic granite must therefore have incorporated some older crust. This intrusion was emplaced through a Grenville-age Precambrian (~1 Ga) basement and a thick sequence of folded lower Paleozoic rocks, any of which could have provided the isotopic contaminant. The Shefford intrusion has risen through a similar crust, but its mode of formation and emplacement must have been different from that of the Megantic granite for it to have preserved a mantle isotopic signature (and its modal nepheline).

The isotopic data for Nd and Sr can be conveniently combined into a single diagram as illustrated in Figure 13.10. The

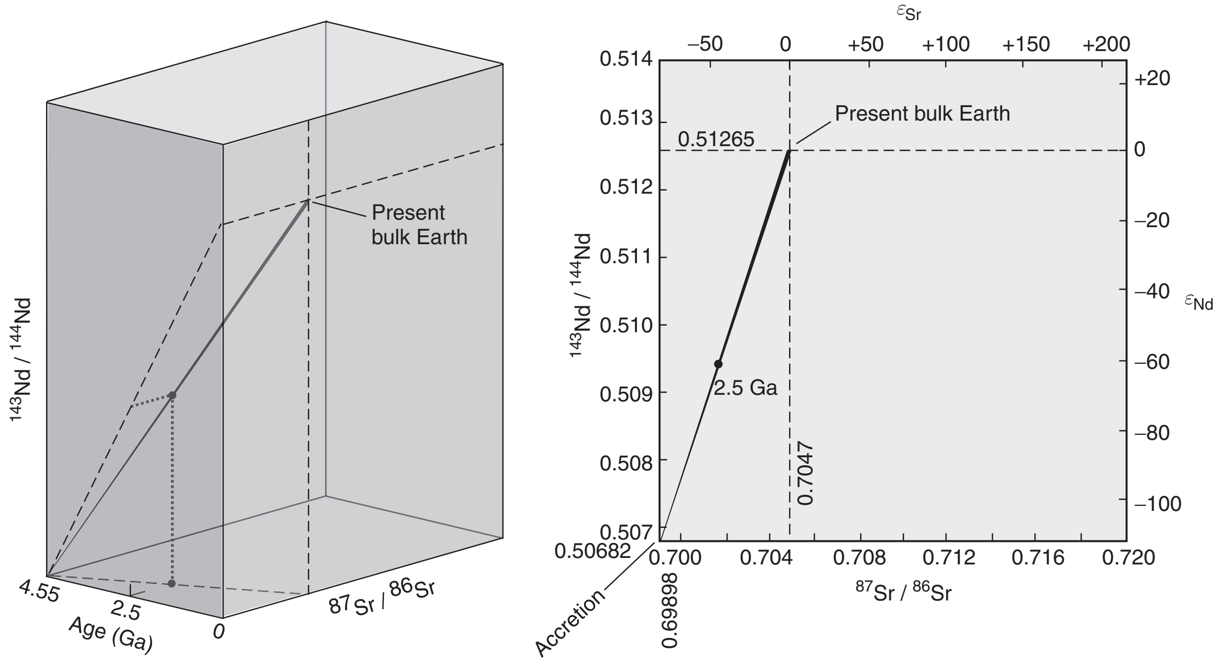


Fig. 13.10 Schematic representation of the evolution of the $^{87}Sr/^{86}Sr$ and $^{143}Nd/^{144}Nd$ ratios in the bulk Earth (solid heavy line) throughout geologic time. This line is projected onto a plane of constant time (the present) in which $^{143}Nd/^{144}Nd$ is plotted against $^{87}Sr/^{86}Sr$. See text for definition of ε values.

front left face and base of the block diagram are, respectively, Figures 13.8 and 13.9. For simplicity, only the line for the evolution of the bulk Earth has been included in this diagram along with a single point for 2.5 Ga. The evolution of the isotopic composition of the bulk Earth can be traced through the block diagram from its value at the time of accretion (4.55 Ga) to its value at the present day. The data can then be projected onto the front right face of the block diagram in terms of the ^{143}Nd/^{144}Nd and ^{87}Sr/^{86}Sr ratios. This face, which is shown separately to the right of the block diagram, is constructed for the present time, but other planes representing earlier times can be constructed. On this plane, the bulk Earth is represented by a point (+), but the evolutionary line leading to that point can be projected into this diagram, as shown by the dashed line. The isotopic ratio of the bulk Earth at any time in the past can also be read off this line, as indicated for the one age of 2.5 Ga.

Finally, some geochemists also display this diagram in terms of parameters known as the $\varepsilon_{\mathrm{Nd}}$ and $\varepsilon_{\mathrm{Sr}}$, which are the fractional deviations ($\times 10^4$) of the ^{143}Nd/^{144}Nd and ^{87}Sr/^{86}Sr ratios, respectively, from the values in the bulk Earth at the same time; that is,

$$\varepsilon_{\mathrm{Nd}} = \left[\frac{\left({}^{143}\mathrm{Nd}/{}^{144}\mathrm{Nd}\right)_{\mathrm{sample}} - \left({}^{143}\mathrm{Nd}/{}^{144}\mathrm{Nd}\right)_{\mathrm{bulk\ Earth}}}{\left({}^{143}\mathrm{Nd}/{}^{144}\mathrm{Nd}\right)_{\mathrm{bulk\ Earth}}}\right] \times 10^4 \tag{13.22}$$

The $\varepsilon_{\mathrm{Sr}}$ is defined in the same way but using the ^{87}Sr/^{86}Sr ratios. These values are also shown in Figure 13.10. Note that the ε values can be calculated for any desired age by adjusting the isotopic ratios for the decay of ^{147}Sm and ^{87}Rb.

A number of important rock types have been plotted in terms of their present ^{143}Nd/^{144}Nd and ^{87}Sr/^{86}Sr ratios in Figure 13.11. This diagram contains much information that needs to be worked through slowly if it is to be appreciated. To help orientate yourself, first locate the present composition of the bulk Earth (+) and the evolutionary line for the bulk Earth; these are identical to those in Figure 13.10. The lines marking the present bulk Earth ratios of ^{143}Nd/^{144}Nd and ^{87}Sr/^{86}Sr divide the diagram into four quadrants. Most mantle-derived rocks plot in the upper left quadrant ($+\varepsilon_{\mathrm{Nd}}$, $-\varepsilon_{\mathrm{Sr}}$), and most continental crustal rocks plot in the lower right quadrant ($-\varepsilon_{\mathrm{Nd}}$, $+\varepsilon_{\mathrm{Sr}}$). The quadrants are respectively depleted and enriched in LIL elements.

Passing through the bulk Earth composition is a remarkably linear array of isotopic ratios in rock types ranging from mid-ocean ridge basalt (MORB), oceanic island intraplate basalt (OI), kimberlite (K), garnet peridotite xenoliths (GP), and continental flood basalts (Co). All of these rocks, without doubt, have formed from magma originating in the mantle. Thus, they provide a sampling of mantle isotopic ratios, at least from those parts tapped by the magmas. For this reason, this range of isotopic values is commonly referred to as the *mantle array*. It indicates a strong negative correlation between Sr and Nd isotopes. Indeed, in light of the previous discussion of the relative degrees of fractionation of Sm–Nd and Rb–Sr, this variation is to be expected; that is, the continental crust was formed of material that had a lower Sm/Nd ratio and a higher Rb/Sr ratio than the bulk Earth. Upon decay of ^{147}Sm to ^{143}Nd and ^{87}Rb to ^{87}Sr, high values of

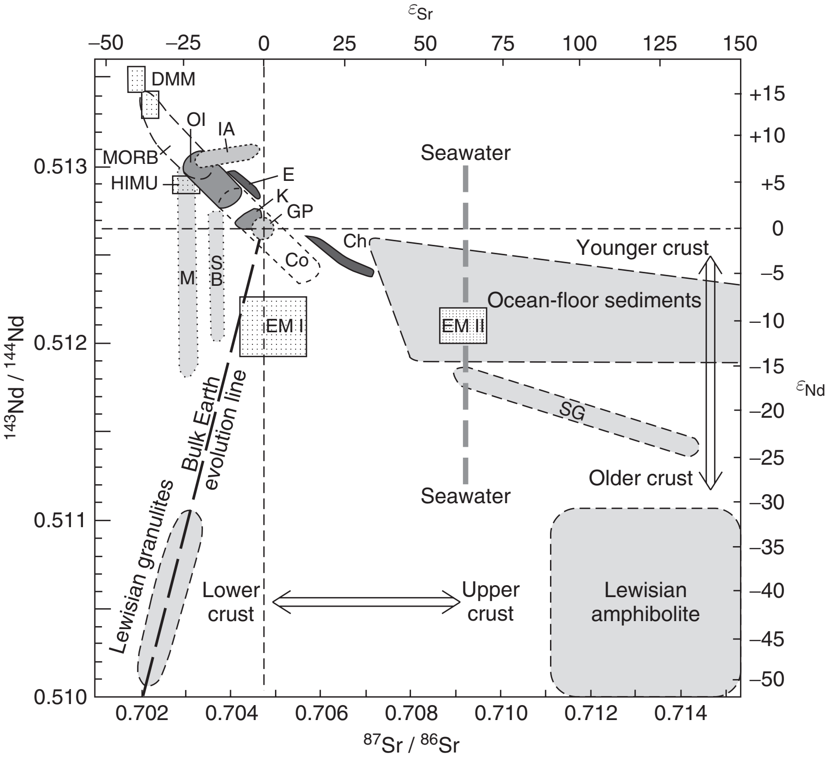


Fig. 13.11 Plot of important rock types in terms of ^{143}Nd/^{144}Nd and ^{87}Sr/^{86}Sr. Mid-ocean ridge basalt (MORB) and oceanic islands (OI) (O'Nions *et al.*, 1977; Norry and Fitton, 1983); island-arc rocks (IA) (DePaolo and Johnson, 1979); kimberlite (K), garnet peridotite mantle nodules (GP), and continental flood basalts (Co) (Hawkesworth *et al.*, 1983); central volcanic complexes of Mull (M), Skye basalts (SB), Skye granites (SG), and Lewisian granulites and amphibolites (Carter *et al.*, 1978); andesites of Ecuador (E) and Chile (Ch) (Hawkesworth *et al.*, 1979); and ocean floor sediments (White *et al.*, 1985). Possible mantle source components for oceanic magmas shown as stippled rectangles are from Zindler and Hart (1986): depleted MORB mantle (DMM), high μ (HIMU), enriched mantle I (EM I), and enriched mantle II (EM II). See text for discussion.

^{143}Nd/^{144}Nd correlate with low values of ^{87}Sr/^{86}Sr; that is, radiogenic Nd correlates with unradiogenic Sr, and vice versa.

The rock types that constitute the mantle array have formed from sources showing variable degrees of depletion in the LIL elements. This variation may occur laterally or vertically within the mantle. MORBs have formed from the most strongly depleted source, and their restricted compositional range indicates that a rather homogeneous, depleted mantle exists beneath all ocean ridges. Basalts with similar isotopic composition do rarely occur in continental extensional regions, such as the Basin and Range province of the western United States. Oceanic islands that occur within plates (i.e. those not associated with divergent plate boundaries) are composed of rocks with a much wider range of isotopic composition along the mantle array than MORB. Most lie within the region marked OI in Figure 13.11, but they can extend to compositions that are slightly enriched over that of the bulk Earth. Their source is not nearly as depleted as that of MORBs. Their range of composition may result from mixing of material derived from undepleted and depleted sources. Because oceanic islands are believed to develop over mantle plumes, their magmas probably tap a deeper part of the mantle than do MORBs, and thus they have more pristine compositions. Recall that some of these plume-related rocks also have elevated ^{187}Os/^{188}Os, which may indicate a contribution from as deep as the outer core. Alkaline basalts within continental plates have isotopic signatures identical to those of oceanic islands. Both must therefore have developed their signature at depths greater than the base of the thick continental lithosphere, which is so different from the oceanic lithosphere.

Continental flood basalts cover a range of composition along the mantle array, but most have compositions near that of the bulk Earth. This has had various interpretations. Their source could be in a part of the mantle that still preserves bulk Earth isotopic compositions. On the other hand, their source could be in a depleted mantle, but the magma assimilates enough continental crustal material to give a bulk Earth ratio. They could also have formed in a depleted mantle that becomes enriched by metasomatism prior to magmatism. There is, as yet, no resolution to this problem; perhaps all factors play a role in varying degrees in different regions. Kimberlites and their enclosed garnet peridotite xenoliths do have isotopic compositions that are very close to that of the bulk Earth. This implies that mantle with a bulk Earth isotopic composition does exist at depths of about 200 km beneath the continental crust in those regions where kimberlites originate.

The negative correlation between Nd and Sr isotopes in the mantle array and the mantle depletion in LIL elements, which it implies, must be the result of a long-lived phenomenon. For instance, had the depletion in ^{87}Rb occurred only recently, the decay product, ^{87}Sr, would still be present in the source region. But it is not, and the amount of ^{87}Sr that is found is consistent with there having been a deficiency of ^{87}Rb in the source for a considerable period of time. All indications are that this depletion occurred during the formation of the crust, which the Nd and Sr data indicate has a mean age of between 2.46 and 2.14 Ga (Allègre *et al*., 1983). Note that this is a mean age; it does not rule out the possibility that the crust-forming process started much earlier (Wilde *et al*., 2001).

It was previously mentioned that the bulk Earth composition for ^{87}Sr/^{86}Sr cannot be determined as accurately as that for ^{143}Nd/^{144}Nd because of the greater susceptibility of Rb and Sr to change during metamorphism. The strong negative correlation between these two isotopic ratios in the mantle array, however, allows the accurately known ^{143}Nd/^{144}Nd ratio to be used to determine the ^{87}Sr/^{86}Sr ratio, which at present is 0.7047.

Continental crustal rocks range more widely in isotopic composition than do mantle-derived ones, in part because of the strong elemental fractionation involved in crust-forming processes (partial melting, weathering, sedimentation), but also because of the relative volumes of the continental crust and mantle. Most crustal rocks plot in the lower right quadrant of Figure 13.11. In general, the greater the age of a segment of crust the farther to the lower right of the diagram it plots; this will be evident from inspection of the growth curves in Figures 13.8 and 13.9. Within the crust itself, there appears to be fractionation of some isotopes. In Chapter 1 it was shown, from heat flow calculations for the continental crust, that U, Th, and K must be concentrated toward the Earth's surface. Rubidium, which behaves geochemically much like K, also appears to be concentrated upward in the crust. Consequently, the ^{87}Sr/^{86}Sr ratio generally increases upward in the crust. No such crustal differentiation, however, affects Sm and Nd and thus the ^{143}Nd/^{144}Nd ratio is unaffected. The result in the plot of Figure 13.11 is a shift from right to left with increasing depth in the crust. This is illustrated by the ancient Lewisian rocks (2.95 Ga) of Scotland. The deepest exposed Lewisian rocks, which are in the granulite facies, have very low ^{87}Sr/^{86}Sr values (0.702 to 0.703). Shallower amphibolite facies rocks, on the other hand, have high ^{87}Sr/^{86}Sr values (0.711 to 0.720).

Because mantle-derived magmas and continental crustal rocks plot in very different positions in Figure 13.11, the effects of assimilation or magma mixing between these two should be readily evident. The isotopic variation exhibited by the mantle array is expected to result from partial melting or fractional crystallization. Isotopic variations in other directions imply mixing of different isotopic compositions. If the materials to be mixed have the same Sr/Nd ratios, mixing in Figure 13.11 results in a simple linear array between the two end-member compositions. If the Sr/Nd ratios are different in the two materials, intermediate compositions lie along curved lines (DePaolo and Wasserburg, 1979). In general, continental crustal materials have lower Sr/Nd ratios than do mantle-derived rocks, and as a result mixtures of mantle and crustal rocks lie along gently curved lines that are concave toward the upper right in Figure 13.11.

The basalts of the Islands of Mull and Skye on the northwest coast of Scotland formed during the early opening of the North Atlantic. Unlike lavas that have since erupted along the

Mid-Atlantic Ridge, these early basalts ascended through continental crust, which consists of Lewisian metamorphic rocks. The almost vertical trends of the Mull and Skye basalts in Figure 13.11 are not consistent with fractional crystallization. They do, instead, form a simple linear trend from a depleted magma toward a lower crustal Lewisian granulite composition. This is a clear case of lower crustal contamination of a mantle-derived magma.

On both Mull and Skye, there are large central igneous complexes containing granites, which some geologists have interpreted as having formed from melted crustal rocks or contamination of mantle-derived magmas. Only the granites of Skye (SG) are included in Figure 13.11, and these plot along a curved line between the mantle array and the most radiogenic Lewisian amphibolites. These granites, then, can be interpreted as a mixture of a mantle magma and crustal material. In contrast to the basalts, however, the granites assimilated higher-level crustal rocks.

Contamination of large bodies of mafic magma that have crystallized in the crust can also be demonstrated by their isotopic composition. The mid-Proterozoic (1.1 Ga) Kalka intrusion in central Australia, for example, was emplaced in quartzo-feldspathic granulite facies gneisses. Although some of the rocks of this intrusion have Nd and Sm isotopic compositions that plot very near the bulk Earth composition on the mantle array, most fall along a linear trend extending toward high ^{87}Sr/^{86}Sr ratios but maintaining approximately the bulk Earth ^{143}Nd/^{144}Nd ratio. This trend is not included in Figure 13.11, but it approximately follows the $\varepsilon_{\mathrm{Sr}}=0$ line to the right of the bulk Earth composition. This trend is directly toward the isotopic composition of the gneisses surrounding the intrusion, and is undoubtedly the product of crustal contamination of a mafic magma (Gray *et al*., 1981). This assimilation is believed to have been responsible for the copious precipitation of orthopyroxene in this intrusion. Other large mafic intrusions that have crystallized within the crust and that also contain abundant orthopyroxene, such as the Duluth Gabbro, the Bushveld Complex, and the Palisades Sill, have isotopic trends similar to that of the Kalka rocks, plotting to the right of the mantle array. Crustal assimilation seems likely in these intrusions as well.

The amount of crustal contamination of mantle-derived magmas indicated by the isotopic ratios of Nd and Sr is commonly much greater than is permissible based on the major element chemistry. The silica content of basaltic magmas, which is typically about 50 wt %, is changed significantly when continental crustal material, which typically contains >70 wt% SiO_2, is assimilated. Many basaltic magmas whose isotopic composition indicates there has been considerable assimilation of continental crustal material (high ^{87}Sr/^{86}Sr values) still contain almost pristine silica contents. The most likely explanation for this paradox is that contamination involves only a small fraction of the crustal rocks, but that fraction is particularly enriched in the isotopes concerned, especially Sr. This fraction might be derived, for example, by partial fusion of the crustal rocks or by extraction of hydrous fluids.

The igneous rocks developed above a subduction zone – the calcalkaline series – form one important group of rocks for which large-scale crustal contamination is commonly invoked to explain their composition. Five possible components may play roles in the formation of calcalkaline magmas: the mantle wedge overlying the subduction zone, subducted rocks of the oceanic crust (principally MORB), subducted sedimentary rocks, seawater, and crustal contaminants picked up during magma ascent. Each of these components is potentially identifiable from its isotopic signature. Thus, isotopic analysis of calcalkaline rocks may shed light on the origin of these rocks. We will examine the isotopic composition of calcalkaline rocks in two very different environments: first, where an island arc develops entirely between oceanic plates at a sufficient distance from continental areas that there can be no possible contribution from continental material, and second where a continental plate overrides an oceanic plate.

The New Britain island arc, which has been studied isotopically by DePaolo and Johnson (1979), is sufficiently far removed from continental masses to eliminate any possible continental contribution to these rocks. They nonetheless include rocks ranging from basalt to rhyolite. Their Nd and Sr isotopic compositions fall in a very small area in Figure 13.11(IA). They form a linear trend with a very slight positive slope extending from compositions on the mantle array at the depleted end of the oceanic island field to $\varepsilon_{\mathrm{Sr}}$ values of almost zero. Other island arcs have similar isotopic compositions. The trend is not one that can be explained by fractional crystallization or varying degrees of partial melting. It must involve mixing of various components. The primary source appears to be mantle that is not as depleted as that from which most MORBs are derived, but it is similar to material on anomalous ridge segments that are not as depleted as normal MORB or to some intraplate oceanic islands. The source is not simply melted average MORB, which is too depleted. The most likely major source for these rocks, then, is in the mantle wedge above the subducted plate.

The most striking feature of the isotopic composition of the island-arc rocks is their trend toward increasing ^{87}Sr/^{86}Sr ratios with almost constant ^{143}Nd/^{144}Nd ratios. Mixing of a mantle source with ocean floor sediments would not produce such a trend. Most ocean floor sediments are ultimately derived from continental areas and thus their isotopic composition places them in the lower right quadrant of Figure 13.11. Mixing of this material with mantle material would produce rocks that plot along a smooth curve with a negative slope. The most likely explanation for the isotopic trend is the incorporation of seawater in the magmas. At present, seawater has an ^{87}Sr/^{86}Sr ratio of 0.7092, but its content of Nd is so small ($<3\times10^{-5}$ ppm) that addition of seawater to any rock on the mantle array results in the isotopic composition simply being shifted to the right. Seawater can be incorporated directly as pore water trapped in the subducted plate, but it can also be stored in the ocean floor rocks in the form of hydrous alteration minerals. Its transport into the overlying mantle wedge upon subduction can be either as rising solutions or as partly melted MORB. The Nd and Sr isotopes imply that most of the magma generated in oceanic island arcs must come from the mantle wedge above the subduction zone, fluxed by seawater that is released from the downgoing oceanic plate. The slight positive slope to the island arc trend in Figure 13.11 could be produced from a small contribution of MORB that has been shifted to the right as a result of seawater alteration. A small amount of melting of ocean floor basalt may therefore occur in the subducted plate, especially if that ocean floor is young and still hot, as would occur if a mantle ridge were subducted. Plank and Langmuir (1993) have shown that trace element characteristics of some island arcs can be correlated with the composition of sediments being subducted. Some elements may be transferred from sediments into the mantle wedge by fluids released during subduction. The contribution of material from ocean-floor sediments, however, must be small, otherwise their strong isotopic signature would be detected.

Calcalkaline rocks developed on continental crust have a very different isotopic pattern from those developed in oceanic island arcs. As an example, data for the Andes have been included in Figure 13.11 (Hawkesworth *et al*., 1979). In Ecuador (E in Fig. 13.11), andesites have Nd and Sr isotopic compositions that plot slightly to the right of the mantle array, and they are more depleted than the bulk Earth composition. Andesites in Chile (Ch in Fig. 13.11) also plot to the right of the mantle array, but they are enriched relative to the bulk Earth composition. In contrast to the oceanic island arc rocks, these continental andesites form trends that roughly parallel the mantle array. Their trends can be interpreted as resulting from the mixing of three major components – a depleted mantle source, seawater, and a component that is enriched relative to the bulk Earth composition – but the interpretation is far from unequivocal.

Melting of subducted MORB does not appear to have played an important role in generating andesites. The trace element enrichment in andesites relative to that in MORB would require less than 5% partial melting of MORB for their generation. Yet 30% melting would be required to provide the major element chemistry. The trace element and isotopic compositions of the andesites are therefore best provided by a mantle source similar to that of the island arc rocks.

The simplest explanation for the andesites plotting to the right of the mantle array is the involvement of seawater. Further explanation of their trend, however, is complicated by numerous factors. First, trends paralleling the mantle array could be due to fractional crystallization, varying degrees of partial melting, or mixing of appropriate compositions. Clearly, compositions that are enriched relative to the bulk Earth composition are involved, but this could be subducted oceanic sediments, continental crustal material, or even mantle that was enriched in incompatible elements. There are, unfortunately, too many possible components for the Nd and Sr isotopes to define the sources uniquely. There is no question, however, that andesites consist predominantly of a mantle-derived component.

Finally, the Sm–Nd and Rb–Sr data can also be used to show that episodes of magmatic activity may be preceded by periods of mantle metasomatism. This has been well documented for intraplate alkaline magmatism (Norry and Fitton, 1983). In Figure 13.11, oceanic island rocks plot on the mantle array in the depleted quadrant; that is, they have relatively high ^{143}Nd/^{144}Nd and low ^{87}Sr/^{86}Sr ratios, which imply that the source regions must have had, for some considerable time, low relative concentrations of ^{144}Nd to ^{147}Sm and ^{87}Rb to ^{86}Sr. Many oceanic island and continental alkaline complexes, however, have relatively high ^{144}Nd and ^{87}Rb contents. Indeed, many of these have Sm/Nd (low) and Rb/Sr (high) ratios that, if present in their source regions for any length of time, would have generated low ^{143}Nd/^{144}Nd and high ^{87}Sr/^{86}Sr ratios. The fact that they have not can only mean that the more incompatible elements were recently introduced; that is, magmatism was preceded by a period of metasomatism in which the more incompatible elements, Nd and Rb, were introduced into a depleted mantle.

Another long-lived isotopic system that is commonly used to investigate the age and provenance of rocks involves the decay of ^{176}Lu to ^{176}Hf with a half-life of 35.7 Ga. Lutetium is the heaviest of the rare earth elements, whereas Hf is an element similar to zirconium, and commonly substitutes in the mineral zircon. Both the parent and daughter isotopes are measured with respect to the stable ^{177}Hf isotope. The ^{176}Hf/^{177}Hf ratio steadily grows with decay of the ^{176}Lu. However, when partial melts are extracted from the mantle, the Lu/Hf ratio in the mantle is increased several times over that found in the complementary crust, with the result that the ^{176}Hf/^{177}Hf ratio in the mantle grows more rapidly than in the crust. Its growth is similar to that of ^{143}Nd/^{144}Nd, but because of the stronger portioning of Lu/Hf between mantle and crust, the contrast in the ^{176}Hf/^{177}Hf is greater. When zircon crystallizes, hafnium is strongly partitioned into the crystal, but lutetium is not. Consequently, zircon crystals tend to preserve the ^{176}Hf/^{177}Hf ratio at the time of crystallization, which remains essentially unchanged due to the low Lu/Hf ratio. In contrast, garnet crystals take in large amounts of lutetium, while rejecting hafnium. Consequently, the Lu–Hf system provides an excellent way of dating garnets, which have the benefit of a relatively high closure temperature of ~600 °C. Lutetium and hafnium have also been used in the same way as samarium and neodymium to determine the source of magmas above subduction zones. Handley *et al*. (2007) use both Lu–Hf and Sm–Nd analyses to show that subduction-related magmas in Indonesia were derived primarily from a fertile mantle wedge to which a small amount of fluid was added from subducted altered ocean floor rocks and <1% pelagic sediment.

Although the remarkably uniform major element composition of ocean-floor basalts suggests that the mantle beneath oceans is homogeneous, isotopic studies of these basalts and oceanic islands show, unequivocally, that the mantle is heterogeneous. These basalts can have components derived from possibly as many as five different sources, each with its own characteristic isotopic signature (Zindler and Hart, 1986; Hofmann, 1997;

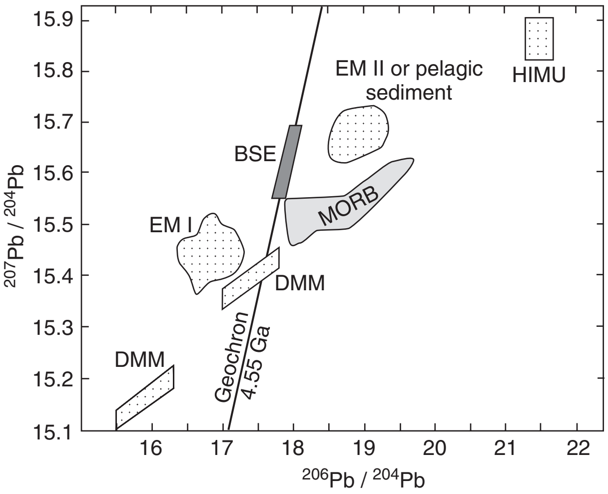


Fig. 13.12 Plot of possible mantle source components for oceanic magmas in terms of $^{207}Pb/^{204}Pb$ and $^{206}Pb/^{204}Pb$. Two possible depleted MORB mantle sources (DMM), high μ (HIMU), enriched mantle I (EM I), and enriched mantle II (EM II) are shown by stippled areas. Also shown is the typical range for MORBs. The composition of the bulk silicate Earth (BSE) is shown on a geochron for 4.55 Ga.

Shirey and Walker, 1998). We have already seen in Figure 13.11 that volcanic rocks have a wide range of compositions in terms of $^{143}Nd/^{144}Nd$ and $^{87}Sr/^{86}Sr$. Although some of these sources have been plotted in Figure 13.11, other isotopic systems must be involved to distinguish between some of them. By adding to this array the isotopic data from ^{238}U–^{206}Pb, ^{235}U–^{207}Pb, ^{232}Th–^{208}Pb, ^{176}Lu–^{176}Hf, and ^{187}Re–^{188}Os systems, the five mantle sources can be characterized.

The most abundant volcanic rock of the ocean's floor is the normal mid-ocean ridge basalt (N-type MORB). This is derived from a source that has undergone repeated melt removal, which has depleted it in incompatible elements; that is, elements that prefer to enter a melt rather than remain in the crystals of the mantle (olivine, pyroxene, spinel, garnet). This depleted MORB mantle (DMM), which is the most depleted of all the sources, is characterized by high $^{143}Nd/^{144}Nd$, and low $^{87}Sr/^{86}Sr$ and $^{206}Pb/^{204}Pb$ relative to the bulk silicate Earth (BSE), which is based on a chondritic parent from which the metallic core has separated. The DMM can be further divided into two possible sources, one with slightly higher $^{206}Pb/^{204}Pb$ than the other (Fig. 13.12). The other sources all show some enrichment in incompatible elements relative to the BSE (Figs. 13.11 and 13.12). One source has a very high U/Pb (= μ) and is therefore referred to as high μ (HIMU). Decay of the ^{238}U causes this source to have a high $^{206}Pb/^{204}Pb$ but otherwise it is similar to DMM in having high $^{143}Nd/^{144}Nd$ and relatively low $^{87}Sr/^{86}Sr$. HIMU is thought to be mantle that contains recycled oceanic crust. Enriched mantle I (EM I) has high $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$ for a given $^{206}Pb/^{204}Pb$, low $^{143}Nd/^{144}Nd$, but low $^{87}Sr/^{86}Sr$. The enrichment in this source may have been caused by metasomatizm or the recycling of lower continental crust. Enriched mantle II (EM II) is similar to EM I except that $^{87}Sr/^{86}Sr$ is high. This mantle may contain recycled upper continental crust.

The fifth source was referred to by Hart *et al*. (1992) as the focus zone (FOZO) and it is characterized by high $^3He/^4He$. This isotopic signature is found at hot spots, such as Hawaii and Iceland. The primitive ^3He is thought to come from a deep mantle source entrained in the mantle plume.

13.5 STABLE ISOTOPES

For most elements, the difference in mass between their isotopes is so small that isotopic fractionation is not likely to occur, especially at elevated temperatures within the Earth. Relative mass differences between isotopes of the light elements, however, are large, and consequently these isotopes can be fractionated by both chemical and physical processes. With the present resolution of mass-spectrometric techniques, isotopic fractionation can be detected in elements with atomic numbers up to 20 (Ca), but work has focused mainly on the stable isotopes of H, He, O, and S.

As with other isotopic analyses, measurements of the light stable isotopes are more conveniently and accurately made using isotopic ratios rather than absolute amounts. Analyses are expressed in terms of the *delta value*, *δ*, which is the per million (‰) fractional difference between the isotopic ratio in the sample and a standard. The heavier isotope is made the numerator. Thus for the $^{18}O/^{16}O$ ratio, the delta value is expressed as

$$\delta^{18}O = \frac{^{18}O/^{16}O_{\text{sample}} - {}^{18}O/^{16}O_{\text{standard}}}{^{18}O/^{16}O_{\text{standard}}} \times 10^3 \qquad (13.23)$$

Isotopic fractionation can result from chemical or physical processes, and it can reflect equilibrium or kinetic conditions. Take, for example, the simple exchange reaction of oxygen isotopes between quartz and water:

$$\begin{array}{llll} Si^{16}O_2 + & 2H_2^{18}O \rightleftharpoons & Si^{18}O_2 + & 2H_2^{16}O \\ \text{(quartz)} & \text{(water)} & \text{(quartz)} & \text{(water)} \end{array} \qquad (13.24)$$

This reaction does not go all the way to the right or to the left. Instead, some equilibrium distribution exists that is determined by the free-energy change of the reaction (see Chapter 9). Because the free-energy change is a function of temperature, the distribution of the isotopes between the two phases can serve as a geothermometer. This, then, is an example of equilibrium fractionation. In contrast, the more rapid evaporation of $H_2{}^{16}O$ relative to $H_2{}^{18}O$ from water is a kinetic fractionation. An important difference between the equilibrium and kinetic fractionation process is that kinetic fractionation is transient, and if carried to completion results in no fractionation. In the previous example, evaporation of water may at first favor the molecules with the lighter oxygen, but if all of the water is evaporated, the isotopic composition of the vapor will be the same as that of the initial water. At any temperature, however, there would be some equilibrium distribution of oxygen isotopes between the liquid and vapor phases that would be independent of time.

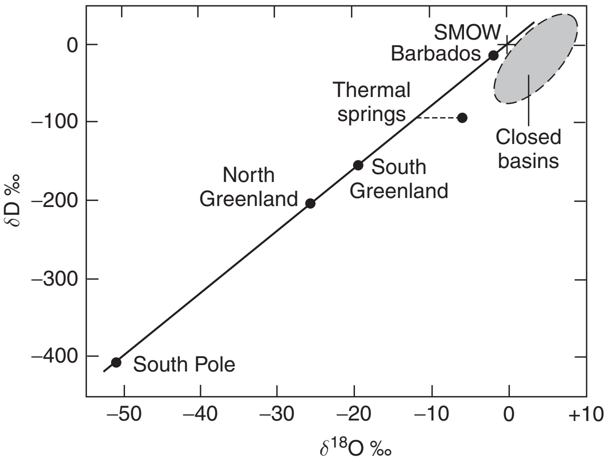


Fig. 13.13 Variation in the deuterium (^{2}H) to hydrogen (^{1}H) ratio (δD, see Eq. (13.23)) versus the δ^{18}O value for surface waters.

Fractionation has produced such varied isotopic compositions in different environments that in many cases the isotopes can be used to identify the source of material or the environment in which a rock formed. The following is a brief introduction to those stable isotopes that are commonly encountered in the petrologic literature. For a detailed discussion see Sharp (2006) or Valley (2001).

13.5.1 Oxygen and hydrogen isotopes

Isotopic ratios of ^{18}O/^{16}O and ^{2}H/^{1}H (D/H) are both measured with respect to the standard known as standard mean ocean water (SMOW). δ^{18}O values range from about −50 to +30 and δD values range from −400 to 0. A very strong correlation exists between the δ^{18}O and δD in natural waters and the latitude (Fig. 13.13); toward the poles, oxygen and hydrogen become progressively enriched in lighter isotopes. This is a consequence of the preferential precipitation of heavy isotopes from clouds. As clouds move away from the equatorial region toward the poles, they become progressively depleted in the heavier isotopes, which are removed by the rain. The same trend can be seen as a function of the distance clouds travel into a continent from an ocean. The line relating the δD and δ^{18}O in Figure 13.13 has a slope of 8, which is the value that would be predicted from the equilibrium fractionation of these isotopes between liquid and vapor. Two sources of natural surface water that fall off the main trend are closed basins and hot springs. Evaporation of light ^{16}O from water in closed basins increases the δ^{18}O value. Hot springs are enriched in ^{18}O relative to normal groundwater because of interaction with surrounding rocks, which invariably have higher δ^{18}O values.

One of the most spectacular successes in the use of oxygen isotopes was in showing that the Pleistocene glaciations could be correlated with the orbital mechanics of the Earth around the Sun, the so-called *Milankovitch cycles*. This correlation, which is documented in all introductory geology textbooks, is based on the oxygen isotopic composition of ocean floor sediments. During periods of major glacial advance, precipitation, which is always richer in ^{16}O than seawater, becomes trapped on continents in the form of ice, and consequently seawater becomes slightly enriched in ^{18}O. Ocean-floor sediment formed from pelagic organism reflect this elevated δ^{18}O, with the degree of enrichment reflecting the amount of ice on the continent and hence the severity of the glaciation. Profiles of δ^{18}O through the ocean-floor sediment are almost identical to Earth's calculated temporal change in heat flux from the Sun, which has cycles of 23 000, 41 000, and 100 000 years (Milankovitch cycles), due respectively to the precession and change in tilt of Earth's axis and the change in the eccentricity of its orbit about the Sun.

Typical δ^{18}O values for mantle-derived rocks (peridotites, gabbros, basalts, andesites, and syenites) and chondritic meteorites are from +5‰ to +7‰, whereas continental-derived rocks are higher and more variable. For example, Peck *et al*. (2000) found that the oxygen isotopes in zircon crystals in Archean rocks of the Canadian Shield had a remarkably restricted range of δ^{18}O of 5.77‰ ± 0.6‰, which suggests that they crystallized from magmas that had equilibrated with primitive mantle. In contrast, zircon crystals in Proterozoic rocks had higher and more variable δ^{18}O of 8.2‰ ± 1.7‰, which indicates that by that time material from a continental crust must have been transported to depth and incorporated into magmas. Surface processes, such as weathering, that are involved in forming sediments produce elevated δ^{18}O, and as a result continental rocks typically have higher δ^{18}O. Because most granites are formed from partially melted continental crust they have δ^{18}O ranging from +7‰ to +13‰. Rhyolites in oceanic island arcs, on the other hand, have δ^{18}O values similar to those in the associated basaltic rocks, which indicates that they differentiated from a common mantle source. δ^{18}O values for sedimentary rocks are positive, and for limestones can be as high as +30‰. Pristine igneous rocks do not have the negative values of δ^{18}O that characterize natural surface water. Interaction of igneous rocks with meteoric water can therefore be detected easily from the oxygen isotopes. For example, Tertiary basalts around the central complex of the Isle of Mull on the west coast of Scotland have had their δ^{18}O values lowered to −6.5‰ by convecting groundwater systems set up during the cooling of the central complex. Similarly, during the cooling of the Skaergaard intrusion, East Greenland, a convecting groundwater cell was set up by the cooling intrusion and meteoric water that was able to percolate easily through the surrounding fractured basalts exchanged oxygen with the Skaergaard rocks and lowered their δ^{18}O values to −5‰ (Norton and Taylor, 1979). Where the intrusion is in contact with the less fractured gneissic country rocks (see Fig. 15.15) the δ^{18}O values of the igneous rocks still retain their primary character (δ^{18}O > 5.7‰).

Oxygen isotopes provide one of the most reliable geothermometers, and, considering the abundance of oxygen in minerals, one of the most widely applicable, especially for rocks formed at low to moderate temperatures. Equation

(13.24) expresses the exchange reaction of ^{18}O and ^{16}O between quartz and water. Similar reactions can be written for all oxygen-bearing minerals. We can generalize such a reaction as

$$\mathrm{A^{16}O + H_2^{18}O \rightleftharpoons A^{18}O + H_2^{16}O} \tag{13.25}$$

where AO is some oxygen-bearing mineral. The equilibrium constant, *K*, for this reaction can be expressed as

$$K = \frac{\left(^{18}\mathrm{O_A}\right)\left(^{16}\mathrm{O_{H_2O}}\right)}{\left(^{16}\mathrm{O_A}\right)\left(^{18}\mathrm{O_{H_2O}}\right)} = \frac{\dfrac{^{18}\mathrm{O_A}}{^{16}\mathrm{O_A}}}{\dfrac{^{18}\mathrm{O_{H_2O}}}{^{16}\mathrm{O_{H_2O}}}} \tag{13.26}$$

where $^{18}\mathrm{O_A}$ is the concentration of ^{18}O in phase A. But, from Eq. (13.23) it follows that

$$\frac{^{18}\mathrm{O_A}}{^{16}\mathrm{O_A}} = \frac{^{18}\mathrm{O_{SMOW}}}{^{16}\mathrm{O_{SMOW}}}\left(\frac{^{18}\mathrm{O_A}}{1000} + 1\right)$$

Therefore,

$$K = \frac{\delta^{18}\mathrm{O_A} + 1000}{\delta^{18}\mathrm{O_{H_2O}} + 1000} \tag{13.27}$$

For most of the common *rock-forming* minerals, $K > 1$; that is, ^{18}O is concentrated in the mineral rather than in the water, with quartz effecting the largest fractionation. The one prominent exception is magnetite, which preferentially takes in the lighter isotope. The value of *K* varies with temperature ($K = e^{-\Delta G/RT}$), approaching a value of 1 at high temperatures for all mineral–water pairs.

If minerals and the fluids from which they crystallized could be analyzed for their oxygen isotopes, Eq. (13.27) would form the basis for a geothermometer. Unfortunately, the fluid phase is normally no longer available for analysis. We may still use Eq. (13.27), however, if two minerals crystallize together, preferably minerals that have very different equilibrium constants. Quartz and magnetite best satisfy this condition, but other mineral pairs can be used. For the quartz–magnetite pair, Eq. (13.27) can be rearranged as follows:

$$\left(\delta^{18}\mathrm{O_{H_2O}} + 1000\right) = \frac{\delta^{18}\mathrm{O_{Qz}} + 1000}{K_{\mathrm{Qz}}} \tag{13.28}$$

and

$$\left(\delta^{18}\mathrm{O_{H_2O}} + 1000\right) = \frac{\delta^{18}\mathrm{O_{Mt}} + 1000}{K_{\mathrm{Mt}}} \tag{13.29}$$

Combining these two equations, the delta value of the fluid is eliminated, leaving

$$\frac{\delta^{18}\mathrm{O_{Qz}}}{K_{\mathrm{Qz}}} = \frac{\delta^{18}\mathrm{O_{Mt}}}{K_{\mathrm{Mt}}} \tag{13.30}$$

Because the values of *K*, which are functions of temperature, are known from experiments and the delta values of quartz and magnetite can be measured mass-spectrometrically, the equation can be solved for temperature. The general thermometer expression valid for metamorphic and igneous temperatures is (e.g. Valley, 2001; Sharp, 2006)

$$1000 \ln \alpha_{A-B} = a_{\mathrm{o}} + \frac{a_1 \times 10^6}{T^2} \tag{13.31}$$

where a_0 is a constant, a_1 is the temperature coefficient of fractionation, *T* is in kelvin, and α_{A-B} is defined as

$$\alpha_{A-B} = \frac{K_A}{K_B} = \frac{1000 + \delta^{18}\mathrm{O}_A}{1000 + \delta^{18}\mathrm{O}_B}$$

for the two coexisting phases *A* and *B*. The a_0 term is usually taken as zero for fractionations between anhydrous minerals, and nonzero if hydrous minerals or water are involved. Thermometers are calibrated using experiments, empirical field data, or theoretical calculations. More complex thermometer expressions exist, but Eq. (13.31) has proven to be very useful for a wide range of geological applications.

Temperatures determined in this way record the temperature when the isotopes last equilibrated, which may be different from the temperature at which the minerals first formed. Isotopic ratios may later be disturbed by exchange with circulating groundwater. Normally, unless concordant temperatures are obtained from several different mineral pairs in the same rock, little reliance can be placed on a single measurement.

Geothermometry using oxygen isotopes has proved reliable over the entire range of normal crustal temperatures; that is, from magmatic to surface temperatures. It has been used successfully on igneous, metamorphic, sedimentary, and hydrothermal rocks, and has even been used to determine paleoclimatic temperatures.

13.5.2 Sulfur isotopes

The four isotopes of sulfur have masses of 32, 33, 34, and 36. Only the ratio of the two most abundant isotopes (^{34}S/^{32}S) is normally measured. The δ^{34}S is measured with respect to the sulfur in troilite (FeS) from the Canon Diablo meteorite. The delta values defined in this way cover a wide range of values in sulfides and sulfates from different environments. These values are of particular importance to the study of sulfide ore bodies, especially with regard to the light they may shed on the origin of these deposits.

Most δ^{34}S values of terrestrial sulfur average around the values found in meteorites, which probably closely approaches a primordial solar value. It is not surprising, therefore, that magmatic sulfides in rocks derived from the mantle have delta values near zero. Sulfates in evaporites formed from seawater, on the other hand, have delta values of about +20, which is similar to that of seawater. Even sulfates deposited in veins have much higher delta values (about +20) than coexisting sulfides. Some magmatic sulfide deposits in mafic igneous rocks have delta values as high as +10. The sulfide deposit at Norils'k, Russia, provides a good example. Here diabase intruded and assimilated sedimentary gypsum. This increased the delta value of the primary sulfur, and produced a magmatic sulfide deposit with anomalously heavy sulfur. Sulfides deposited in sediments by the action of bacteria can have delta values that range widely, from −30 to +30.

Fractionation of sulfur isotopes between coexisting sulfides (and sulfates) is significant and can be used as the basis of a geothermometer in the same way that oxygen isotopes are used. Sphalerite and galena, pyrite and chalcopyrite, pyrite and sphalerite, and sulfide and sulfate all provide sensitive geothermometers. Considerable care must be taken, however, to ensure that the mineral pairs analyzed did indeed crystallize together. Pyrite, for example, can crystallize over a very wide temperature range.

13.5.3 Helium isotopes

Helium, which has two isotopes with atomic masses 3 and 4, is an inert gas and is therefore not chemically bound in the Earth. Moreover, its mass is low enough that it is able to escape from the Earth into space. Helium in the atmosphere, consequently, must be replenished by outgassing from the Earth's interior. Within the Earth, helium has two principal sources: primordial helium trapped at the time of Earth accretion, and radiogenically produced helium, which because of the concentration of radioactive elements toward the Earth's surface, is produced largely at shallow depths. Both of these sources have characteristic isotopic signatures, which can serve as a means of identifying sources.

The isotopic composition of helium produced by radioactive decay is very different from that of primordial helium. The isotopic ratio of ^{3}He/^{4}He (referred to as R) in the primordial solar system, which was the result initially of nuclear fusion, can be estimated from its value in meteorites, which is about 3×10^{-4}. Radioactive decay, on the other hand, produces mainly ^{4}He through α decay. Some ^{3}He is produced by fission of ^{238}U and some by the interaction of α particles with nuclei of light elements. This interaction produces a neutron flux that can combine with lithium, for example, to produce ^{3}He. The reaction is as follows: ^{6}Li combines with a neutron to produce an α particle plus ^{3}H; through β emission the ^{3}H decays to ^{3}He. The combined effect of these various radioactive processes is to produce an R value of 2×10^{-8}, which is very much lower than the primordial ratio. Cosmic rays are also capable of generating helium, but this is not a significant source on Earth.

Rocks that have risen directly from deep in the mantle have R values of about 10^{-5}. In contrast, ancient granitic crustal rocks have R values of about 10^{-8}, which simply reflects their high content of radiogenic helium. The ratio in the atmosphere is 1.4×10^{-6}, which must result from a mixture of primordial and radiogenic helium. In general, the higher the R value, the larger must be the contribution from a deep source. MORB typically has R values from seven to nine times that of the atmosphere, whereas oceanic island values can be as much as 30 times larger. Hot springs in areas of recent volcanism also have high R values. At Yellowstone, for example, R values are relatively constant at seven times the atmospheric value within the caldera but drop off sharply outside it. Some values within the caldera are as high as 16 times atmospheric. Similar values are obtained in Icelandic hot springs. In recent volcanic areas, helium isotopes are therefore capable of indicating a mantle component. In older rocks, however, the primordial helium component is soon diluted and swamped by the radiogenic helium derived from surrounding crustal rocks; then the mantle signature is no longer detectable.

13.6 PROBLEMS

13.1 For the following radioactive decay reactions determine the type of decay process involved.

$$^{115}_{49}\mathrm{In} \rightarrow {}^{115}_{50}\mathrm{Sn} \quad {}^{142}_{58}\mathrm{Ce} \rightarrow {}^{138}_{56}\mathrm{Ba} \quad {}^{138}_{57}\mathrm{La} \rightarrow {}^{138}_{56}\mathrm{Ba}$$

13.2 ^{147}Sm decays to ^{143}Nd with a decay constant of 6.54×10^{-12} a^{-1}. ^{144}Nd is a stable isotope whose abundance in the bulk Earth has remained constant throughout time. If the present ^{147}Sm/^{144}Nd ratio in the bulk Earth is 0.1870, what would it have been at the time of Earth accretion (4.55 Ga)?

13.3 From the data presented in Figure 13.3, show that the isochron for the Megantic granite does correspond to an age of 127.6 Ma. The most Rb-rich analysis in this plot has an ^{87}Rb/^{86}Sr ratio of 11.04 and an ^{87}Sr/^{86}Sr ratio of 0.7252.

13.4 Using Eq. (13.12), construct a concordia curve in the ^{206}Pb/^{238}U versus ^{207}Pb/^{235}U diagram. Where in this diagram might zircon crystals plot that were initially formed in a 2.6-Ga-old granite that was metamorphosed at 1.0 Ga?

13.5 At the time of accretion (4.55 Ga), the Earth is believed to have had a ^{143}Nd/^{144}Nd ratio of 0.50682 and a ^{147}Sm/^{144}Nd ratio of 0.1926. Plot a graph of the value of the ^{143}Nd/^{144}Nd for the bulk Earth throughout geologic time and record the value of this ratio at an age of 2.5 Ga before present.

13.6 Using the information from Problem 13.5, calculate the evolution of the ^{143}Nd/^{144}Nd ratios in the crust and mantle if the crust is taken as having formed 2.5 Ga before present. The ^{147}Sm/^{144}Nd ratios in the crust and mantle at time of formation can be taken as 0.1435 and 0.2245, respectively.

13.7 Repeat Problem 13.5 but for ^{87}Sr/^{86}Sr. The initial ^{87}Sr/^{86}Sr for the Earth was 0.69898 and the ^{87}Rb/^{86}Sr was 0.0857.

13.8 Repeat Problem 13.6 but for ^{87}Sr/^{86}Sr. The ^{87}Rb/^{86}Sr ratios in the crust and mantle at time of formation were 0.48423 and 0.0332, respectively.

14 Magmatic processes

14.1 INTRODUCTION

Magmas that reach Earth's surface to form lavas are highly varied, ranging in composition from ultramafic komatiites, through basalts and andesites, to rhyolites and feldspathoidal felsic rocks. Although the compositions of these lavas may not represent all magmas formed in the Earth (some may be too dense to rise to the surface, and others may require high pressures to keep volatile fluxes in solution), they do indicate the enormous diversity of magmas. Explaining the origin of this diversity has been the dominant goal of petrology.

Early in the history of the science, most of the different magmas were thought to have independent origins; some were interpreted as the products of magma mixing, and still others the products of magma splitting (immiscible fractions). From early in the twentieth century, however, the trend has been to interpret the wide diversity of igneous rocks as being derived from only a few primary magmas. The process by which these magmas are modified is known as *magmatic differentiation*. For example, N. L. Bowen (1928), who championed this new interpretation, argued that basalt was the primary magma from which other magmas were derived. So persuasive were his arguments that during the first half of the twentieth century, the interpretation that compositional variations in magmas might reflect primary variations in the source region was almost completely neglected. Although petrologists now recognize the importance of the source region, magmatic differentiation is still considered the major cause for variations in the composition of suites of igneous rocks.

Lavas erupted from a single volcano at different times can reasonably be interpreted to share a common origin. The same is true of different igneous rocks within a single pluton or even in separate, but closely related bodies, especially when the rock types share some striking geochemical signature. Igneous rocks that are related to a common source are said to be *comagmatic* or *consanguineous*. The latter term implies a clear genetic lineage; that is, one rock type is derived in some way from another. For example, on the basis of field evidence, geochemistry, or experimental work, a particular magma might be interpreted as being derivative from another magma, which we can refer to as a *parental magma*. That parent may, in turn, have been derived from another magma, and so on. As the lineage is traced backward toward the primary magma, magmas are said to become more *primitive*. Conversely, magmas late in the lineage are said to be *evolved* or *differentiated*. The *primary magma* itself is the one derived directly from the source region without being affected by any process of differentiation. Although easy to define, primary magmas are, in fact, very difficult to identify, because of the uncertainty in the composition of the source region and hence in the composition of magma with which it would be in equilibrium.

Given that a group of igneous rocks, and the magmas from which they were derived, are comagmatic, we may ask how their differentiation occurred. Undoubtedly, this is the most fundamental question in petrology today, yet it is far from being completely answered because, for the most part, magmatic processes occur at depths where they are not accessible to direct observation, or they take place at rates that are far slower than can be achieved in laboratory experiments. Nonetheless, igneous rocks preserve many textural and chemical characteristics that shed light on the differentiation processes involved in their formation. Few of these characteristics, however, are without multiple interpretations. Herein, then, lies the reason for much of the interest and excitement in studying petrology.

Many different processes have been invoked to account for magmatic differentiation. Some of these involve changes in the composition of magmas while entirely in the liquid state. The components of a magma, when placed in a potential field, may have to modify their concentrations to achieve an equilibrium distribution. The effect of the Earth's gravitational field in this respect is very small, but steep temperature gradients can cause a significant redistribution of elements by what is known as the Soret effect. Steep temperature gradients, however, can exist only at the extreme outer margins of magma bodies, and then only for short time periods. The Soret effect is therefore likely to be only of local and transient importance. Some magmas are capable of splitting, into immiscible fractions, producing liquids of contrasting compositions. For the most part, however, such unmixing is restricted to magmas of evolved compositions and is not of importance to the fractionation of more primitive magmas. Thus, although differentiation does take place in the liquid state, it cannot be the main cause for the differentiation of igneous rocks.

Ever since the evidence was laid out so clearly by Bowen (1928) in his influential book *The Evolution of the Igneous Rocks*, most petrologists have accepted crystal–liquid fractionation as the major cause for igneous differentiation. As was evident from the discussion of phase diagrams in Chapter 10, the composition of a multicomponent liquid is changed dramatically by the crystallization of a mineral, especially when the composition of the mineral differs significantly from that of the liquid. What makes this process particularly attractive is that most lavas contain phenocrysts (magmas are not superheated), and the calculated composition of liquids that would be formed if the phenocrysts are removed commonly matches the composition of associated lavas. So convincing is this evidence, especially when documented for many different elements, that crystal fractionation is undoubtedly the dominant process of magmatic differentiation. How the crystals actually separate from the magma is, however, a hotly debated question.

As proposed by Bowen (1915b), crystals were thought to separate from their host magma by sinking or floating. This appears self-evident for minerals such as the ferromagnesian ones, which are considerably denser than the magmas in which they form. Other minerals, such as plagioclase in basaltic magma, however, may show no density contrast with the liquid. Furthermore, we now know that many crystallizing magmas behave as Bingham liquids, and thus even the dense ferromagnesian minerals may not be able to sink if they are unable to overcome the yield strength of the magma. For this reason, additional processes have been invoked, most of which involve the movement of magma. For example, *flowage differentiation* calls upon shear stresses in magma to help move crystals; *convecting* magma, on the other hand, can transport crystals in suspension to distant depositional sites; and *filter pressing* can separate liquid from crystals in much the same way that coffee is separated from coffee grinds. Each of these processes is likely to leave telltale evidence of its action.

Finally, the separation of a vapor phase from a magma can bring about marked differentiation, as evidenced by the high concentrations of certain rare elements in pegmatites (Section 11.6). Separation of a vapor phase, however, tends to occur in rather evolved magmas or during the final stages of crystallization of more primitive ones. It therefore cannot be responsible for the major trends of magmatic differentiation.

14.2 COMPOSITIONAL VARIATION IN SUITES OF VOLCANIC ROCKS

Lavas are important in the study of petrology because they provide samples of unequivocal magmatic liquids (or liquid plus phenocrysts). Many plutonic igneous rocks, although formed from magmas, are not, themselves, representative of the composition of liquids. They may, for example, be formed by the accumulation of early crystallizing minerals, as happens in the case of dunite, which is formed by the accumulation of olivine. Lavas from a single volcano may show progressive changes in composition both during a single eruptive episode or during successive eruptions. This compositional variation, then, is the most direct evidence we have of magmatic differentiation.

Almost any well-studied volcano can be used to illustrate the type of compositional variation found in volcanic rock suites. We will use, as an example, the Kilauea volcano, located on the southeastern side of the Island of Hawaii. This basaltic shield volcano, which is very active (erupting ~0.1 km^3 of magma per year), has been monitored continuously by the U.S. Geological Survey for many years. Not only are the compositions of its recent and historic eruptions well known, but geophysical monitoring of the volcano has provided much information about the magmatic plumbing and the movement of magma beneath the volcano.

As described in Section 4.3, seismic evidence indicates that the ultimate source of Kilauean lavas is at a depth of at least 60 km; from there, magma rises into a shallow chamber between depths of 6.5 to 5.7 km directly beneath the summit caldera (Fig. 4.13). As this chamber fills, the summit of the volcano expands and does so until either a summit eruption occurs or the magma moves laterally into the east or southwest rift zones. Movement of magma into the rift zones is not always accompanied by an eruption, but when it is, the first magma to be extruded is not of the newest addition to the rift zone but is derived from pockets of magma emplaced during earlier periods of activity. The most recent eruption of Kilauea at Pu‘u ‘O‘o, in the east rift zone (which began in January 1983 and continues at the date of publication of this book), appears to have bypassed the summit reservoir and flowed directly into the east rift zone. For a detailed analysis of the compositional and temporal variation in this most recent eruption, see Garcia *et al*. (2000).

Table 14.1 lists analyses of the full range of lavas erupted from Kilauea (mainly from Wright and Fiske, 1971). These are all basaltic rocks ranging in composition from picrite (no. 1) to quartz tholeiite (no. 15). Rocks containing more than 7.0 wt% MgO contain phenocrysts of olivine (Fo_{87-85}), a typical analysis of which is given in no. 16. Rocks containing less than 7.0% MgO contain phenocrysts of plagioclase and augite. Below 5.5% MgO, orthopyroxene phenocrysts are present, which take the place of olivine phenocrysts. At still lower MgO (<4.0%), ilmenite becomes a phenocrystic phase.

From a cursory inspection of Table 14.1, which is arranged in order of decreasing MgO (17.4 – 3.9%), certain relations are readily apparent. For example, with decreasing MgO, the TiO_2 and alkalis increase. The table, however, contains far too much data for any but the most obvious relations to be found by inspection. To simplify the task, petrologists utilize plots known as *variation diagrams*, which graphically portray compositional trends between certain groups of elements or oxides in igneous rock suites.

One of the most widely used diagrams is the triangular *FMA* plot, which shows the simultaneous variation in the weight percentages of FeO (may also include ferric iron recalculated as FeO, that is, $0.9 \times Fe_2O_3$), MgO, and total alkalis ($Na_2O + K_2O$). The positions of points in this diagram are determined by summing the FeO, MgO, and $Na_2O + K_2O$, and recalculating to 100%. For example, $F = 100 \times FeO/(FeO + MgO + Na_2O + K_2O)$.

The FMA diagram for the Kilauean rocks (Fig. 14.1) reveals a remarkable, systematic variation, in which the iron

Table 14.1 *Analyses of lavas from Kilauea, Hawaii*

	Analysis															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16[c]
SiO_2	48.05	48.43	47.92	48.21	49.16	49.20	49.71	50.10	50.37	50.56	50.74	50.85	50.92	51.24	53.42	40.01
TiO_2	2.04	2.00	2.16	2.24	2.29	2.57	2.68	2.71	3.09	3.16	3.35	3.36	3.61	3.74	3.36	0.04
Al_2O_3	10.33	10.70	10.75	11.37	13.33	12.77	13.65	13.78	14.02	13.92	13.57	14.02	13.80	13.60	13.75	1.13
Fe_2O_3	1.34	1.15	1.08	1.50	1.31	1.50	1.19	1.89	1.88	1.78	1.36	1.90	1.85	1.87	1.96	0.30
FeO	10.19	10.08	10.65	10.18	9.71	10.05	9.72	9.46	10.07	10.18	10.63	10.44	10.71	11.19	10.45	12.33
MnO	0.17	0.17	0.18	0.18	0.16	0.17	0.17	0.17	0.17	0.18	0.18	0.18	0.19	0.18	0.18	0.17
MgO	17.39	16.29	15.43	13.94	10.41	10.00	8.24	7.34	6.75	6.33	6.16	5.68	5.46	5.12	3.92	44.77
CaO	8.14	8.67	9.33	9.74	10.93	10.75	11.59	11.46	10.39	10.24	9.94	9.71	9.45	9.03	7.75	1.23
Na_2O	1.66	1.71	1.79	1.89	2.15	2.12	2.26	2.25	2.35	2.61	2.69	2.77	2.80	2.81	3.34	0.01
K_2O	0.36	0.35	0.44	0.44	0.51	0.51	0.54	0.57	0.62	0.64	0.67	0.74	0.75	0.83	1.10	0.00
P_2O_5	0.19	0.18	0.23	0.22	0.16	0.25	0.25	0.27	0.32	0.33	0.37	0.38	0.40	0.41	0.59	0.00
Total	99.86	99.73	99.96	99.91	100.12	99.89	100.00	100.00	100.03	99.93	99.66	100.03	99.94	100.02	99.82	99.99
M'[a]	75.3	74.2	72.1	70.9	65.6	63.9	60.2	58.0	54.4	52.6	50.8	49.2	47.6	44.9	40.1	86.6
F	34.0	35.0	38.0	38.0	43.0	44.0	47.0	48.0	51.0	52.0	53.0	53.0	54.0	56.0	56.0	22.0
M	59.0	57.0	55.0	53.0	46.0	44.0	40.0	37.0	34.0	32.0	31.0	29.0	28.0	26.0	21.0	78.0
A	7.0	7.0	8.0	9.0	12.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	18.0	18.0	24.0	0.0
D.I.[b]	16.18	16.54	17.75	18.59	21.20	20.95	22.31	23.52	26.01	27.93	28.97	30.64	31.54	32.92	41.44	
CIPW norm																
Q	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.11	2.46	2.10	2.25	2.83	3.42	4.24	6.68	
Or	2.13	2.07	2.60	2.60	3.01	3.01	3.19	3.37	3.66	3.74	3.96	4.37	4.43	4.90	6.50	
Ab	14.05	14.47	15.15	15.99	18.19	17.94	19.12	19.04	19.89	22.09	22.76	23.44	23.69	23.78	28.26	
An	19.67	20.49	20.00	21.24	25.22	23.82	25.51	25.82	25.88	24.38	22.97	23.64	22.87	22.04	19.28	
Cpx	15.60	17.14	19.95	20.65	22.66	22.61	24.74	23.87	19.30	19.92	19.69	18.17	17.66	16.63	12.77	
Opx	21.12	21.12	14.87	15.99	14.45	18.66	16.92	18.28	19.51	18.32	18.84	17.48	17.41	17.67	15.74	
Ol	21.04	18.57	21.20	16.51	9.98	6.21	3.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
Mt	1.94	1.67	1.57	2.17	1.90	2.17	1.73	2.74	2.73	2.58	1.97	2.75	2.68	2.71	2.84	
Il	3.87	3.80	4.10	4.25	4.35	4.88	5.09	5.15	5.87	6.00	6.36	6.38	6.86	7.10	6.38	
Ap	0.45	0.43	0.54	0.52	0.38	0.59	0.59	0.64	0.76	0.78	0.88	0.90	0.95	0.97	1.40	

[a] M′ is the magnesium number (100 × Mg/(Mg+Fe″)).
[b] D.I. is the Thornton and Tuttle (1960) differentiation index.
[c] Analysis 16 is of a typical olivine phenocryst in Kilauean basalts.
Data mainly from Wright and Fiske (1971).

and alkali contents increase as the magnesium content decreases. In addition, all but the least magnesian of the lavas (no. 15 in Table 14.1) lie on a straight line that radiates from the composition of the typical olivine phenocryst in these rocks. Based on our experience with ternary-phase diagrams and the lever rule (Section 10.14), this variation could be explained simply by subtraction of olivine from a magnesium-rich magma. The fact that the least magnesian lava falls off what is known as the *olivine control line* implies that, for this rock, some additional factor must have played a role in its differentiation. Indeed, this lava is the only one in this group to contain ilmenite phenocrysts. Ilmenite plots near the F corner in Figure 14.1 and if subtracted from a basaltic magma, would displace derivative magmas to the iron-poor side of the olivine control line.

Although the evidence from Figure 14.1 for olivine fractionation in the Kilauean rocks may be persuasive, it is rash to draw such a conclusion based only on the parameters F, M, and A. If correct, other components in the rocks should vary in a predictable manner. For this reason it is useful to compare all the components in the rock series. A variation diagram in which this is done (Fig. 14.2) plots the weight percentage of each oxide in the rock against the atomic ratio of 100 × Mg/(Mg + Fe″), which is determined by dividing the oxide weights in the analyses by their molecular weights. This ratio, which is known as the *magnesium number* and designated by M', is particularly useful as a *differentiation index* because, as we know from numerous phase diagrams in Chapter 10, ferromagnesian minerals that crystallize at high temperatures are more magnesian than those that crystallize at lower temperatures. Consequently, we can expect rocks with high magnesium numbers to have formed at higher temperatures and be more primitive than those with low magnesium numbers.

A number of important oxides in the Kilauean rocks are plotted against M' in Figure 14.2 along with the composition of typical phenocrystic olivine. For compositions with high values of M', the effect of olivine fractionation is readily apparent, with the concentration of all oxides (except MgO) increasing with decreasing M'. For M' values of less than ~67, however, the rate of increase in CaO with falling M' decreases. This change occurs where augite first appears in

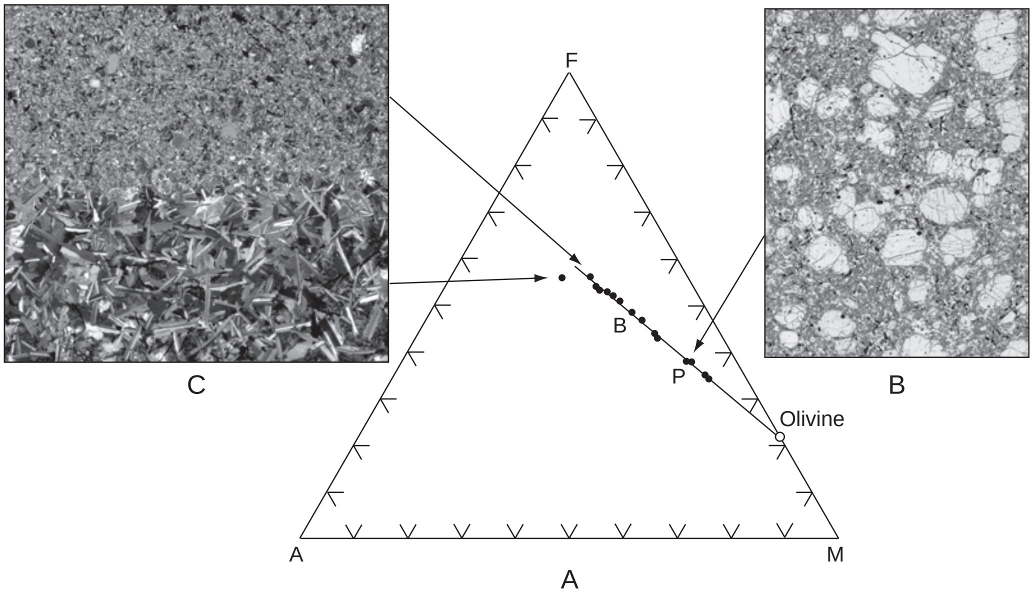


Fig. 14.1 (**A**) *FMA* variation diagram for Kilauean lavas (analyses given in Table 14.1). B and P mark compositions that may correspond to parental basalt and picrite magmas, respectively. (**B**) Photomicrograph of a typical sample from the lower part of the 1959 Kilauea Iki lava lake (see Fig. 2.2) in which an accumulation of olivine phenocrysts in the fine-grained basaltic matrix has created a picritic composition. (**C**) Photomicrograph of a typical sample from the upper part of the Kilauea Iki lava lake from which olivine crystals have been removed by sinking into the lower part of the lava lake. This has produced an iron-rich basaltic composition, which plots toward the end of the olivine control line. Prior to complete solidification, this basaltic mush was ruptured and residual liquid filled a fracture to form a coarse-grained segregation sheet. The segregation sheet falls off the olivine control line, because by the time it separated from the host basalt, ilmenite (opaque plates) had already started to crystallize.

the rock series as phenocrysts. Separation of augite phenocrysts from the magmas would diminish the increase in CaO caused by the separation of olivine alone. At slightly lower *M′* values, plagioclase also becomes a phenocrystic phase, and its separation can account for the Al_2O_3 content remaining constant and the CaO content falling in the less magnesian rocks. Oxides such as P_2O_5, K_2O, Na_2O, and TiO_2 do not enter the early-crystallizing minerals in the Kilauean basalts except plagioclase, which does incorporate some Na_2O, but even here, sodium is concentrated in the residual melt. Because these oxides do not enter early-crystallizing minerals, they are described as *incompatible*. With decreasing *M′*, the concentrations of incompatible elements steadily increase in the derivative liquids. The only exception is in the basalt with the lowest *M′*, where ilmenite becomes a phenocrystic phase, and thus TiO_2 is no longer an incompatible component at this stage of differentiation (a crystalline phase is present into which TiO_2 can enter) and thus the TiO_2 content begins to fall.

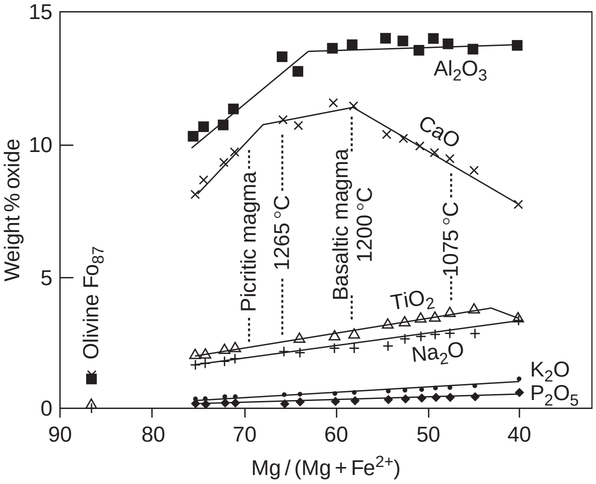


Fig. 14.2 Variation diagram of weight percentages of oxides versus the magnesium number of Kilauean lavas (Table 14.1). Temperatures are those measured during eruptions or are laboratory-measured liquidus temperatures.

Based on direct measurements of eruption temperatures and the laboratory determination of liquidus temperatures, we can label the *M′* variation diagram with typical magmatic temperatures. As would be expected, temperatures decrease with falling *M′*. Maximum eruption temperatures, which occur at the summit of Kilauea, are near 1208 °C; flank eruptions are typically cooler than this. Picritic rocks, which contain high concentrations of olivine phenocrysts, may represent primitive liquids, or they may simply be basaltic liquids in which olivine phenocrysts have accumulated. Evidence indicates that primary magmas beneath Kilauea may be picritic in composition, in which case their liquidus temperature would have to be higher than the 1200 °C eruption temperatures at the summit of Kilauea (Problem 14.2).

In Figure 14.2, the decrease in the rate of increase in CaO for rocks with *M′* less than ~ 67 was attributed to the crystallization of augite, followed by the crystallization of plagioclase in lavas with *M′* less than ~ 63. Although augite can be found as a phenocrystic phase in Kilauean lavas, it is not common. Instead, most augite occurs in the groundmass and crystallizes after plagioclase appears on the liquidus. The apparent sequence of appearance of minerals on the liquidus in Figure 14.2, therefore, seems at odds with the sequence of crystallization in the lavas (see the results from MELTS in Problem 14.2). However, at the higher pressure in the magma chamber beneath the summit caldera (Fig. 4.13(D)), augite crystallizes before plagioclase, thus explaining the sequence of changes in Figure 14.2 (Problem 14.3).

The hypothesized compositions of a primary picritic magma and of a basaltic parent magma are shown in Figures 14.2 and 14.1 (*P* and *B*). Compositions with M' greater than 70 are formed by magmas becoming enriched in olivine and are not representative of liquid compositions. Compositions with M' between 70 and 58 may also reflect accumulations of olivine in basaltic magma. Lavas with M' greater than 58 erupt primarily from the summit of Kilauea and have compositions that are determined largely by variations in the amount of olivine added or subtracted; that is, they lie on the olivine control line. Lavas with M' less than 58 erupt primarily along the rift zones, where magma has had time to cool to temperatures where plagioclase and pyroxene have crystallized and played a role in determining the composition of lavas.

The M' index is particularly useful for studying compositional variations in basalts, because the ferromagnesian components constitute such a large fraction of these rocks. In granitic rocks, however, where there may be only a few percent ferromagnesian minerals, this index is likely to be less sensitive to differentiation, and thus a different index is required. Two of the most commonly used ones are the *Differentiation Index* (D.I.) of Thornton and Tuttle (1960), which is simply the sum of the normative minerals that are concentrated in residual magmas, that is, quartz, orthoclase, albite, nepheline, leucite, and kaliophilite (kalsilite), and the Harker diagram, which simply uses the weight percent of SiO_2. The resulting diagrams for the Kilauean rocks (Figs. 14.3 and 14.4) are very similar, and the explanation for the compositional trends they exhibit is the same as for the M' diagram.

Variation diagrams can be used in a quantitative way to determine possible relations between rock types. To do this, however, care must be used to plot conservative quantities, such as the weight percentages of the oxides. Consider, for example, Figure 14.5, in which analyses 2 and 8 of Kilauean lavas and analysis 16 of the phenocrystic olivine from Table 14.1 are plotted against their weight percent of MgO. With the exception of a small deviation in the amount of CaO, equivalent oxides in these three analyses lie on straight lines. This indicates that the intermediate analysis, no. 2, could be formed by combining the two extreme analyses, or, conversely, analysis 8 could be formed from no. 2 by subtracting no. 16. Furthermore, using the lever rule, the relative lengths of the lines between the three analyses indicate the proportion of phases involved. For example, analysis 2 could be formed by combining 24 wt% of olivine (line 2–8/line 8–16) with 76% of lava 8. The small deviation in the CaO plot is probably attributable to too high a value of CaO in the olivine

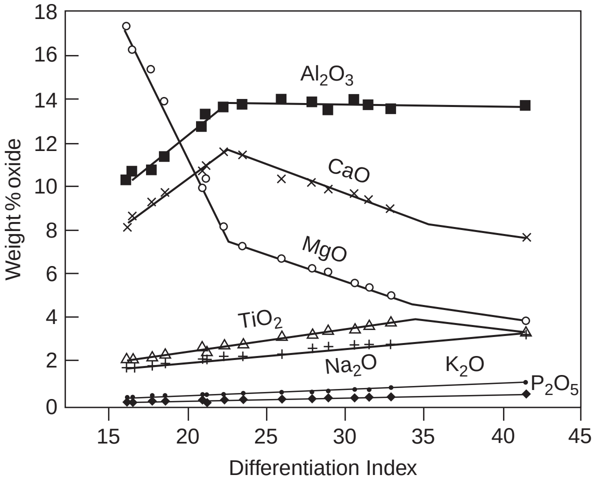


Fig. 14.3 Variation diagram of Kilauean lavas (Table 14.1) using the Thornton and Tuttle (1960) Differentiation Index.

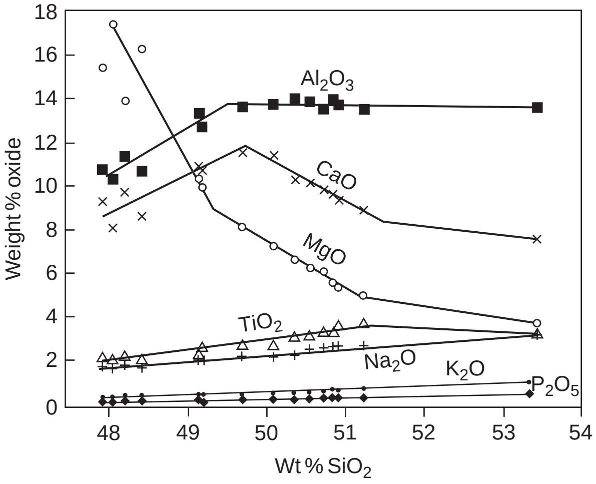


Fig. 14.4 Variation diagram of Kilauean lavas (Table 14.1) using the weight percent SiO_2 (Harker diagram).

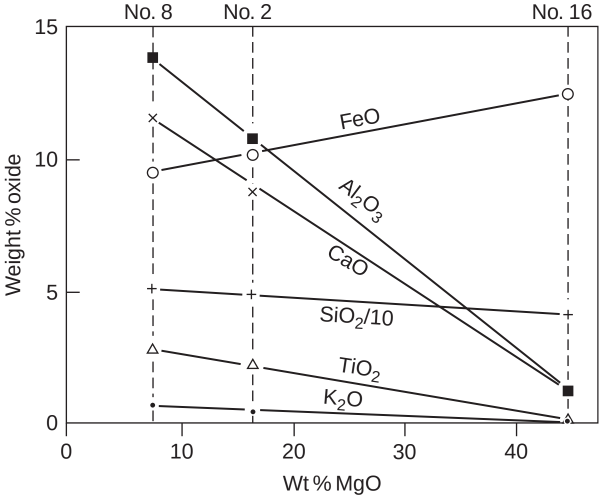


Fig. 14.5 Variation diagram showing the linear relation between Kilauean basalts 2 and 8 from Table 14.1 and the phenocrystic olivine, no. 16. See text for discussion.

analysis. Na_2O and P_2O_5, which have been omitted from Figure 14.5 for clarity, also lie on linear trends in this plot.

The lines in Figure 14.5 were visually fitted to the data. In this case the task was simple. But consider how much more difficult it would be if we were trying to account for the derivation of one lava from another by subtracting some unknown ratio of analyzed olivine and pyroxene crystals or possibly a combination of olivine, pyroxene, and plagioclase crystals. Even if the supposed mechanism were correct, unavoidable errors in the analyses would mean that no exact solution would exist. Only a best fit could be obtained. Calculations of this type are tedious and are therefore done on computers using least squares techniques to fit the data (Wright and Doherty, 1970; Gray, 1973).

Calculations such as these do not prove that mixing or separation occurred; they simply show that it is mathematically possible. Where olivine addition or separation is suspected, however, we have a means of testing the hypothesis. Roeder and Emslie (1970) showed that in a wide range of magmas and simple silicate systems, the composition of olivine can be related to the mole fractions of Mg and Fe′ in the melt with which the olivine is in equilibrium by a simple distribution coefficient, K_D, which remains relatively constant over a wide range of composition and temperature. Toplis (2005) has shown that this distribution coefficient does change significantly in iron-rich compositions, and Putirka (2005) has investigated its variation with temperature and pressure. In typical basaltic compositions, however, the distribution coefficient has values between 0.30 and 0.35. It is defined as

$$K_D = \frac{\left(X_{Mg}/X_{Fe''}\right)^{\text{liquid}}}{\left(X_{Mg}/X_{Fe''}\right)^{\text{olivine}}} \tag{14.1}$$

Thus, if the mole fractions of Mg and Fe″ in the magma are known, we can calculate the composition of the olivine that would be in equilibrium with it, or, conversely, given the composition of an olivine, we can calculate the Mg/Fe ratio in the liquid from which it formed.

To use the olivine distribution coefficient to test the two possible mechanisms suggested by Figure 14.5, we have first to decide with which liquid the olivine crystals would have been in equilibrium. For example, if analysis no.2 was formed by the addition of olivine crystals to analysis 8 the olivine crystals must have been in equilibrium with liquid 8. Using the Mg/Fe ratio of analysis 8 we calculate a value of K_D of 0.21, which is well outside the possible range of K_D values. We can conclude, therefore, that formation of lava 2 by the addition of olivine crystals with the composition of 16 to magma of composition 8 is not thermodynamically possible. Conversely, if we assume that lava 8 was formed from a picritic magma of composition 2 by the subtraction of olivine crystals, the olivine crystals would have to have been in equilibrium with a liquid of composition 2. This gives a K_D value of 0.45, which is equally impossible. Presumably the true mechanism, then, lies somewhere between these two extreme hypotheses; that is, the original magma was not as magnesian as no. 2, which must have had some olivine accumulate in it but not as much as 24%. If we take the value of K_D to be 0.33, the magma with which olivine of analysis 16 would be in equilibrium would have a magnesium number of 68. This is how the parental picrite magma in Figure 14.1 was determined (see Irvine, 1979).

Although rocks from only Kilauea have been discussed in this section, suites of rocks from other volcanoes reveal the same general features; that is, the compositional variation in lavas from a given volcano can be explained in terms of separation of phenocrystic phases from a parental magma. The conclusion that crystal fractionation must be responsible for the differentiation of these rocks seems inescapable. It is to the processes taking place within the magma chambers and conduits beneath volcanoes that we must now turn our attention in order to learn how this fractionation occurs.

14.3 CRYSTAL SETTLING IN MAGMA

Although many geologists had considered crystal settling as a possible mechanism of igneous differentiation, it was not until Bowen's 1915 paper on the "crystallization differentiation in silicate systems" that incontrovertible evidence of the process was documented. Bowen showed from experiments in the system forsterite–diopside–silica that crystals of olivine and diopside could sink (and crystals of tridymite float) a centimeter or two in periods of less than an hour. Although the experiments were carried out at high temperatures just below the liquidus where viscosities are much lower than they would be at typical magmatic temperatures, the magnitude of the effect was so great that Bowen concluded that crystal sinking (and to a lesser extent floating) must be of general importance to igneous differentiation. Although ferromagnesian minerals have densities that are significantly greater than those of magmas and may therefore sink, plagioclase has a density that is very close to that of basaltic magma and would therefore be neutrally buoyant (Problem 14.6).

It is not surprising that once crystal settling had been confirmed in the laboratory, field evidence for the process became more apparent. One of the most famous bodies in which crystal settling was believed to have occurred was the Palisades Sill, which forms the prominent escarpment on the west side of the Hudson River facing New York City. This sill, which is over 300 m thick, has a layer containing abundant olivine crystals approximately 15 m above its base. Olivine that was not trapped in the upper chilled margin was believed to have sunk and accumulated to form the olivine layer. Recent studies, however, have shown that the Palisades Sill has had a complicated history involving at least three injections of fresh batches of magma (Shirley, 1987), and although gravitative settling may have played a role in its differentiation, a simple model of olivine settling is no longer believed to be valid (see Section 14.4). Nonetheless, the Palisades Sill is a large enough body that it would have taken several hundred years to solidify, and if olivine sinking occurs at a perceptible rate, some accumulation of olivine would be expected.

The rate at which a crystal sinks or floats depends on the difference between the buoyant force caused by the density contrast between the crystal and the magma ($\Delta\rho = \rho_s - \rho_m$) and the viscous drag of the liquid on the crystal. If we consider, for simplicity, a spherical crystal of radius r, the buoyant force on the crystal is

$$\text{buoyant force} = 4/3\pi r^3 g\Delta\rho \tag{14.2}$$

where g is the acceleration of gravity. The viscous drag resulting from the laminar flow of the liquid (viscosity η) across the surface of the crystal at a constant velocity of dz/dt, where z is measured downward, is given by *Stokes' formula* (Turcotte and Schubert, 1982a):

$$\text{drag force} = 6\pi\eta r\frac{\mathrm{d}z}{\mathrm{d}t} \tag{14.3}$$

Equating these two forces, we obtain an expression for the *steady state* or *terminal* velocity (v_t) of the crystal:

$$v_t = \frac{\mathrm{d}z}{\mathrm{d}t} = \frac{2g\Delta\rho r^2}{9\eta} \tag{14.4}$$

which is commonly referred to as *Stokes' law*. This law is valid only when the crystal moves slowly enough that no turbulent vortices develop. This condition is met when the Reynolds number is less than 1. The Reynolds number is defined as

$$\text{Re} \equiv \frac{\rho_m 2 r v_t}{\eta} \tag{14.5}$$

For most geological cases this number is much smaller than 1. For nonspherical crystals, the factor $^2/_9$ in Eq. (14.4) can be changed to account for shape.

We can use Stokes' law to determine how far crystals of olivine might have sunk in the Palisades Sill. Following Shirley (1987), we assume that the viscosity of the magma was 400 Pa s and its density was 2620 kg m^{-3}. If the olivine crystals, which have a density of 3500 kg m^{-3}, had a radius of 1 mm (note that grain size is normally given as grain diameter), their settling velocity would have been 4.8×10^{-6} m s^{-1}, or 1.7 cm h^{-1}, or 41.4 cm day^{-1}, or 151 m a^{-1}. This is a significant settling velocity, and in the hundreds of years the intrusion took to solidify, olivine crystals should have been able to sink to the floor of the magma chamber. Note that the Reynolds number for these conditions is 6.2×10^{-8}, so the use of Stokes' law is justified.

To calculate the distance traveled by a sinking crystal, it is necessary to integrate Eq. (14.4), which if we assume ρ and η to be constant, can be written as

$$\int \mathrm{d}z = \frac{2g\Delta\rho}{9\eta}\int_0^t r^2 \mathrm{d}t \tag{14.6}$$

The radius of a crystal is unlikely to remain constant during sinking; instead, the crystal will either grow or dissolve. Crystals nucleate near the roof of the intrusion at some degree of undercooling, which allows them to sink into hotter magma without necessarily dissolving. Furthermore, the rise in pressure with increasing depth in a magma chamber causes the liquidus temperature to rise. The general tendency, then, is for crystals that form near the roof of the magma chamber to continue growing as they sink.

To evaluate the right-hand side of Eq. (14.6), we need to express the radius of a crystal as a function of time, which can be done by making use of the growth rate equations in Chapter 12 (Eqs. (12.7) to (12.10)). For example, if the growth of olivine crystals were controlled by a phase boundary reaction, integration of Eq. (12.8) would give

$$\int_{r_i}^{r_t} \mathrm{d}r = k_R \int_0^t \mathrm{d}t \quad \text{from which} \quad r_t = r_{\text{initial}} + k_R t$$

The initial and final radii of the crystal might be determined, for example, from the size of crystals trapped in the upper chilled margin and in the layer of accumulation. Determining the value of k_R is more difficult and requires that we know how long the crystal took to grow to its final size. A rough estimate might be obtained from the length of time required for the magma to solidify up to the layer of olivine accumulation (see Problem 14.8). Equation (14.6) could then be rewritten as

$$\int \mathrm{d}z = \frac{2g\Delta\rho}{9\eta}\int_0^t (r_{\text{initial}} + k_R t)^2 \mathrm{d}t \tag{14.7}$$

Even without precise knowledge of the effect of crystal growth rates on settling velocities, the simple application of Stokes' law indicates that ferromagnesian minerals should sink in basaltic magmas with significant velocities that are on the order of centimeters per day. The question then is: why do more sills not show evidence of crystal settling? Many sills of the same age, composition, and thickness as the Palisades Sill, but in other parts of the Mesozoic basins of eastern North America show little, if any, evidence of crystal settling. Clearly, other factors must be involved.

McBirney and Noyes (1979) have drawn attention to the fact that magmas, on cooling into the crystallization range, become non-Newtonian liquids with a significant yield stress that increases with falling temperature, especially when plagioclase and pyroxene begin crystallizing. They found, for example, that a Columbia River basalt had a yield strength of 60 Pa at 1195 °C. A dense crystal in such a magma would not be able to sink until it first grew large enough to overcome the yield strength of the magma. We can calculate this critical size for a spherical crystal of radius r by equating the gravitational force on the crystal (Eq. (14.2)) with the shear force imposed by the yield strength (σ_y) acting on the surface of the crystal

$$\frac{4}{3}\pi r^3 \Delta\rho g = \sigma_y 4\pi r^2$$

or for the case where the gravitational force overcomes the yield strength

$$\frac{r\Delta\rho g}{3} > \sigma_y \tag{14.8}$$

If the magma in the Palisades Sill had a yield strength of 60 Pa, the radius of an olivine crystal would had to have exceeded 2 cm (grain diameter 4 cm) for the crystal to sink. This grain size is an order of magnitude greater than occurs in the Palisade Sill. Clearly, no settling would have occurred in this intrusion if, or when, the yield strength of the magma reached significant values. Indeed, crystal settling in any magma chamber may be restricted to a short period while the magma is still at high temperatures and has Newtonian behavior.

So far, we have discussed crystal settling in terms of stationary magmas. If, however, magma is moving, either because of emplacement or convection, the flow may counteract, or add to, any movement due to gravitative settling of crystals. In extreme cases, convection could become turbulent, and then the effects of crystal settling would be completely neutralized by the movement of magma. As will be seen in the next section, convection in magma chambers is common, and it may therefore be another reason why simple crystal settling is not more prevalent in igneous bodies.

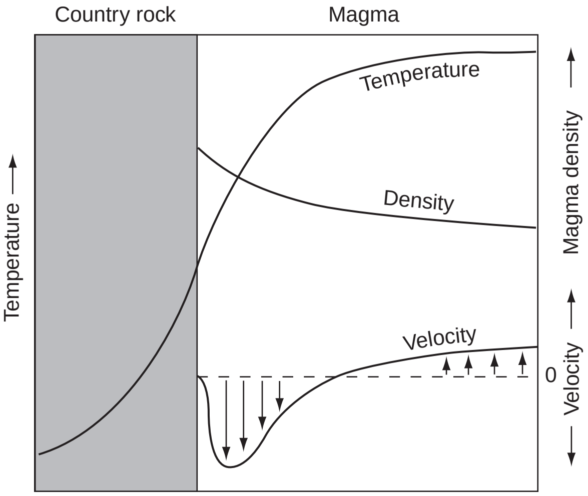


Fig. 14.6 Temperature and density distribution near the vertical contact of an igneous intrusion. The higher density of the cooler magma near the contact relative to that farther into the intrusion produces gravitational instability, with magma convecting down the wall and up in the center of the intrusion.

14.4 MAGMA CONVECTION

Magma in an isolated chamber may convect because of density differences resulting from thermal or compositional variations within the magma (Martin *et al*., 1987). This is known as *free convection*. If the magma chamber is still connected to a feeder conduit, influx of fresh magma may cause another type of convection known as *forced convection*. Forced convection is undoubtedly important in chambers that are repeatedly replenished with magma, such as those beneath the mid-ocean ridges, and the vorticity associated with this type of convection can be an effective means of mixing magmas (Campbell and Turner, 1986). The effects of forced convection, however, are transient compared with those of free convection, which may last through a considerable part of the cooling history of an intrusion. The remainder of this discussion deals only with free convection; forced convection is discussed in Section 14.10.

Consider magma near a vertical wall of an intrusive body. Following intrusion, a temperature gradient, such as that shown in Figure 14.6, soon develops due to the conduction of heat into the surrounding rocks. Because magma contracts on cooling (Eq. (2.3)), lower temperatures toward the contacts cause magma there to be denser than that a short distance into the intrusion. The system is therefore gravitationally unstable, and the dense magma near the wall sinks, while the hotter magma in the center rises, with the convection continuing as long as magma remains fluid enough to move.

The approximate rate at which magma sinks near a cool, vertical wall is given by

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \left(\frac{g\Delta\rho kL}{\eta}\right)^{1/2} \tag{14.9}$$

where g is the acceleration of gravity, $\Delta\rho$ the density contrast that exists over a distance L resulting from the difference in temperature over that distance, k the thermal diffusivity, and η the viscosity. For example, if a temperature difference of 400 °C exists across a distance of 2.0 m near the vertical contact of a dike of basaltic magma for which the coefficient of expansion is 6×10^{-5} °C^{-1}, the density contrast would be 64.8 kg m^{-3} (Eq. (2.3)). If the thermal diffusivity, viscosity, and density of the magma are 10^{-6} m^2 s^{-1}, 100 Pa s, and 2700 kg m^{-3}, respectively, the convective velocity would be 0.4 cm s^{-1}. This velocity is many orders of magnitude greater than the velocities at which ferromagnesian minerals would sink based on Stokes' law.

The velocity calculated with Eq. 14.9, in reality, would be too great, because the equation does not take into account the increase in viscosity that would accompany the decrease in temperature toward the contact. In addition, lower temperatures would be accompanied by higher degrees of crystallization, which increases the bulk viscosity (Eq. (14.24)) and might impart a yield strength to the magma; this, then, would have to be overcome before convection could occur. These factors have been taken into account by Spera *et al*. (1982) in calculating the velocity profile in the thermal boundary layer of a silicic magma chamber whose temperature is maintained constant by the intrusion of mafic magma into its base. As would be expected, the low temperatures near the contact produce a highly viscous crystal mush that is unable to sink, despite its negative buoyancy. Toward the inner side of the thermal boundary layer, however, magma becomes less viscous and yet has negative buoyancy. The downward convecting velocity of the magma therefore increases rapidly away from the contact, reaching a maximum of about 12 km a^{-1} (137 cm h^{-1}) at 10 m from the contact on the inside of the thermal boundary layer. This velocity is still orders of magnitude greater than the rate at which crystals will sink in a silicic magma.

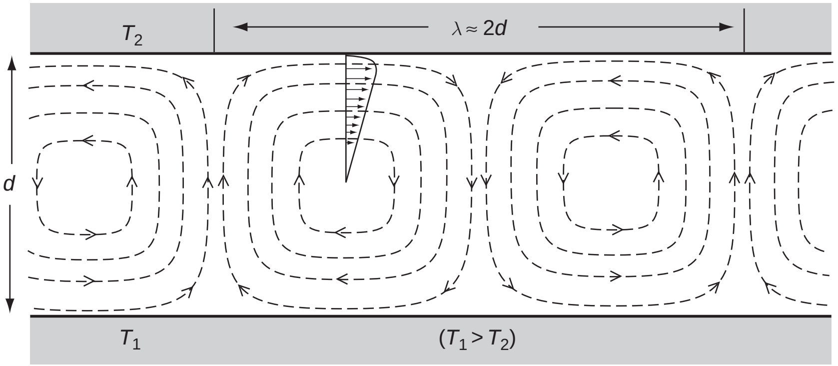


Fig. 14.7 Bénard convection cells produced in a sheet of liquid that is gravitationally unstable because its base is kept at a higher temperature than its roof. In the third dimension, cells form rolls at low Rayleigh numbers, polygonal cells at intermediate values, and chaotic turbulent cells at high Rayleigh numbers. See text for discussion.

Magma cooling along a vertical wall is always gravitationally unstable. If, however, the contact is horizontal, as in a sill, the situation is more complex, and convection may, or may not, occur depending on a number of factors (Turner, 1973). Consider a horizontal sheet of liquid between upper and lower boundaries that are maintained at constant temperatures T_1 and T_2 (Fig. 14.7). If $T_1 > T_2$, magma near the base of the sheet has a lower density than magma near the roof and so experiences a buoyant force. The magnitude of this force depends on the coefficient of thermal expansion of the magma, α, the temperature difference ($T_1 - T_2$), and the thickness of the sheet (d). This buoyant force causes magma near the base to rise and displace the denser magma near the roof. Resisting this movement, however, is the viscous drag of the magma. Also, the diffusion of heat through the magma diminishes the temperature difference that causes the density inversion. Thus the higher the thermal diffusivity, the less likely is a temperature (and hence density) difference to be maintained.

Just as the Reynolds number was defined as the ratio of the inertial and viscous forces, another dimensionless number, the *Rayleigh number* (Ra), can be defined as the ratio of the factors driving thermal convection to those opposing it.

$$\mathrm{Ra} \equiv \frac{\rho g d^3 \alpha (T_1 - T_2)}{\eta k} \tag{14.10}$$

When the Rayleigh number exceeds the critical value ($\mathrm{Ra_c}$) of 1708 for the situation illustrated in Figure 14.7, the sheet of liquid becomes unstable and breaks up into a series of convective cells, known as *Bénard cells*, which have wavelengths of approximately twice the thickness of the sheet. If magma in the 330-m-thick Palisades Sill, for example, had a density of 2620 kg $\mathrm{m^{-3}}$, a coefficient of expansion of $2.5 \times 10^{-5}\ \mathrm{K^{-1}}$, a viscosity of 400 Pa s, and a thermal diffusivity of $4.4 \times 10^{-7}\ \mathrm{m^2\ s^{-1}}$ (Shirley, 1987), the Rayleigh number would be $10^{11} \times (T_1 - T_2)$. Consequently, with only a fraction of a degree difference in temperature between the bottom and top of the Palisades Sill, the Rayleigh number would exceed the critical value and convection would occur. This is likely to be the case in all but the thinnest sheets.

When the Rayleigh number just exceeds the critical value, the Bénard cells form long rolls in plan view. When the Rayleigh number exceeds 10^4 to 10^5, these rolls break up into polygonal convection cells, and at values of Ra in excess of 10^6, the convection becomes turbulent with chaotic unsteady motion (Sparks *et al*., 1984). Thus, in the Palisades Sill the convection would have most definitely been turbulent, at least during the early stages of convection. As the magma cooled and become more viscous and the sheet of remaining liquid became thinner, the Rayleigh number would have decreased and more regular, steady patterns of convection may have developed.

Once convection cells become established and achieve a steady velocity, the dissipation of potential energy caused by the density distribution in the magma exactly balances the viscous dissipation of mechanical energy. The potential energy ultimately results from the temperature distribution, and the convection cell provides a far more efficient and rapid means of transferring heat down the temperature gradient than does conduction. If no convection occurred in the sheet of liquid between the two surfaces illustrated in Figure 14.7, heat could be transferred only by conduction (assuming that radiation is unimportant), which according to Eq. (5.3) would be

$$J_z = -K \frac{\mathrm{d}T}{\mathrm{d}z} = \frac{K(T_1 - T_2)}{d} \tag{14.11}$$

where J_z is the heat flux and K is the thermal conductivity of the magma. When the magma convects, heat is physically transported in the form of hot magma from the bottom to the top of the chamber. Thermal diffusion has then to transport the heat only a short distance through the roof. The heat flux through the roof of a convecting chamber is therefore greater than through the roof of a chamber with stationary magma; that is, $J_z > K(T_1 - T_2)/d$. The ratio of the actual heat flux (J_z) to that which would have been released had the magma been stationary gives a measure of the vigor of the convection and is known as the *Nusselt* number (Nu); that is,

$$\mathrm{Nu} \equiv \frac{J_z d}{K(T_1 - T_2)} \tag{14.12}$$

When the Rayleigh number is greater than 10^5, which is usually the case for most magma bodies, the Nusselt number is given by $0.10\mathrm{Ra}^{1/3}$ (Turner, 1973).

Because convection transports magma from the hot lower parts to the cooler upper parts of a chamber, the temperature gradient in a convecting body of magma is far less than the gradient in a nonconvecting one. The thermal gradient in a convecting cell approaches the adiabatic gradient (see Problem 7.7), which can be expressed as

$$\frac{\mathrm{d}T}{\mathrm{d}P}=\frac{T\bar{V}\alpha}{C_{\mathrm{p}}} \tag{14.13}$$

where $\bar{V}$ is the molar volume ($\bar{V}=1/\rho$), and C_{p} is the heat capacity of the magma. But we know from Eq. (1.1) that $\mathrm{d}P=\rho g\mathrm{d}z$, so Eq. (14.13) can be rewritten as

$$\frac{\mathrm{d}T}{\mathrm{d}z}=\frac{gT\alpha}{C_{\mathrm{p}}} \tag{14.14}$$

Again, using data for the magma of the Palisades Sill as an example, and assuming a temperature of 1373 K, and a heat capacity of 1.2 kJ $\mathrm{kg}^{-1}\,°\mathrm{C}^{-1}$, the temperature gradient in the convecting magma would approach 0.3 K km^{-1}. This is a remarkably low gradient compared with the typical geothermal gradient of 30 K km^{-1}.

The melting points of most minerals rise with increasing pressure. In a basaltic magma, the melting points of the early-crystallizing minerals increase, on average, about 3 °C for every kilometer of depth (slightly less in granitic magma). This rise in melting points is approximately one order of magnitude greater than the rise in temperature with depth in a convecting basaltic magma (Fig. 14.8). Consequently, crystallization takes place near the base of a convecting cell where the pressure is greatest. This is illustrated schematically in Figure 14.8 for a superheated convecting magma that cools and eventually intersects the liquidus. Because the convecting magma is essentially the same temperature throughout, the magma first intersects the liquidus at the lowest point in the convecting cell. Some crystallization would occur in a narrow boundary layer near the roof where there is a steep temperature gradient. This material, being cool and laden with crystals, would be dense and thus gravitationally unstable. Periodically, plumes of this material would sink from the roof to the floor of the chamber. The result is that most accumulation of crystals in a convecting magma chamber takes place on the floor, either by direct crystallization or from high-density crystal-laden slurries that sink from the upper, cooler parts of the chamber.

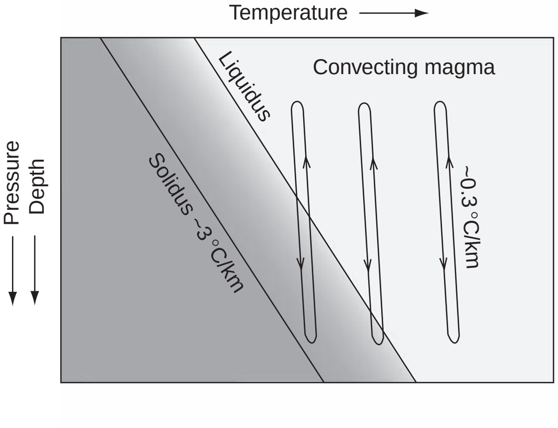


Fig. 14.8 On cooling, a convecting magma, which has a shallow adiabatic temperature gradient, first intersects the steeper liquidus at its deepest part; convecting magmas consequently solidify from their base upward.

The rate of cooling of a convecting magma body is far greater than that of a nonconvecting one. Because of the low-temperature gradient within a convecting cell, temperatures near the roof of a chamber are almost as high as those deeper in the body (Eq. (14.14)). The resulting steep temperature gradient across the upper contact causes the heat flux from the magma body to be high. The rate of cooling of the body as a whole depends almost entirely on this heat flux through the roof. The length of time for a convecting magma chamber to crystallize completely is, according to Brandeis and Jaupart (1987),

$$\text{convecting crystallization time}=\frac{15d^2}{\mathrm{Ra}^{1/3}k} \tag{14.15}$$

whereas the time for a nonconvecting one is

$$\text{nonconvecting crystallization time}=\frac{d^2}{4k} \tag{14.16}$$

Thus for a convecting body, the length of time to crystallize is proportional to the thickness of the body (substitute Eq. (14.10) into Eq. (14.15)), whereas for a nonconvecting body it is proportional to the square of the thickness.

The Rayleigh number for many magma bodies is well in excess of the critical value, and thus convection in magma chambers is common. The importance of convection depends on its rate relative to the rates of other processes. We have already seen that the rate of convection of even viscous granitic magma near a vertical wall is orders of magnitude greater than rates of crystal settling. The rate at which magma convects in a sheet-like body (Turcotte and Schubert, 1982a) is given by

$$\text{convective velocity}=\frac{0.271k(\mathrm{Ra})^{2/3}}{d} \tag{14.17}$$

If the Rayleigh number for the Palisades magma was 10^{11}, the convective velocity could have been as high as 3 km day^{-1}. Thus a packet of magma in a convective cell that occupied the entire thickness of the sill might have made two complete revolutions around the cell per day. Clearly, these rates are orders of magnitude greater than those calculated for crystal settling, and may actually be rapid enough to keep magma relatively homogenized during convection. Crystal concentrations can, however, develop in convecting cells. If crystals are denser than magma, they tend to concentrate in regions of convective upwelling, whereas if they are less dense, they concentrate in regions of convective downwelling (Marsh and Maxey, 1985).

As stated at the beginning of this section, the density differences that drive convection are caused not only by temperature variations but by compositional differences

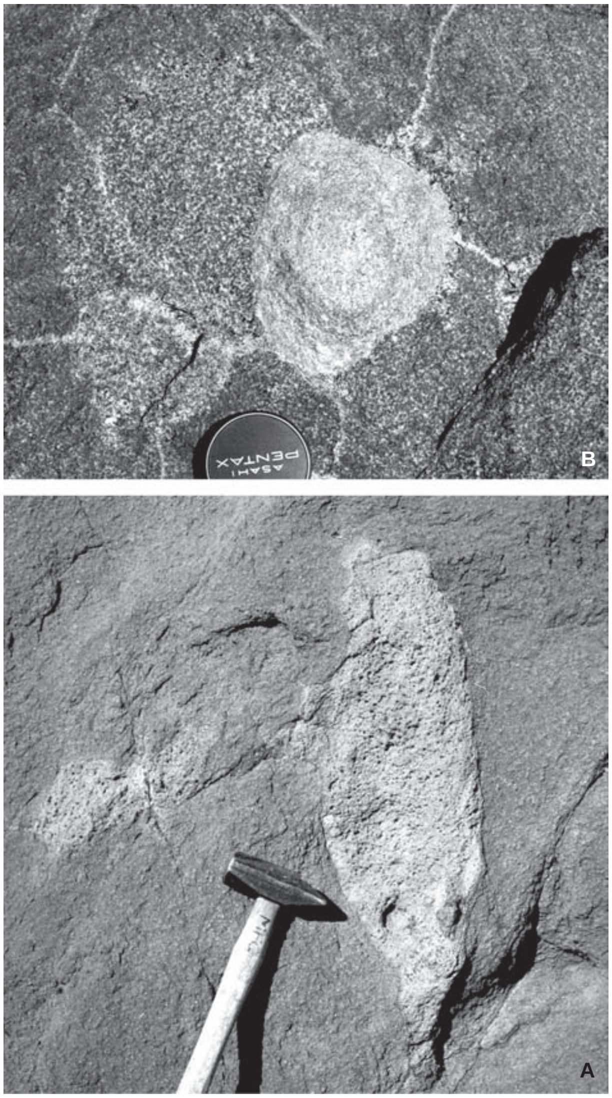


Fig. 14.9 (**A**) Near-vertical oblique section through a pipe-like body of nepheline monzonite–syenite in alkali gabbro in the Rattlesnake Mountain Sill, Big Bend National Park, Texas. These pipes appear to have risen to feed horizontal sheets of felsic rock in the upper part of the sill (Carman *et al.*, 1975). (**B**) Horizontal section through pipes showing an earlier more mafic pipe cut by a later more felsic one.

(Fig. 14.9). For example, basaltic magma that first crystallizes plagioclase near the wall of a chamber produces residual liquid that is more dense than the initial magma and would therefore sink. Conversely, if olivine or pyroxene were to crystallize first, the residual liquid would be less dense, so it would rise. In the right-hand inset photograph of Figure 14.10, crystallization of Na_2CO_3 on the sloping wall of a laboratory tank produced less dense aqueous solution that rose and set up convection, which can be seen to have distorted the vertical trails of sinking crystals of dissolving potassium permanganate crystals.

Convection near the wall of a magma chamber can be complex if partial melting or assimilation of wall rock produces density instabilities. For example, the downward thermal convection of basaltic magma may cause partial melting of country rocks to form granitic melt, which, being less dense than the basaltic one, would rise (Fig. 14.10(A)). Similarly, if a boundary layer of magma assimilates water from country rocks, the less dense hydrated magma would rise. Crystallization of Fe at the inner core–outer core boundary produces liquid residue that is less dense than the overlying liquid. The resulting convection may contribute to generating Earth's magnetic field. Any process, then, that changes the composition of a magma may induce convection. The volume of magma affected by this process at any one time is small because of low diffusion rates, but over the life span of an intrusion, its cumulative effect may be large. Magma in thin boundary layers along vertical walls of a chamber could, for example, float or sink and accumulate at the top or bottom of an intrusion to form large volumes of differentiated magma (Fig. 14.10(A) and (B)).

The density difference that drives thermal and compositional convection results from the diffusion of heat and chemical constituents respectively. When both factors play roles, the convection is described as *double-diffusive convection* to indicate that two diffusion coefficients are involved (Turner, 1973). Heat diffuses about 10 000 times more rapidly than do chemical species ($k = 10^{-6}$ m^2 s^{-1}, whereas $D = 10^{-10}$ m^2 s^{-1} as seen in Fig. 5.13). Consequently, thermal convection develops first in an intrusion, and thermal boundary layers are always much thicker than compositional ones. The effect of composition on magma density, however, is far greater than the thermal effect, because of the small coefficient of thermal expansion of magmas. Compositional changes in magma brought about by crystallization can therefore produce liquids of very different densities, which can float or sink rapidly (Problem 14.15).

One example where compositional convection may play an important role is in piles of crystal mush on the floor of magma chambers (Tait and Jaupart, 1992; Jaupart and Tait, 1995). In experiments using aqueous salt solutions, the accumulation and growth of crystals on the floor of a tank can produce residual liquids that are less dense than the overlying body of liquid. In such cases, the residual liquid rises through the porous and permeable crystal mush and forms low density fingers of liquid that rise into the denser overlying liquid, which at the same time must sink into the crystal mush. The expulsion of liquid from the crystal mush can be augmented by compaction (Section 14.5). The convective rise of liquid through the porous medium can become channelized and the flow so rapid that small conical mounds of salt crystals can form around the exit points of the channels. In a similar manner, low density residual melt in piles of crystal mush on the floor of magma chambers is expected to rise and be replaced by denser magma from the overlying main body of magma. When channelized, this flow may be responsible for the formation of vesicle cylinders in basaltic lava flows (Fig. 4.12(A)) and the pipe-like bodies of more evolved rock (Fig. 14.9) that are found in the lower parts of some sheet-like basaltic intrusions (Carman *et al.*, 1975; Bédard *et al.*, 2007). Of course, if the residual liquid were denser than the main body of liquid, the system would be gravitationally stable,

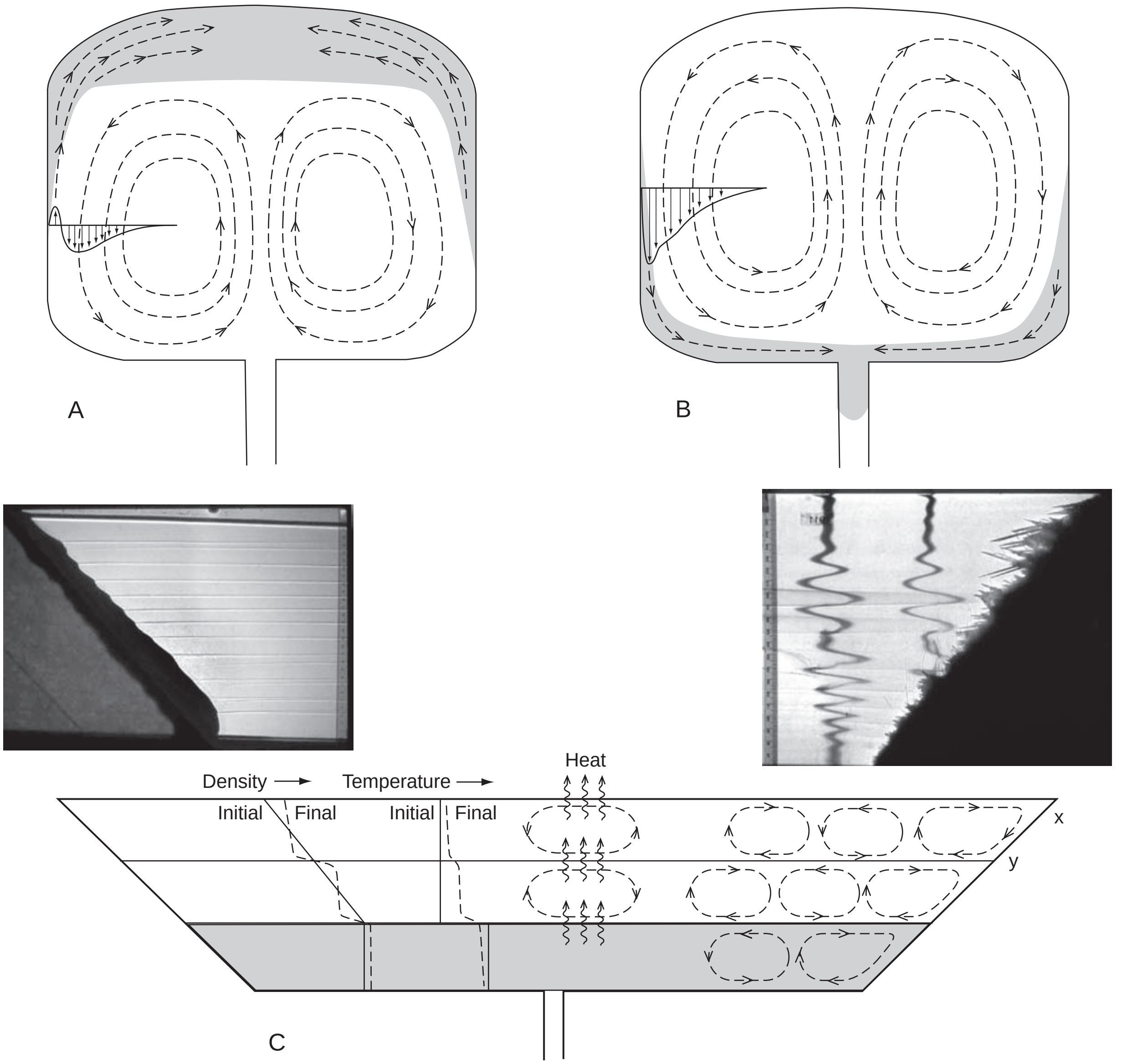


Fig. 14.10 Double-diffusive convection cells. (**A**) Magma that assimilates water from wall rocks produces a low-density boundary layer that rises and accumulates in the upper part of the chamber, even though a wider thermal boundary layer causes magma to the inside of the compositional boundary layer to convect downward. (**B**) Magma cooling and convecting down the walls and crystallizing plagioclase onto the walls produces a dense residual liquid that can descend rapidly and pond on the floor of the chamber. (**C**) A laterally extensive sheet of magma that initially has almost constant temperature throughout and steadily increasing density with depth has a second more dense, hot magma emplaced at its base. Heat from this second intrusion rises into the base of the overlying magma, which then buoyantly rises, but as it does so, it eventually reaches magma of the same density (because of initial density gradient), and then it is forced to move laterally. As it does so, heat is transferred by diffusion into the overlying, less dense magma. Little, if any diffusion of chemical constituents takes place across the boundary because of the very low mass diffusion rates. Magma in this upper cell now convects, but on rising, it too encounters lower-density magma, which prevents it from rising farther. In this way, many separate sheets of convecting cells can form. Near the outer contact of the chamber, magma cooling at x becomes more dense and sinks, but on reaching point y, its density matches that of the main body of magma at this depth (because of initial density gradient), and the downward-convecting magma moves laterally away from the wall. The inset photograph to the left shows many double diffusive convection cells developed in a tank of aqueous Na_2CO_3 that was cooled along the inclined plane to the left. Each cell is marked by a sharp change in refractive index (bright line). The inset photograph to the right shows a similar experiment in which two crystals of potassium permanganate were dropped into the tank, and as they sank, their dissolution trails became distorted by the convective flow of the liquid in the individual horizontal cells. Some Na_2CO_3 crystals can be seen to have grown in from the cooling surface to the right. The growth of these crystals produced a lower density residual liquid that rose along the sloping wall and out into the tank at the top of each cell. (Photographs by J. R. Wilson.) Although stratified double-diffusive convection cells of this type can be produced with salt solutions in carefully adjusted laboratory experiments, their existence in magma chambers is yet to be proven.

and the residual liquid would remain in place amongst the crystal. In the roof zone of an intrusion, the situation would be reversed, and dense residual liquids could drain out of a crystal mush, whereas less dense ones would remain in place. This type of compositional convection may play important roles in the solidification of rocks and result in significant bulk compositional changes.

Complex stratified sequences of Bénard cells (Fig. 14.10(C)) can form through double-diffusive convection (Turner, 1973; Huppert *et al*., 1986, 1987). Such layered convection cells have been found in seawater that has special temperature and salinity gradients. Although sequences of double diffusive convection cells have been invoked to explain some of the layering in large gabbroic intrusions (Irvine *et al*., 1983; Sparks *et al*., 1984; Wilson and Larsen, 1985), the special conditions necessary for this type of convection may not occur in typical magmas (McBirney, 1985).

We have discussed convection so far in terms of the movement of liquid whose density has been changed as a result of cooling (thermal convection) or compositional changes due to crystal growth (compositional convection). However, the largest density change that magma undergoes is due to crystallization. As magma crystallizes, its bulk density increases. Along the walls of magma chambers, dense piles of crystal mush can become gravitationally unstable and slump down and out across the floor of magma chambers (Wager and Deer, 1939). In the roof zone of sheet-like intrusions, crystal mush can separate from the roof and descend as dense plumes to the floor of the body (Brandeis and Jaupart, 1986; Marsh, 1988b, 1989; Jaupart and Tait, 1995). This type of convection, which is known as *dripping instabilities*, has been modeled by Bergantz and Ni (1999). They show that when plumes of crystal mush descend into low viscosity magmas, such as alkali basalt, they tend to break up and the crystals become dispersed throughout the magma. In more viscous magma such as tholeiitic basalt, the plumes maintain more coherence and a larger fraction settles to the floor of the body. In either case, this type of convection is an efficient means of cooling a body of magma from above and dispersing many crystals throughout the body.

Some petrologists believe that despite the dense nature of the crystal mush in the roof zone of sheet-like intrusions, its viscosity and strength prevent it from escaping from the roof (Hort *et al*., 1999). Marsh (1988b) divides the roof zone into three parts based on crystallinity and viscosity. At the top is a solid crust where crystallinities are >50%; below this is a zone of crystal mush where crystallinities are between 50% and 25%; and beneath this is a suspension where crystallinities are $<25\%$. The boundary between the suspension and the mush is marked by a rapid increase in viscosity caused by the crystals contacting one another and forming an interconnected network. Cooling of the body causes this boundary to advance and overtake any slowly settling crystals. Bulk convection is possible only in the suspension, but as discussed above it may occur by porous flow through the mush. Marsh believes that only crystals that were initially present in the magma at the time of intrusion, that is, phenocrysts, have a chance of escaping capture by the advancing solidification front. Crystals that nucleate and start growing in the advancing solidification front become trapped.

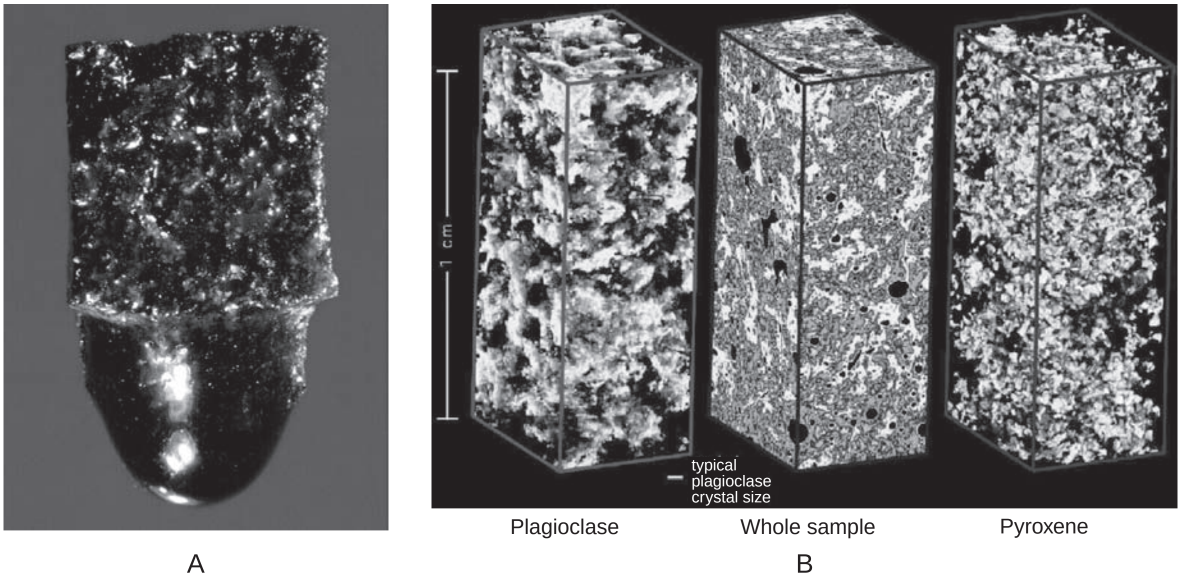
Some sheet-like bodies show clear evidence that material was not moved up or down during solidification (Gibb and Henderson, 1992) and therefore convection could not have occurred. Many alkali basalt sills have central concentrations of olivine that were formed during emplacement (see flowage differentiation, Section 14.6), and these concentrations remain in place during solidification to produce D-shaped modal profiles rather than the S-shaped profiles that would result from crystal settling or sinking plumes. Gibb and Henderson suggest that convection did not occur in such bodies because they may have formed by multiple pulses of magma that inflated the bodies to their final thickness, but at no time was there a thick enough body of magma in which convection could occur.

Sparks (1990) believes that convection in sheet-like intrusions and lava lakes is common and cites as evidence the fact that the final thermal maximum in these bodies is displaced well to the top of the sheet. In a body such as the Palisades Sill, the roof rocks constitute only one seventh the thickness of the sill, and although a few layers of accumulation of early phenocyrstic orthopyroxene occur in the lower part of the sill, the majority of the rock is formed of plagioclase and pyroxene crystals that grew as the sill cooled. This consequently requires that material sank from the roof to the floor during cooling. Similarly, in thick flood-basalt flows, such as the Holyoke basalt of Connecticut (Philpotts *et al*., 1999), which contains only a few percent of neutrally buoyant plagioclase phenocrysts, the entablature/colonnade boundary occurs at two-thirds of the height of the flow. Since the surface of a flow cools more rapidly than its base (see Problem 5.12), the upward displacement of the entablature/colonnade boundary from one-third to two-thirds the height of the flow requires material to have sunk from the roof to the floor during solidification.

In summary, filed evidence indicates that some type of convection occurs in many bodies of magma. Although thermal convection may be weak, that due to compositional change resulting from crystallization can be strong, even though it may operate over only short distances. Convection resulting from dripping instabilities in the roof zone of sheet-like bodies of magma seems to be an efficient way of transferring large quantities of material from the roof to the floor of intrusions.

14.5 CRYSTAL-MUSH COMPACTION

We have seen that despite its attractiveness, crystal settling may not be able to cause magmatic differentiation if crystals are neutrally buoyant, such as plagioclase in basalt, or the magma has a yield strength that can be overcome only by unnaturally large crystals. However, buoyant forces can still cause differentiation and overcome the problems of neutral buoyancy and yield strength if a large number of crystals can act in unison.



Plagioclase Whole sample Pyroxene

Fig. 14.11 (**A**) One-centimeter cube of Holyoke flood basalt that was heated until 70% liquid. The liquid drained out of the bottom of the cube leaving behind a highly permeable crystal mush that still had enough strength to maintain the shape of the cube (Philpotts *et al*., 1998). (**B**) The structure of the mush is revealed by an X-ray CT image. In the center rectangular prism, the darker glass (melt) separates the light colored plagioclase from the intermediate colored pyroxene (see text's web site, www.cambridge.org/philpotts, for animated colored version of this diagram). The right-hand prism shows only the pyroxene crystals, which can be seen to form isolated clusters of grains. The left-hand prism shows only the plagioclase crystals, which form a sponge-like interconnected network that provides the crystal mush with its strength (Philpotts *et al*., 1999).

For example, in the dripping instabilities discussed in the previous section, a large number of crystals, either in a suspension (separate crystals) or acting as a mush (connections between crystals), sink as a dense plume. Although the individual crystals in the plume might not be able to overcome the yield strength of the liquid, the plume itself is large enough that it can exert a force far greater than the yield strength of the liquid. The plume sinks because its bulk density (crystals + liquid) is greater than that of the liquid beneath it. In this way, neutrally buoyant plagioclase can be transported downward as long as it is accompanied by dense minerals such as pyroxene and olivine that make the plume negatively buoyant. From Stokes' law (Eq. (14.4)), the settling velocity is proportional to the square of the radius of the sinking body. Thus, a large plume (radius of meters) can sink very much more rapidly than individual dense crystals (radius of millimeters), even if the magma has no yield strength.

When such a plume arrives on the floor of a magma chamber, it forms a layer of mush consisting of from 25% to 50% crystals. The initial porosity depends on the shape of the crystals. Equidimensional olivine crystals, for example, pack more closely than do plate-like plagioclase crystals. The mush can be quite deformable, and if its bulk density is greater than that of the liquid, it can undergo compaction, with upward expulsion of the interstitial liquid. In this way, differentiation is achieved with separation of crystals from liquid, but without the problems encountered by individual crystals.

Some idea of what such a crystal mush might look like near the base of a basaltic sheet can be gained from examining a partially melted rock from the zone of compaction in the lower part of a thick flood-basalt flow (Philpotts *et al*., 1998). As will be shown below, crystal mush in the lower part of thick flood-basalt flows can undergo compaction. Flood basalts are texturally very simple rocks compared with coarse-grained intrusive rocks, and consequently the textural and chemical effects of compaction are much easier to interpret.

Figure 14.11(A) shows a 1-cm cube of the Holyoke flood basalt from Connecticut that was heated for 24 hours at 1115 °C until 70% melted. At the start of the experiment, the cube was placed over a hole in the bottom of a graphite crucible, so that as the melt formed it was able to drain from the cube and form the large droplet on its lower side. The cube was therefore supported only by its extreme corners. Clearly, the partially melted cube was highly permeable. The remarkable feature of this simple experiment is that, despite the very large amount of melting (70%), the cube maintained its shape due to the structure of the remaining refractory crystal mush. The three-dimensional structure of this mush is revealed in the X-ray CT scan shown in Figure 14.11(B) (Philpotts *et al*., 1999). A 1-cm high rectangular prism from the center of the partially melted cube is shown in the center of the figure. Light colored areas are plagioclase crystals, intermediate gray areas are clusters of pyroxene crystals, and the darker material separating the plagioclase and pyroxene is glass (melt in the experiment). With a CT scan, it is possible to have the computer display each of these phases separately. Thus, on the right of Figure 14.11(B), the image shows all of the pyroxene grains in the prism. Pyroxene tends to occur in isolated clusters. The

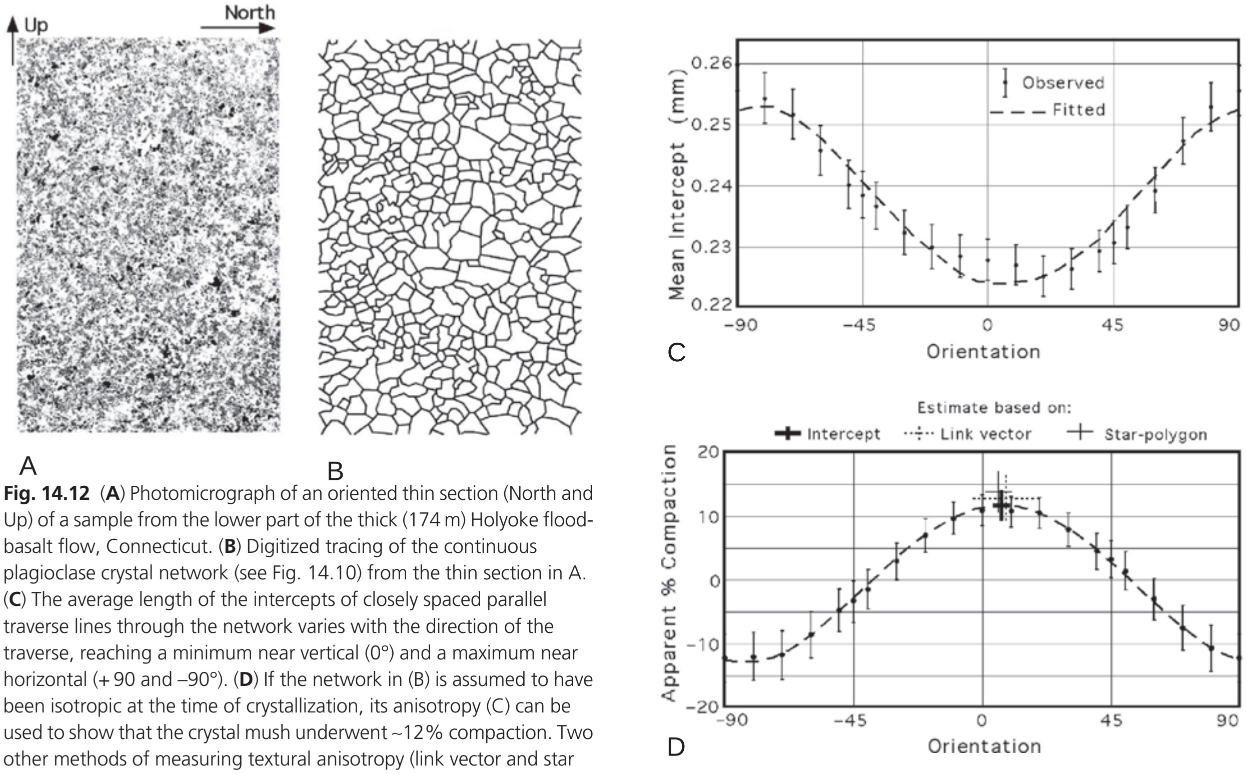


Fig. 14.12 (**A**) Photomicrograph of an oriented thin section (North and Up) of a sample from the lower part of the thick (174 m) Holyoke flood-basalt flow, Connecticut. (**B**) Digitized tracing of the continuous plagioclase crystal network (see Fig. 14.10) from the thin section in A. (**C**) The average length of the intercepts of closely spaced parallel traverse lines through the network varies with the direction of the traverse, reaching a minimum near vertical (0°) and a maximum near horizontal (+ 90 and −90°). (**D**) If the network in (B) is assumed to have been isotropic at the time of crystallization, its anisotropy (C) can be used to show that the crystal mush underwent ~12% compaction. Two other methods of measuring textural anisotropy (link vector and star polygon) indicate the same degree of compaction (see text for discussion).

image on the left of the figure shows the plagioclase crystals, which form a continuous sponge-like three-dimensional network. This network gives the crystal mush in the partially melted cube its coherence. The size of the individual plagioclase crystals making up the network is shown by the white line to the lower right of the plagioclase image. The 3-D nature of the network is more clearly visible in an animated CT image on the book's web site (www.cambridge.org/philpotts).

The continuous 3-D network of plagioclase crystals shown in Figure 14.11(B) is a common texture of slowly cooled basaltic rocks and gabbros (see Fig. 14.24(B)). When the crystal mush first develops, this network tends to be isotropic; that is, it shows no preferential shape. However, if the mush undergoes compaction, the network becomes flattened. The degree of flattening, and consequently the amount of compaction, can be measured directly from thin sections.

Figure 14.12(A) shows an oriented thin section (North and Up are marked) from the lower part of the Holyoke flood-basalt flow, which was used in the melting experiment shown in Figure 14.11. A high resolution scanned image of the thin section is imported into a drawing program on a computer and the plagioclase network is traced by drawing lines along the center line of the clusters of plagioclase crystals. This is a tedious task, but some computer programs will do it for you (e.g. the skeletonize function in NIH Image). The result is a digitized tracing of the plagioclase network as shown in Figure 14.12(B). We must now determine whether this network is random (isotropic) or is flattened, and if so, to obtain a quantitative measure of the flattening. One way of doing this is to draw a series of evenly spaced parallel lines through the network, say in a vertical direction, and then measure the length of all the intercepts the traverse lines make with the network. In this thin section, the average intercept along vertical lines is 0.225 mm. The direction of the traverse lines can then be rotated to a new angle and the measurements repeated. In this way we find that the average intercept in this thin section steadily increases to a maximum of ~0.25 mm in a horizontal direction (Fig. 14.12(C)). If the plagioclase network were initially isotropic, this degree of flattening of the network corresponds to ~12% compaction. The degree of flattening can also be measured by determining the mean of all the ellipses that can be fitted to the polygonal areas in the digitized drawing of the network. This result is shown as "star polygon" in Figure 14.12. The anisotropy can also be measured by examining the orientation and length of all the line segments used to draw the polygonal network. Each line segment is a vector and can be represented by a point in an *xy* (North–Up) plot. The resulting scatter of points can then be analyzed statistically and a moment ellipse fitted to the data from which we can measure the compaction. This results in the "link vector" point shown in Figure 14.12. These three independent ways of measuring textural anisotropy all

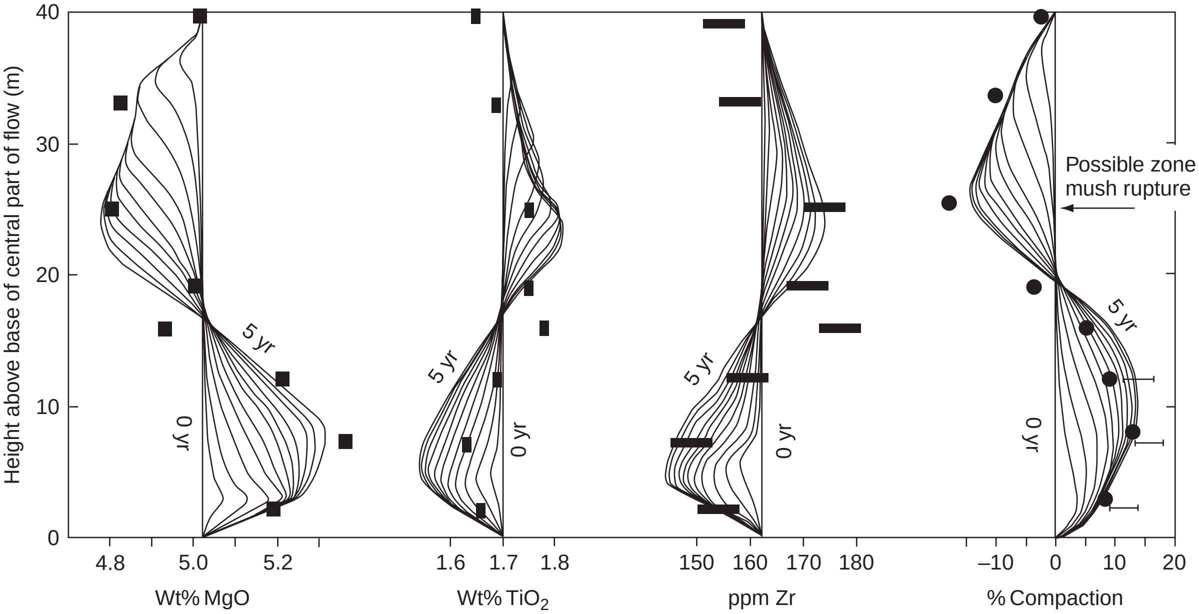


Fig. 14.13 Evolution of MgO, TiO_2 and Zr and amount of compaction/dilation during solidification of the central part of the Cohassett flow according to the IRIDIUM program of Boudreau (2003). Results are presented in 5-month time steps for the first five years, after which there is no significant change. The measured chemical profiles are shown by the black rectangles. The compaction measurements based on the amount of interstitial glass are shown as black circles and those based on the anisotropy of the plagioclase crystal network (Fig. 14.11) are shown by horizontal lines (Philpotts and Philpotts, 2005).

indicate that this particular sample of basalt underwent ~12% compaction. The details and computer programs for making these measurements are given by Gray *et al*. (2003).

The physical process of compaction has been quantitatively modeled by McKenzie (1984) and Shirley (1987) (see Section 3.9). The rate of compaction is determined by the permeability and the viscosity of the solid matrix. At low degrees of crystallinity, the mush is highly permeable and weak, so that compaction is rapid, but as crystallization progresses, the permeability decreases, the mush becomes stronger, and the compaction rate decreases. Boudreau (2003) has combined McKenzie's and Shirley's treatment of the physical aspects of compaction with the chemical equilibrium determined from the MELTS program to create a comprehensive model of the compaction process. This computer program, known as IRIDIUM, can be downloaded from the web at www.nicholas.duke.edu/people/faculty/boudreau/DownLoads.html. The program allows one to trace the compositional changes that take place in a sheet of liquid as it crystallizes and undergoes compaction.

To illustrate the results given by the IRIDIUM program and to indicate the efficacy of compaction in causing differentiation, we will examine the effects of compaction during the solidification of a thick Columbia River flood-basalt flow (Philpotts and Philpotts, 2005). The Cohassett flow is a ~70-m-thick flood basalt, which covers >5000 km^2 of south-central Washington. It attained its final thickness following several pulses of inflation, the last of which formed the central 40 meters of the flow. Within this central zone, the compatible and incompatible elements form, respectively, S- and Z-shaped profiles, which are characteristic of compaction (Fig. 14.13). For example, compaction enriched the lower part of this central zone with early crystallizing pyroxene and plagioclase, thus enriching it in MgO and Al_2O_3. Residual liquid expelled from the zone of compaction rose into, and dilated, the overlying crystal mush, thus enriching it in incompatible components, such as TiO_2 and Zr. Direct measurement of the amount of compaction and dilation were obtained using the intercept, star-polygon, and link-vector methods outlined above and by direct measurement of the amount of interstitial liquid (Fig. 14.13). The original composition of the magma in the central part of the flow was obtained by integrating the chemical profiles through this zone. This constituted the starting composition for the IRIDIUM program. The sheet of magma was assumed to have cooled at 2 °C day^{-1} from the top and 0.2 °C day^{-1} from the bottom. After reasonable values were selected for the permeability and viscosity of the crystal mush, the IRIDIUM program tracked how the composition of the sheet of basalt changed as it underwent compaction and solidified. Figure 14.13 shows the profiles of MgO, TiO_2 and Zr, and the amount of compaction or dilation (negative compaction) through the central part of the flow in 5-month time steps for the first five years, after which there is no significant change. The computer model matches closely the chemical and compaction/dilation profiles. In still thicker flows, compaction has time to expel more liquid from the zone of compaction, which results in the crystal mush in the overlying zone of dilation being ruptured with the formation of sheets of residual liquid (Boudreau and Philpotts, 2002). If compaction can cause differentiation in thick lava flows, its effects should be far greater in slowly cooled intrusive bodies.

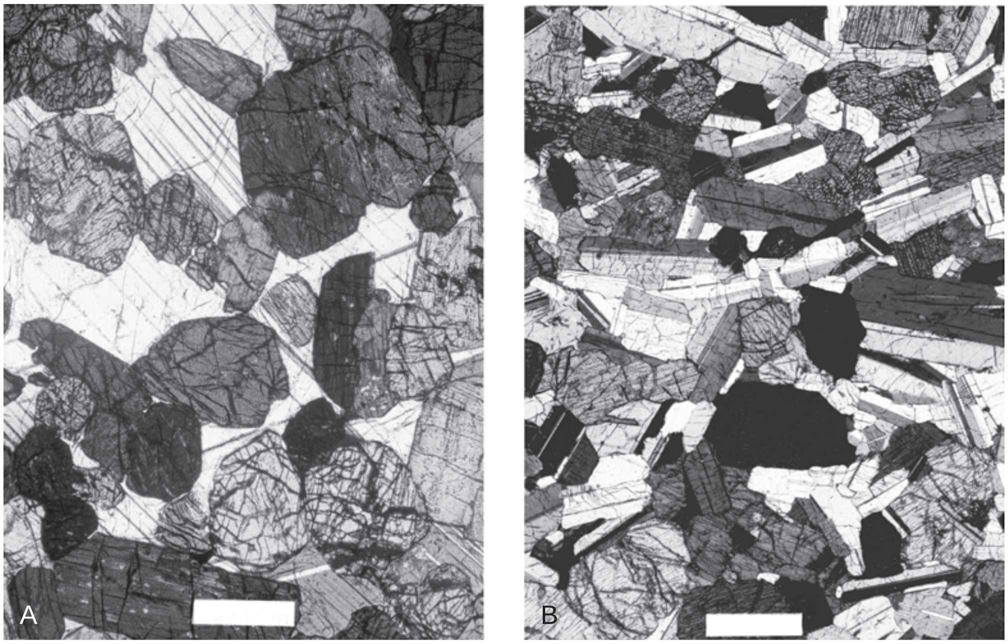


Fig. 14.14 Photomicrographs of two rocks composed of pyroxene and plagioclase from the Great Dyke of Zimbabwe. In (**A**), pyroxene crystals accumulated from the magma, whereas plagioclase crystallized from the intercumulus liquid to form one single large poikilitic grain that cements the pyroxene grains in place. In (**B**), plagioclase was also a cumulus phase; many plagioclase laths aligned themselves parallel to the floor of the intrusion during accumulation possibly as a result of compaction. Crossed polars; scale bars are 1 mm.

14.6 IGNEOUS CUMULATES

The compositions of certain igneous rocks differ significantly from those of known magmas that reach the surface as lavas. These rocks are enriched in early crystallizing minerals, and their textures (Fig. 14.14) provide clear evidence that crystals of these minerals were concentrated by accumulation. Rocks of this type, therefore, are referred to as *cumulates* (Wager *et al*., 1960). They were originally thought to form by crystals sinking to the floor of a magma chamber. This also accounted for the prominent horizontal or gently dipping layering that most cumulates exhibit (Figs. 14.16 and 14.18), which was interpreted to result from magmatic sedimentation. Although such a process may occur, it cannot be the only means by which cumulates form, because some contain plagioclase crystals that would certainly have floated in the magmas from which they crystallized. Regardless of the process by which crystals are concentrated in cumulates, fractional crystallization is clearly involved in the formation of these rocks.

Cumulates have a characteristic texture formed by two contrasting types of mineral grains (Irvine, 1982). First, there are the more abundant *cumulus* grains, which tend to be euhedral and rest against each other to form a cumulus framework. Between these grains are minerals that form during a *postcumulus* period from the *intercumulus* liquid. A randomly stacked pile of crystals would have a porosity of 25% to 50%, and cumulates that contain this volume of postcumulus material are described as *orthocumulates* (Fig. 14.15). Some cumulates, however, contain far less postcumulus material, and these are termed *mesocumulates* if the percentage is between 7% and 25% and *adcumulates* if it is less than 7%. In mesocumulates, cumulus grains become enlarged by overgrowths, and many grains contact one another along mutual interference boundaries rather than at the point contacts that are more typical of cumulate grains in orthocumulates. In adcumulates, all cumulate grains contact

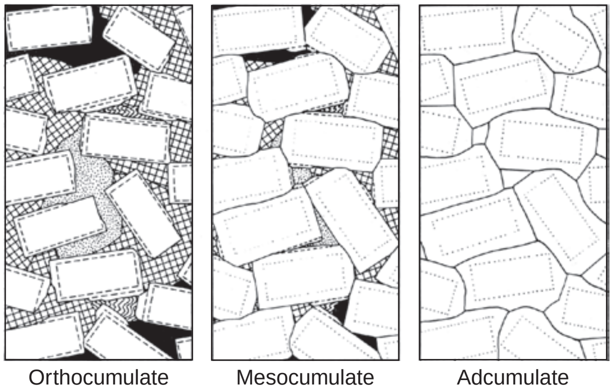


Fig. 14.15 Three different types of plagioclase cumulate. In each, plagioclase is the only cumulus phase. In the *orthocumulate*, intercumulus liquid crystallizes to an intergrowth of pyroxene, olivine, and magnetite; a small amount of plagioclase forms thin postcumulus rims on cumulus plagioclase grains. In the *mesocumulate*, overgrowths on cumulus grains are thicker, and the percentage of other postcumulus minerals is greatly decreased. In the *adcumulate*, the amount of postcumulus plagioclase added to cumulus grains is so great that no significant amount of intercumulus minerals is present. (After Wager *et al*., 1960; published with permission of Oxford University Press.)

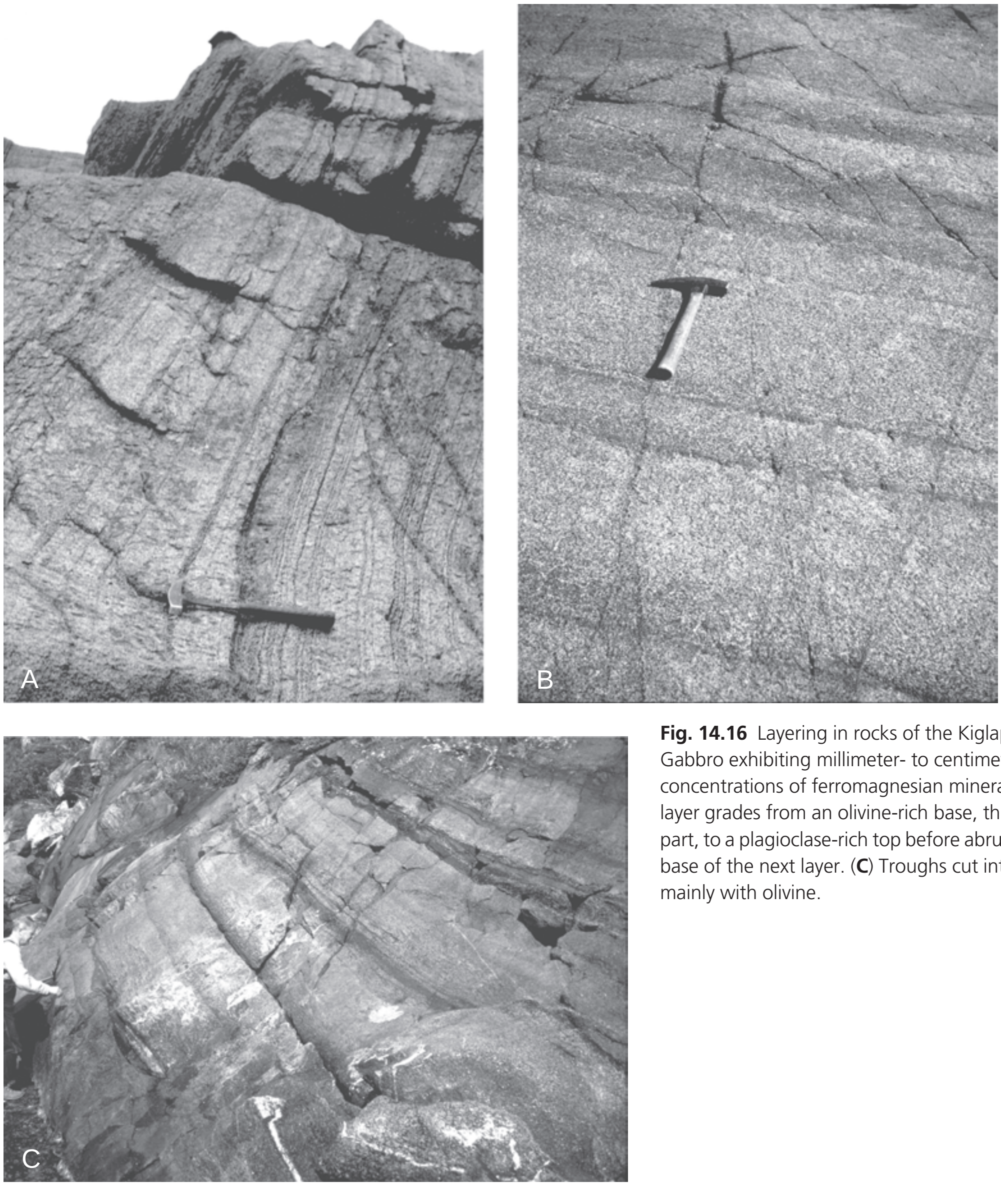


Fig. 14.16 Layering in rocks of the Kiglapait Intrusion, Labrador. (**A**) Gabbro exhibiting millimeter- to centimeter-scale layering produced by concentrations of ferromagnesian minerals. (**B**) Graded layering; each layer grades from an olivine-rich base, through a pyroxene-rich central part, to a plagioclase-rich top before abruptly returning to the olivine-rich base of the next layer. (**C**) Troughs cut into layered gabbro are filled mainly with olivine.

one another along boundaries that are formed either by mutual interference of overgrowths, or by recrystallization or pressure solution of the original grains. Holness *et al.* (2005) have shown that when cumulates are first formed, the faces bounding cumulate grains are usually prominent crystal faces, but as the cumulate matures, these faces are replaced by curved ones that attempt to minimize the interfacial energies, and as a result characteristic dihedral angles develop between the grains (Section 12.7).

An additional, less common type of cumulate known as a *crescumulate* has the cumulus grains grow normal to the layering in which they occur. The resulting texture is similar to the comb structure in pegmatites where minerals nucleate on the wall of the pegmatite and grow inward (Fig. 11.15(B)). Crescumulates can involve felsic as well as mafic minerals. In the steeply dipping marginal border group of the Skaergaard Intrusion, sheets paralleling the contact of the intrusion contain bladed 2.5-cm-long plagioclase crystals oriented perpendicular to the contact. These crystals are believed to have grown during periods of quiescence of the magma; convection of magma down or up the wall would have deflected the crystals. Olivine crystals in some of the peridotite layers of the Rum Intrusion grew perpendicular to the layering, forming vertically elongated crystals that branch upward (Fig. 4.71). This was originally described as a *harrisitic structure*. Apparently, during periods of slow accumulation of cumulus grains, olivine crystals resting on the floor of the intrusion that had their *a*-axis oriented approximately normal to the floor grew more rapidly than did those with other orientations (Wager and Brown, 1968). O'Driscoll *et al.* (2007), however, believe that these elongate olivine crystals grew very rapidly (days) in magma that became supersaturated in olivine. In this respect they resemble the olivine crystals that grow down from the surface of komatiite lava flows to produce the spinifex texture (Section 15.8 and Figs. 15.21 and 12.7(A)).

The common cumulus minerals are olivine, pyroxene, plagioclase, chromite, and magnetite, and the postcumulus minerals typically include pyroxene, plagioclase, magnetite, ilmenite, hornblende, and granophyric intergrowths of quartz and alkali feldspar. The compositional differences between the cumulus and intercumulus grains are consistent with an origin through fractional crystallization; that is, cumulus ferromagnesian minerals are more magnesian than postcumulus ones, and cumulus plagioclase is more calcic than intercumulus plagioclase. Compositional zoning on the rims of cumulus grains is common, especially where there have been adcumulus overgrowths. This is particularly true of plagioclase and to a lesser extent of pyroxene. Olivine homogenizes rapidly enough that zoning is much less common. Cumulates are described in terms of their cumulate phases, which are given in order of decreasing abundance. Thus, describing a gabbro as a plagioclase–olivine cumulate tells the reader that this rock contains cumulate grains of plagioclase and lesser olivine, and that the other minerals, such as pyroxene and magnetite, occur as postcumulus grains.

The most common occurrence of cumulates is in layered gabbroic intrusions, but they also occur in some syenitic and granitic intrusions. Reviews of these occurrences and how the layering may have formed are given by Wager and Brown (1968) and Naslund and McBirney (1996). Much of what is known about cumulates comes from the study of tholeiitic layered gabbroic bodies, such as the Skaergaard Intrusion of East Greenland (Wager and Deer, 1939), the Stillwater Complex of Montana (Hess, 1960), the Muskox Intrusion of the Northwest Territories, Canada (Irvine and Smith, 1967; Irvine, 1980), the Bushveld Intrusion of South Africa and the Great Dyke of Zimbabwe (Rhodesia) (Jackson, 1967), the Kiglapait Intrusion of Labrador (Morse, 1969), the Duke Island Ultramafic Complex of southwestern Alaska (Irvine, 1974), and the Rum Complex of the Inner Hebrides, Scotland (Emeleus, 1987). Cumulates are also found in alkaline gabbroic and syenitic rocks, as for example in those from the Gardar province of southwest Greenland (Sørensen, 1974b) and the Monteregian province of Quebec (Philpotts, 1974).

The layering exhibited by most cumulates ranges in thickness from millimeters to meters and can result from variations in modal abundance of cumulus minerals, grain size, texture, or composition of minerals (Figs. 14.16, 14.18, 4.67, 4.70). Because variations in mineral composition are normally not evident until analyses have been done, such layering is referred to as *cryptic layering*. Variations in modal abundance and grain size are commonly gradational so are described as *graded layers* (Figs. 14.16(B), 14.17, 14.18). In many layers, dense ferromagnesian minerals are concentrated at the base of the layer and their abundance decreases gradually upward as the abundance of less dense minerals (usually feldspar) increases. The sequence then commonly repeats itself to produce *rhythmic layers* (Fig. 4.72). If the sequence of appearance of cumulus grains in passing up though rhythmic layers is identifiable as the fractional crystallization sequence of the minerals involved, the layering is said to be *cyclic*. Cyclic units are typically formed by injections of new batches of undifferentiated magma.

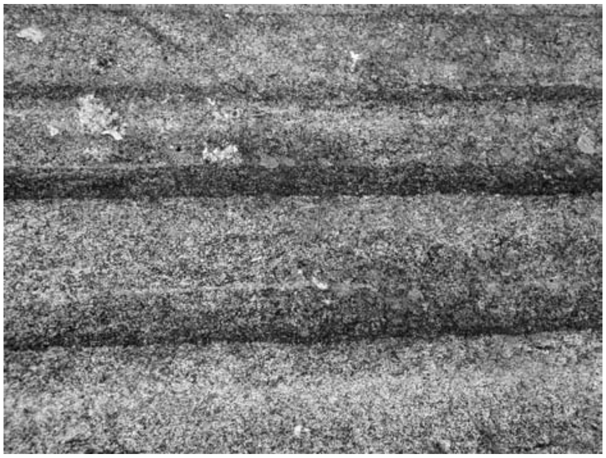


Fig. 14.17 Detail of graded layering in the gabbro of Stavanger, Norway, shown in Figure 4.72. Height of image is 0.75 m.

In large gabbroic intrusions, layers generally dip at shallow angles toward the center of the intrusion (Fig. 14.18). Toward the margins, dips may increase and even become so steep that the layers become unstable and slump down onto the floor of the intrusion, where they disrupt the flat-lying layers (Fig. 4.70). In other cases, channels may cut down through the cumulate layers to produce troughs that later are filled with other cumulates that commonly exhibit strong graded compositional layering (Fig. 14.18(B)). Blocks of related rock derived from elsewhere in an intrusion (cognate xenoliths) or blocks of wall rock (xenoliths) are commonly found in layered cumulates (Fig. 4.68). The way these indent the underlying layers and have other layers deposited around and draped over them has been used by Irvine *et al*. (1998) to argue that in the Skaergaard Intrusion the blocks must have been deposited on a sharp interface between cumulates below and magma above. The disruption or erosion of cumulate layers typically affects cumulates to depths of several meters, providing important evidence about the thickness of crystal mush on the floor of these intrusions.

Graded layers, which are very common in the Skaergaard Intrusion (Fig. 14.18(A)), were interpreted by Wager and Deer (1939) to have had a similar origin to graded beds in sedimentary rocks that are formed by turbidity currents. They postulated that dense slurries of crystals descending the walls of the intrusion spread across the floor where the denser ferromagnesian minerals sank first, leaving the less dense feldspar crystals to be deposited on top. The sedimentary analogy is made still more convincing by the fact that layers commonly truncate underlying layers (Fig. 4.67). In places channels may be cut into layers and then filled with more layered rock. In some cases, this even resembles cross-bedding in sedimentary rock. Graded layering was one of the first pieces of evidence used to argue that cumulates are formed by gravity settling. This interpretation became very popular and is still thought by some to be the main cause of layering in cumulates (e.g. Irvine, 1987; Irvine *et al*., 1998).

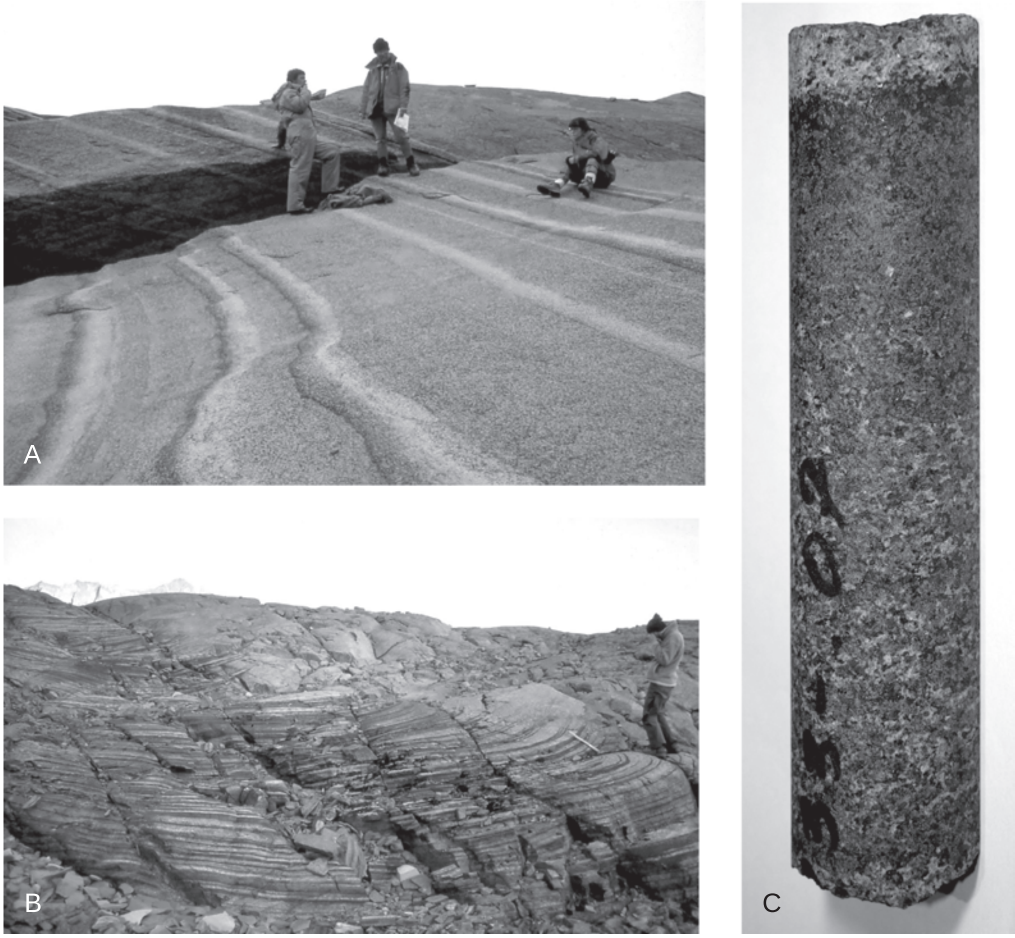


Fig. 14.18 Layering in the Layered Series of the Skaergaard Intrusion, East Greenland. (**A**) Outcrop exhibiting three types of gently dipping modal layering: (1) layers grading from mafic at the base to felsic at the top (lowest three); (2) homogeneous layers composed of average gabbro in near-cotectic proportions; (3) plagioclase-rich layers with no corresponding mafic component. (**B**) Layered sequence forming a trough-like structure trending away to the upper left. The layering thins and disappears as it leaves the trough and becomes homogeneous gabbro (in front of geologist). (**C**) Drill core through reversely graded layer. Pyroxene, olivine, and opaque oxides increase gradually upward as plagioclase decreases, ending with a layer dominated by oxides in the upper ~0.5 cm; the layer is abruptly overlain by a layer of leucogabbro. Core sample is 16 cm long. (Photographs by Christian Tegner, Aarhus University.)

There are many problems, however, with interpreting layering as the result of gravitative settling. In some layered intrusions, graded layers are present, but they grade in the reverse direction; that is, they become more feldspathic downward. Even in the Skaergaard Intrusion (Fig. 14.18(C)), reverse grading is not uncommon (Christian Tegner, personal communication). A detailed study of modal layering in the Fongen-Hyllingen Intrusion, Norway, shows that 50% of the layers are ungraded, 15% are normally graded, but 35% are reversely graded (Josephsen, 2003). In the central part of the Palisades Sill, all of the “inch-scale” layering (see below) is reversely graded (Dickson, 2006), even though these layers exhibit similar cross-bedding features to those exhibited by the normally graded layers in some other intrusions (compare Fig. 14.19 with Fig. 4.67). Figure 14.20 shows profiles in the relative concentrations of feldspar crystals in the normally graded layering shown in the gabbro from Stavanger, Norway (Fig. 14.17) and the reversely graded layers in the Palisades Sill (Fig. 14.19(B)). When the Pallisades’ profile is reversed it is remarkably similar to that in the Stavanger example except that the length scale is an order of magnitude less.

All of the sedimentary-like features that are found in the shallowly dipping layering can also be found in vertically dipping layers near the walls of intrusions where gravity settling could not possibly be a factor. Many alkaline gabbroic intrusions form vertical pipe-like bodies that show prominent graded layering parallel to the contacts. These layers commonly start with a high mafic mineral content and then grade to more feldspathic compositions toward the center of the intrusion (Fig. 14.21(A)). This layering is commonly truncated to produce either “cross-bedding” (Fig. 4.75) or troughs (Fig. 14.21(B)). Even in the steeply dipping layers at the margin of the Skaergaard Intrusion, similar features are found. Such evidence led McBirney and Noyes (1979) to propose that layering in the Skaergaard Intrusion formed by the nucleation and growth of minerals along the walls, roof, and floor of the intrusion in a static boundary layer in which the minerals were trapped in a Bingham liquid.

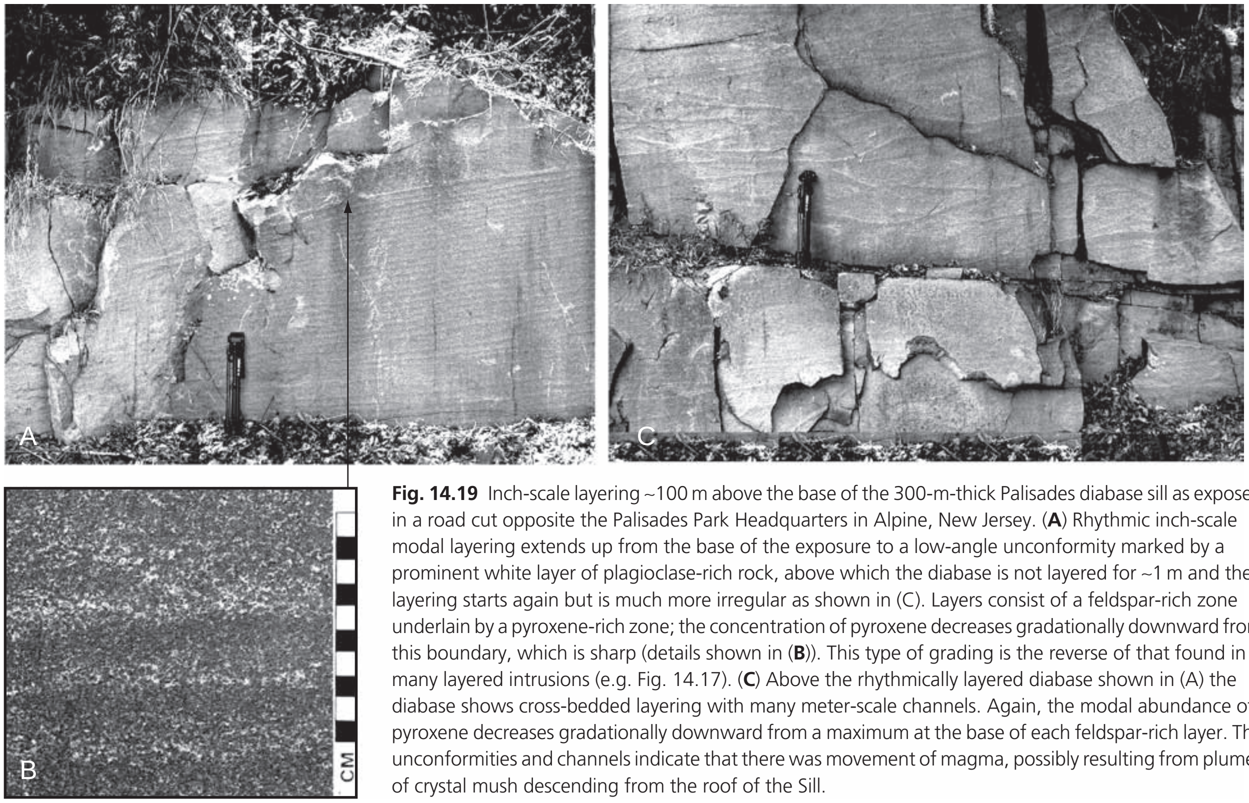


Fig. 14.19 Inch-scale layering ~100 m above the base of the 300-m-thick Palisades diabase sill as exposed in a road cut opposite the Palisades Park Headquarters in Alpine, New Jersey. (**A**) Rhythmic inch-scale modal layering extends up from the base of the exposure to a low-angle unconformity marked by a prominent white layer of plagioclase-rich rock, above which the diabase is not layered for ~1 m and then layering starts again but is much more irregular as shown in (C). Layers consist of a feldspar-rich zone underlain by a pyroxene-rich zone; the concentration of pyroxene decreases gradationally downward from this boundary, which is sharp (details shown in (**B**)). This type of grading is the reverse of that found in many layered intrusions (e.g. Fig. 14.17). (**C**) Above the rhythmically layered diabase shown in (A) the diabase shows cross-bedded layering with many meter-scale channels. Again, the modal abundance of pyroxene decreases gradationally downward from a maximum at the base of each feldspar-rich layer. The unconformities and channels indicate that there was movement of magma, possibly resulting from plumes of crystal mush descending from the roof of the Sill.

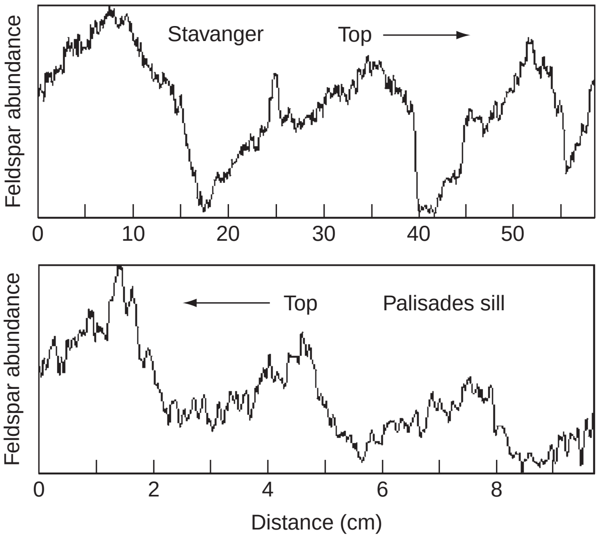


Fig. 14.20 Abundance of feldspar in the graded layers of the gabbro from Stavanger, Norway (Fig. 14.17) and the diabase of the Palisades Sill, New Jersey (Fig. 14.19(B)) as determined by the profiling function in NIH Image (see Problem 14.16). The profiles are similar, but the top is to the right in the Stavanger plot and to the left in the Palisades plot, which means the latter is reversed compared with "normal" graded layering. Note also the different length scales.

Brandeis *et al*. (1984) have shown that the rhythmic layering in cumulates can be developed in such a boundary layer by the interplay of crystal nucleation and the dissipation of the heat of crystallization. Once nucleation occurs, the latent heat of crystallization raises the temperature and prevents further nucleation; crystal growth therefore takes place only on the already formed nuclei. The rate of this growth is determined by the rate at which heat is dissipated from the boundary layer. As the crystals approach their final size, which is determined by phase equilibria and the temperature, they grow more slowly, and the temperature is able to begin to fall. Eventually, a second batch of nuclei form and the process repeats. Differences in nucleation and growth rates produce graded modal abundances across each layer. Calculations indicate that layers produced by this process would be several centimeters thick. This could explain the origin of the common rhythmic layering of cumulates, referred to as "*inch-scale*" layering (Fig. 14.16(A)).

In general, cumulate layering can be formed by two classes of processes, those which are related to flow of magma and those that take place in static magma or piles of crystal mush (McBirney and Nicolas, 1997). Slump structures, cross-bedding, and folds are features that clearly are produced by the flow of magma or the collapse of piles of crystal mush. Flow can align crystals and develop a lineation (Fig. 4.69). This *dynamic layering* can be quite variable in thickness. *Nondynamic* layering can form near walls or roofs where differences in nucleation or crystal growth rates can cause transitory excursions from cotectic crystallization or in piles of compacting crystal mush where pressure solution and textural coarsening can redistribute minerals into layers (Boudreau and McBirney, 1997). Although nondynamic layering can develop strong modal variations, the bulk composition of the layers remains on the cotectic.

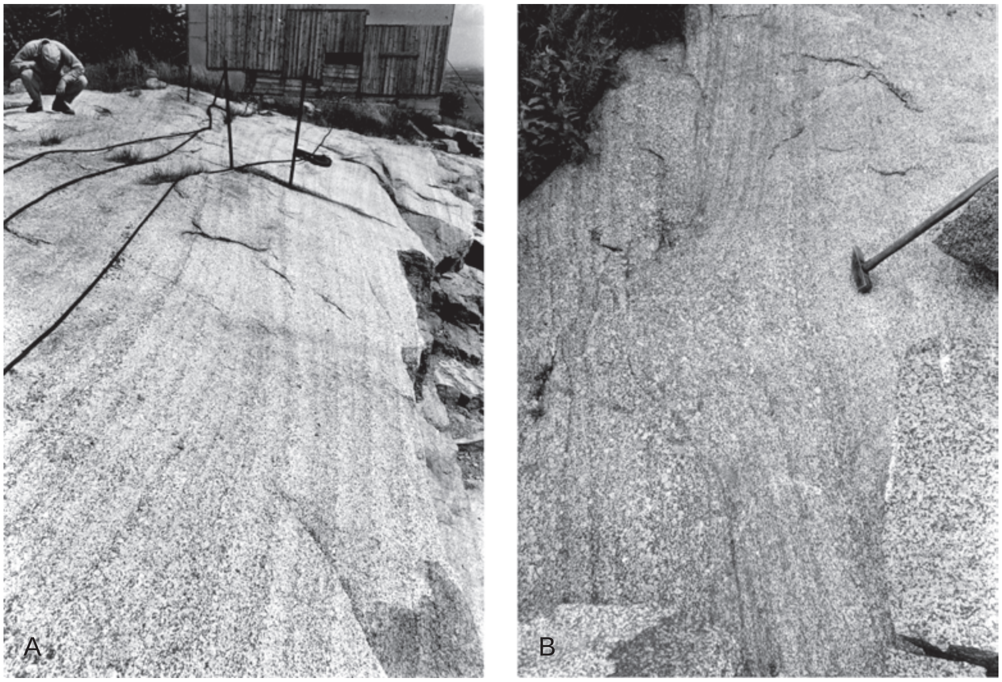


Fig. 14.21 Vertical layering paralleling the walls of the Monteregian alkaline intrusion, at Mount Johnson, Quebec (Philpotts, 1968). In (**A**), layering is graded with respect to abundance of mafic and felsic minerals (center of intrusion to left). Because no tilting has occurred since intrusion, the compositional differences producing the vertical layering cannot be the product of gravity settling. In (**B**), the layered gabbro is cut by a vertically plunging trough that is filled by a younger layered gabbro (center of intrusion to the right).

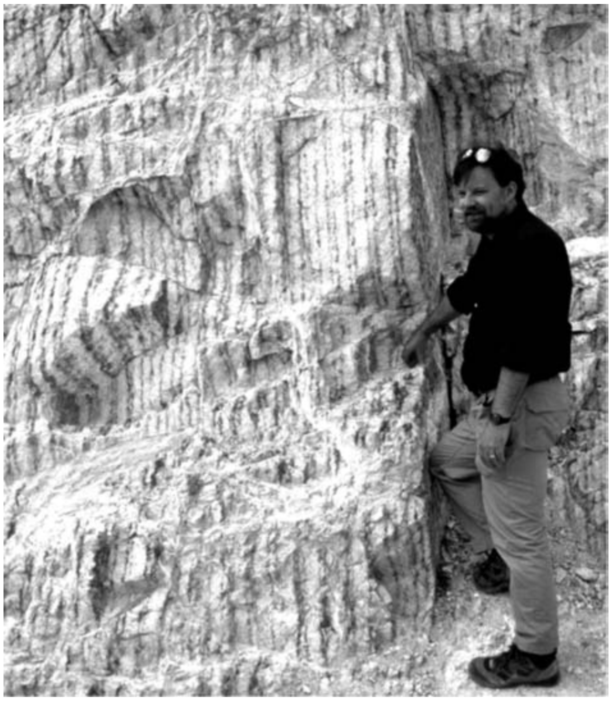


Fig. 14.22 Fine-scale layering in leuco-norite of the Stillwater Complex, Montana. The layers were originally horizontal with top to the left. Note that the layers occur in doublets. Alan Boudreau of Duke University for scale.

In compacting piles of crystal mush, solution and redeposition of minerals, textural coarsening (Ostwald ripening), and flux of residual liquid expelled from the pile can lead to the development of "inch-scale" layering (Boudreau, 1995). Some "inch-scale" layering forms such remarkable repetitive patterns that it is not likely to have formed by any sedimentation process. One of the most remarkable examples of this type of layering is found in the norite of the Stillwater Complex, Montana (Fig. 14.22). In this rock, almost all of the orthopyroxene has been segregated into layers that form doublets and are separated by rock consisting almost entirely of plagioclase. Boudreau has shown that this type of layering can result from the redistribution of minerals during compaction and coarsening of a pile of crystal mush.

Boudreau likens the process of layer formation to that which forms Liesegang rings. If an initially very small difference in grain size exists, aging (Ostwald ripening) causes smaller crystals to dissolve and reprecipitate on the larger ones. In a pile of crystal mush consisting of more than one mineral, the aging process will not be the same for each mineral because of differences in aging kinetics and the rate of diffusion of components through the interstitial liquid. In the case of the norite, for example, the diffusion of components necessary to redistribute the pyroxene are likely to be faster than those necessary to redistribute the plagioclase (aluminum diffuses slowly).

If the pyroxene crystals at the bottom of the pile are slightly coarser than those above, perhaps because they formed first, the overlying smaller crystals would dissolve and their components diffuse downward to precipitate on the larger crystals. However, as the small crystals become smaller, the concentration of their components in the melt increases because of the higher surface energy on the smaller grains. This causes both downward and upward diffusion of components away from the zone of higher concentration. The downward diffusing components are added to the already enlarged crystals, whereas the upward diffusing components are added to crystals above, and a new coarsening layer develops. The process then repeats, with multiple layers being formed. As a coarsening layer is fed by diffusion from both above and below, its upper and lower boundaries coarsen faster than its central part. Once this happens, the central part of the layer begins to dissolve and reprecipitate in the upper and lower boundary zones. In this way a doublet is formed. Boudreau's mathematical model of this process accounts for many of the features and spacings of cumulate layers.

The inch-scale modal layering in the central part of the Palisades Sill has been shown to result from redistribution of pyroxene in a compacting crystal mush of cotectic composition (Dickson, 2006). Clinopyroxene in the sill has three different modes of occurrence: as phenocrysts, aggregates of polygonal grains of augite and pigeonite, and as ophitically intergrowth crystals. The granular aggregates are thought to have formed from oikocrysts of subcalcic augite that sank from the roof of the sill to the floor in dense plumes of crystal mush. Where the diabase is layered, the granular aggregates are completely absent from the feldspar-rich layers, but the underlying rock is enriched in ophitically intergrown pyroxene in the same proportion that the granular pyroxene is absent from the feldspar-rich zone. The result is that the bulk composition of these couplet layers remains on the cotectic. Evidence that layering in the Palisades Sill was at least augmented by processes that redistributed minerals in the pile of crystal mush can be found where penecontemporaneous faults cut the layering (Fig. 14.23). These faults, like the layering, are marked by feldspar-rich and pyroxene-rich zones.

Although the concept of cumulus and intercumulus minerals is simple, the distinction between such grains is often difficult to make in practice. If the intercumulus liquid crystallizes to distinct mineral grains, its identification is easy. Commonly, the intercumulus minerals nucleate at widely separate locations and grow around the cumulus minerals to form very large crystals that may contain many cumulus grains. These large crystals that grow late are known as *oikocrysts*, to distinguish them from phenocrysts, which grow early. A rock with oikocrysts has a *poikilitic* texture (Figs. 14.14(A), 14.24(A) and (B)). Some intercumulus liquid may crystallize as overgrowths on cumulus grains, and then the identification of the postcumulus material must depend on textural interpretations or the identification of compositional zoning in the regions of overgrowth. In still other cases, the intercumulus liquid may react with the cumulus grains, as happened in the peridotite from the Stillwater Complex, where olivine was rimmed by orthopyroxene (Fig. 14.24(A)).

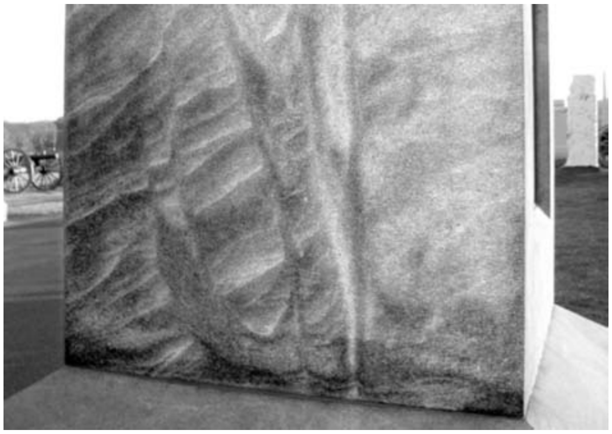


Fig. 14.23 Penecontemporaneous faults cutting layered diabase in the Monument to the Connecticut 14th Regiment on the Gettysburg Battlefield, Pennsylvania. Layering caused by redistribution of pyroxene parallel to the faults shows that some layering can form in piles of crystal mush prior to its complete solidification.

When cumulates contain less than 25% intercumulus material, some process other than random packing of cumulus grains must be involved in their formation. For example, the intercumulus liquid may have been expelled from the crystal mush, or this liquid's composition may have changed so that cumulus minerals could continue crystallizing from it, or the initial cumulate pile may simply have formed with a lower percentage of intercumulus liquid. Petrologists do not yet agree on which of these mechanisms is the most important, but most believe that some change in composition of the intercumulus liquid takes place so as to allow the cumulus grains to continue growing – hence the term adcumulate. Of course, several mechanisms may operate simultaneously, any one of which may play the dominant role in a particular case.

Originally, adcumulates were believed to form when compositional imbalances between the intercumulus liquid and the main body of magma were set up by the continued crystallization of cumulus minerals from the intercumulus liquid. For example, olivine crystallizing onto cumulus olivine grains would enrich the residual liquid in the other components of the melt. These higher concentrations would cause these components to diffuse up into the overlying magma. At the same time, olivine components in the main body of magma would diffuse down into the cumulate pile to replace the olivine components that had already crystallized onto the cumulate grains. With this two-way diffusion process the cumulate could eventually be converted into a monomineralic olivine rock – dunite. Because diffusion rates are so low, the distance between the intercumulus liquid and the interface with the main body of magma would have to be small for this process to operate. This, in turn, would require a slow rate of accumulation of cumulus grains. Morse (1986) estimates that rates should not exceed 0.5 cm a^{-1} if adcumulus growth is to produce monomineralic rocks by this diffusion process.

The change in composition of the intercumulus liquid that causes diffusion may also set up convection. Indeed, the

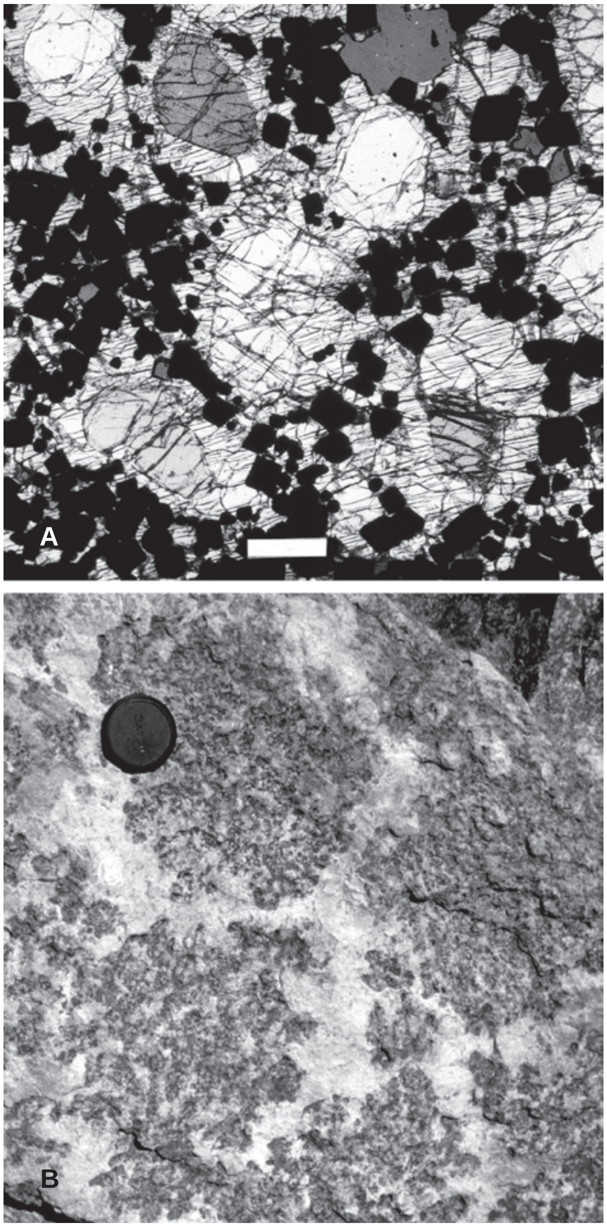


Fig. 14.24 Oikocrysts in the Stillwater Complex, Montana. (**A**) Photomicrograph of harzburgite, where the original outlines of cumulus olivine grains are marked by the position of chromite grains, which accumulated with olivine. Interstitial liquid reacted with olivine to produce postcumulus orthopyroxene, which poikilitically enclosed all cumulus grains within this field of view in a single oikocryst (note common cleavage direction). Crossed polars; scale bar is 1 mm. (**B**) Cumulus plagioclase are enclosed in basketball-size pyroxene oikocrysts, which themselves are surrounded by plagioclase crystals to produce an extremely coarse version of the texture revealed by X-ray CT scans of the basalt shown in Fig. 14.11.

change in density of a liquid resulting from crystallization (Fig. 3.1 and Problem 14.15) is likely to provide a far greater driving force for exchange of intercumulus liquid than is diffusion. For convection to bring about an exchange with overlying magma, the residual liquid must be less dense than the overlying magma. This is true when crystallization of ferromagnesian minerals brings about enrichment of the residual liquid in felsic components, but when plagioclase is the cumulus mineral residual liquids are more dense and thus would remain *in situ* among the cumulus plagioclase grains on the floor of the intrusion. Convective removal of intercumulus liquid from plagioclase cumulates can therefore not be called upon to produce monomineralic layers of anorthosite in layered intrusions. This mechanism, however, may be an effective means of replenishing intercumulus liquid in mafic cumulates.

Irvine (1980) showed from whole-rock analyses and the composition of cumulus minerals that the interstitial liquid in cyclical cumulate units formed from repeated intrusions of batches of unfractionated magma in the Muskox Intrusion was displaced upward by as much as 20 m. The first cumulus mineral to form from each batch of magma was olivine, followed by clinopyroxene, and then plagioclase. The emplacement of a new batch of magma is recorded in the magmatic stratigraphy by the abrupt appearance of olivine (and minor chromite) as the only cumulus phase. Olivine forming from each new pulse of unfractionated magma should have a higher magnesium content than that which formed in the fractionated magma at the top of the underlying unit. Irvine found, however, that the increase in magnesium invariably occurred above the stratigraphic level at which olivine appears as the only cumulus phase. He concluded, therefore, that the more iron-rich residual intercumulus liquid in the underlying rock was displaced upward into the new cumulate layer of olivine, where it reacted with the olivine to lower its magnesium content. He proposed that this *infiltration metasomatism* results from the upward migration of intercumulus liquid expelled from the underlying cumulates by *compaction*. Although the composition of the intercumulus liquid may be changed in this way, it does not necessarily produce adcumulates because this liquid still contains other components of the magma.

Compaction of cumulus grains can reduce porosity by pressure solution and redeposition and by modifying dihedral angles to minimize interfacial energies (Holness *et al*., 2005). The result can be a rock with an exceptionally high concentration of cumulus minerals, which would then be described as an adcumulate, even though no postcumulus material need have been added to the cumulus grains. The compaction process, which has been dealt with in detail by McKenzie (1984), results from the imbalance between the load pressure on the network of cumulus grains and the pressure on the intercumulus liquid. Mafic minerals and mixtures of mafic minerals and plagioclase form cumulates that have higher densities than does the magma from which they crystallize. Compaction can therefore take place in these cumulates. The density of plagioclase cumulates, on the other hand, differs little from the density of the magma from which they form and may even be lower. Compaction could not, therefore, be a factor in the formation of plagioclase cumulates, unless they were overlain by denser cumulates.

Where compaction is effective, porosity can be reduced to less than 1% by deformation and recrystallization of the cumulus grains (Sparks *et al*., 1985). This is because the

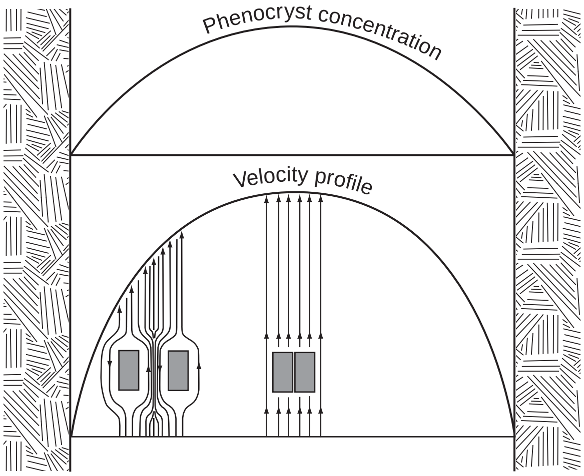


Fig. 14.25 Crystals near margins of dikes are forced apart by differential laminar flow rates of magma around the crystals. Toward the center of dikes, shear strain decreases and so does the force dispersing the grains. As a result, phenocrysts tend to concentrate toward the centers of dikes that are emplaced by laminar flow.

interstitial liquid wets the cumulus grains and produces a small dihedral angle (Section 12.7; Holness, 2006); the interstitial liquid therefore remains connected even when reduced to very small amounts. Compaction, however, does require a cumulate pile that is tens of meters thick to generate the necessary pressure differentials. Morse (1986) argues that such thicknesses are probably unusual. However, Shirley (1987) has shown that the chemical profiles through the Palisades Sill can be explained by compaction, and Gray *et al*. (2003) and Dickson (2006) have measured the textural anisotropy resulting from this compaction. In larger intrusive bodies, Meurer and Boudreau (1998) have shown that compaction can result in the development of planar fabrics, which can give a physical measure of the degree of compaction. Elongate or platy crystals such as feldspar can be rotated toward horizontal (Fig. 14.14(B)). However, the amount of compaction necessary to produce the foliation seen in many plagioclase cumulates requires that in addition to rotation there must be dissolution and reprecipitation of minerals (Higgins, 1991). Where cumulate grains contact one another, the increased pressure can cause dissolution of the grains and reprecipitation on the grains where they extend into low-pressure pore spaces. This is similar to one of the processes that cause foliation to develop in metamorphic rocks (see Section 17.2 and Fig. 16.3).

Concentrations of early crystallizing minerals can occur when phenocryst-laden magma moves. One way in which this takes place is by *filter pressing*. Any constriction to flow may cause phenocrysts to be left behind. Changes in the width of dikes feeding magma into the east rift zone of Kilauea, for example, might be expected to filter out phenocrysts of olivine. Another process that comes into play when magma is in laminar flow and near the wall of a conduit is known as *flowage differentiation* (Komar, 1972). The differential rate of flow of magma around phenocrysts at different distances from the wall of an intrusion produces a force that pushes the crystals apart (Fig. 14.25). Because the rate of change of velocity with distance increases toward the walls of a conduit (differentiate Eq. (3.8) or (3.15)), the *grain dispersive pressure* reaches a maximum at the walls and decreases to a minimum at the center of the conduit. As a result, the concentration of phenocrysts is least near the walls and greatest near the center of conduits. Because large phenocrysts cause a greater distortion of the flow lines than do small ones, the effect is greater for large crystals (Ross, 1986). This mechanism of producing rocks with concentrations of early crystallizing phases is most effective in narrow conduits or in boundary layers of convecting cells where velocity gradients are high.

All of these different processes of concentrating early crystallizing minerals may play roles in producing cumulates. However, the apparent indifference of the cumulate-forming process to density contrasts between cumulus grains and magma and to the dip of the surface on which deposition occurs suggests that gravity does not play an important role. Instead, nucleation and growth of crystals on the walls, roof, and floor of chambers, aided by dense plumes of crystal mush sinking to the floor, seem to be the dominant processes of formation, with diffusion and crystal aging (Ostwald ripening) being responsible for many of the small-scale compositional variation. Gravity does play a role in redistributing cumulates that slump from the walls or roof of a chamber onto the floor. With these controlling factors, it is not surprising that most cumulates form in large, slowly cooled plutonic bodies.

14.7 LIQUID IMMISCIBILITY

The unmixing of magmas into liquids of contrasting composition appealed to early petrologists as a simple mechanism by which magmas might differentiate. However, the field evidence advanced in support of this hypothesis, such as agate geodes in basalt, spherulites in rhyolite, and orbicular granites, which were all supposed to have formed as immiscible liquid globules, was easily shown by Bowen (1928) to be flawed. Furthermore, Bowen argued that the only liquid immiscibility field known at that time to exist in silicate systems (Fig. 10.9) was at too high a temperature to be of importance in the Earth's crust. This, combined with the success that fractional crystallization met with as a means of explaining magmatic differentiation, resulted in liquid immiscibility being almost completely ignored by petrologists during the first half of the twentieth century. The discovery of the low-temperature immiscibility field in the system fayalite–leucite–silica (Fig. 10.24) by Roedder (1951) rekindled interest in this phenomenon, and by the time samples were returned from the Moon, petrologists were again receptive to finding evidence of immiscibility, which was present in a large fraction of the lunar samples (Roedder, 1979).

The residual liquid formed by fractional crystallization of tholeiitic magma is commonly quenched to a glassy mesostasis between the plagioclase and pyroxene crystals of tholeiitic basaltic lavas. In the majority of these rocks this glass is now recognized to consist of two phases, one an iron-rich silicate glass of approximately pyroxene composition, and the other a silica-rich glass of granitic composition (Philpotts, 1982). These glasses, which are quenched immiscible liquids, form globules of one in the other, with diameters reaching 50 μm (Fig. 14.26). The abundance of this two-phase glass varies from a few percent in primitive mid-ocean ridge basalts to approximately 30% in iron-rich basalts.

The compositions of pairs of immiscible liquids in these tholeiitic rocks (Table 14.2) match closely the compositions of conjugate liquids in the system fayalite–leucite–silica (compare Figs. 10.24 and 14.26). The compositional gap is slightly wider in basalts than in the simple system, due to the presence of additional components in the basalts, such as TiO_2, P_2O_5, and Fe_2O_3, that are known to expand the immiscibility field.

Figure 14.26 shows the composition of immiscible liquids in a wide range of tholeiitic lavas; it also shows the bulk compositions of the immiscible liquid pairs and the whole-rock compositions. The dashed line from the whole-rock compositions to the bulk compositions of the immiscible liquid pairs are fractional crystallization trends due to the crystallization mostly of plagioclase and pyroxene. Crystallization of these minerals causes the residual liquid to enter the two-liquid field at a point that depends mainly on the oxidation state of the magma and the stage at which magnetite starts crystallizing (Philpotts and Doyle, 1983). If magnetite crystallizes late, the residual liquid remains iron-rich and the two-liquid field is reached rapidly, with approximately equal proportions of iron-rich and silica-rich

Table 14.2 *Average compositions of immiscible liquids preserved as glassy globules in tholeiitic and alkaline lavas*

	Tholeiitic		Alkaline	
	Fe-rich glass	Si-rich glass	Fe-rich glass	Si-rich glass
SiO_2	41.5	73.3	37.1	65.4
TiO_2	5.8	0.8	8.2	1.0
Al_2O_3	3.7	12.1	5.1	13.9
FeO^t	31.0	3.2	28.9	4.0
MnO	0.5	0.0	0.5	0.0
MgO	0.9	0.0	1.9	0.5
CaO	9.4	1.8	8.7	2.3
Na_2O	0.8	3.1	1.1	4.0
K_2O	0.7	3.3	1.6	4.9
P_2O_5	3.5	0.07	3.4	0.5
Total	97.8	97.67	97.0	96.5

After Philpotts (1982).

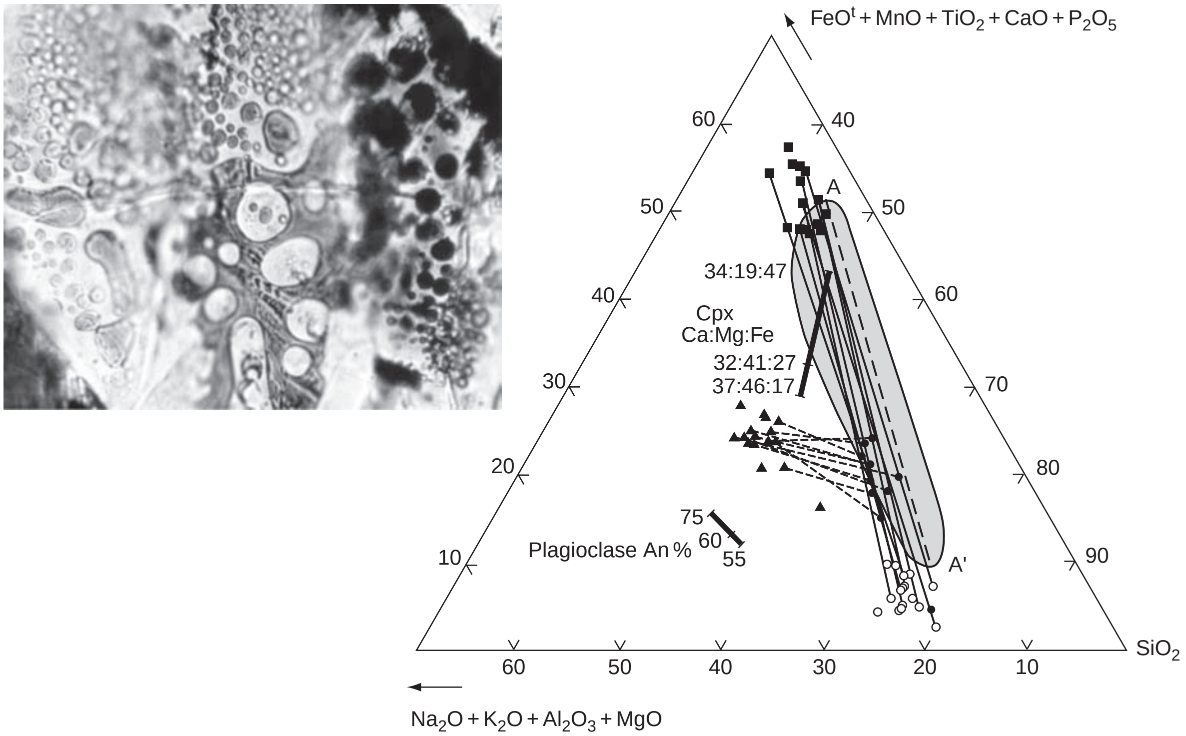


Fig. 14.26 Plot of the composition of immiscible iron-rich (filled squares) and silica-rich (open circles) immiscible liquids preserved as glassy globules in a range of tholeiitic volcanic rocks (triangles) (Philpotts, 1982). The shaded area is the two-liquid field from the system fayalite–leucite–silica and the heavy dashed line from A to A′ is the fayalite tridymite cotectic crossing the two-liquid field (Roedder, 1951). The compositional range of plagioclase and of augite in these rocks is shown, with a mark on each solid solution line indicating the composition at which the magmas are believed to have unmixed. The inset photomicrograph shows iron-rich (dark) and silica-rich (clear) glasses formed as immiscible residual liquids in a tholeiitic basalt from Kilauea, Hawaii. Plane light; width of field is 100 μm. (From Philpotts, 1982.)

immiscible liquids being formed. When magnetite crystallizes early, the residual liquid becomes enriched in silica and the two-liquid field is encountered at a later stage of crystallization, with a smaller fraction of iron-rich liquid being formed. If the oxidation state of the magma rises above the values associated with the nickel–nickel oxide buffer (Fig. 11.21), the fractionated liquid misses the immiscibility field completely.

Although liquid immiscibility produces melts of granitic composition (and of Fe-rich pyroxenite) in the mesostasis of tholeiitic basalts, is there evidence that immiscibility has played a role in the generation of significant amounts of differentiated tholeiitic rocks? It should be emphasized that the immiscibility field in tholeiitic basalts is encountered only in the fractionated liquid residue after at least 70% of the original magma has crystallized. Immiscibility, therefore, cannot be called upon to simply cause magmas to split into contrasting liquids upon cooling, as was done by early petrologists. Once a magma has solidified beyond 70%, immiscible liquids have little likelihood of being able to separate from one another to form sizable bodies of magmas, despite the significant density contrast between the two liquids. Perhaps if a crystal mush were to be compacted, some separation of the immiscible liquids might occur. Granophyric segregations towards the tops of bodies, such as the Palisades Sill, might form in this way.

If immiscibility is to produce sizable bodies of contrasting magmas, it appears necessary for magma first to crystallize fractionally, so that when unmixing does take place, the conjugate liquids are free to separate. These conditions exist during the final stages of crystallization of large layered gabbroic intrusions. Indeed, McBirney (1975) has shown through experiments that late stage liquids in the Skaergaard Intrusion entered the two-liquid field. He postulates that the low-density silica-rich liquid rose to the roof of the residual lens of magma to form granophyre, while the dense iron-rich liquid sank along with dense minerals to form an underlying zone of ferrodiorite. Jakobsen *et al*. (2005) have documented the presence of trapped immiscible droplets in melt inclusions in apatite and fayalite crystals from the upper part of the Skaergaard Intrusion. Although this separation occurs late in the crystallization of tholeiitic magma, in a fractionated intrusion such as the Skaergaard it can play a role in the differentiation of tholeiitic rocks, but only as an additional factor to the more important process of fractional crystallization.

Granitic magma can form from tholeiitic magma simply by fractional crystallization or by a combination of fractional crystallization and immiscibility. In Figure 14.26 these alternate paths involve either fractionating a magma directly from an initial tholeiitic composition (solid triangles) to the granitic residual liquid (open circles), or following one of the dashed lines into the two-liquid field and then producing the granite liquid through immiscibility. These granitic liquids have geochemical signatures that are diagnostic of the path followed.

Elements with large ionic radius and charge (incompatible elements) do not enter the early crystallizing minerals and, therefore, become concentrated in residual liquids (Section 14.12). Granitic magma formed through fractional crystallization should therefore be enriched in elements such as phosphorus and rare earths. Experiments, however, reveal that if a silica-rich liquid is in equilibrium with an immiscible iron-rich liquid, the incompatible elements partition strongly into the less polymerized iron-rich melt (Watson, 1976; Ryerson and Hess, 1978; Veksler *et al*., 2006). Granitic magma formed through immiscibility would therefore have low concentrations of these elements. Despite the simplicity of this distinction, its application is often difficult. A granite cannot simply be said to be rich or poor in incompatible elements; this concentration must be given relative to either the parental magma or, in the case of suspected immiscibility, with the rock type that may have formed from the conjugate liquid. In many occurrences, neither of these is known or their compositions are uncertain. Nonetheless, examples of comagmatic felsic and mafic rocks have been formed where this geochemical criterion of immiscibility has been satisfied (Vogel and Wilband, 1978).

Although the evidence for liquid immiscibility is ubiquitous in the mesostasis of tholeiitic basalts, it is rare in alkaline basalts. This is due, in part, to the tendency of the mesostasis in alkaline basalts to crystallize rather than to quench to the glass, which would be necessary to preserve immiscible globules. Also, magnetite commonly crystallizes earlier in alkaline rocks than it does in tholeiitic ones, and thus residual liquids are not as enriched in iron and are less likely to encounter the immiscibility field. On the other hand, most alkaline rocks contain higher concentrations of the elements that expand the immiscibility field (TiO_2, P_2O_5, Fe_2O_3) than do tholeiitic ones. A small number of immiscible alkaline glasses have been found (Philpotts, 1982). Their appearance is identical to those in tholeiitic rocks and their compositions are similar, except that both conjugate liquids contain less silica and more alkalis than do those in tholeiitic rocks (Table 14.2).

These few examples of alkaline immiscible liquids that are quenched to glasses are important because they provide a compositional link between the common immiscible glasses in tholeiites and the features in some lamprophyric dike and sill rocks known as ocelli (Philpotts, 1976). Ocelli are globular felsic bodies that occur most commonly in lamprophyres of camptonitic composition (Chapter 6). They have diameters of several millimeters (Fig. 14.28(D)), but they are capable of coalescing to form larger globules (Figs. 4.81 and 14.27). They have been interpreted either as gas cavities that were filled with late crystallizing liquid, or as globules of immiscible felsic liquid in mafic alkaline magma. The partitioning of incompatible elements between ocelli and host lamprophyre favors the immiscibility origin for these structures. Calculated viscosities and densities of the ocellar phase based on the shape and distribution of ocelli within sills also favors the immiscibility hypothesis.

Ocelli are capable of rising and coalescing to form sheets near the top of sills. When these sheets attain thicknesses of about 1 cm they become gravitationally unstable and may rise into the dense overlying crystal mush as diapirs (Fig. 3.16). Some ocelli become large enough that they form small magmatic bodies in which gravitative crystal settling takes place as they cool and crystallize (Fig. 14.27). On this small scale, liquid immiscibility appears to play the dominant role in fractionating the lamprophyric magma. In larger intrusions, immiscibility may still be effective, but with slower cooling, other processes may obscure its evidence.

Intrusive plutonic alkaline rocks commonly show bimodal compositions, being either mafic or felsic, with few intermediate rock types. Figure 14.28(B), for example, shows meter-size irregular masses of nepheline syenite in hornblendite from Bon Concielle, Quebec. These two contrasting rock types show no reaction, and large hornblende crystals can be seen to cross the boundary between them. This scarcity of intermediate rocks between the mafic and felsic members of the alkaline series is commonly referred to as the *Daly Gap*, after Reginald Daly, who first drew attention to it. Numerous explanations have been advanced to explain the gap. Differences in the physical properties of the various alkaline magmas (density, viscosity) might increase the chances of mafic and felsic magmas being intruded into the crust. On the other hand, nonideal behavior in the liquids could result in sharp changes in the slope of the liquidus; this, in turn, would cause the rate at which liquid changed its composition with falling temperature to change rapidly. If the system is sufficiently nonideal, liquid immiscibility could be responsible for the gap.

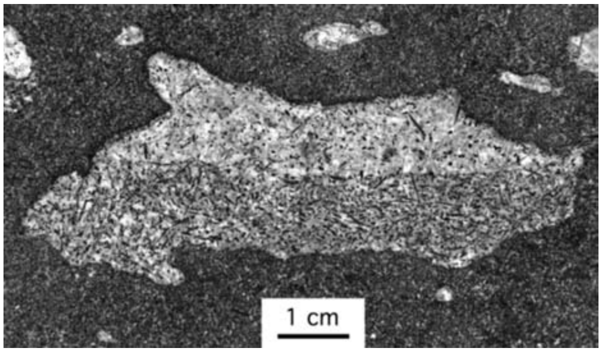


Fig. 14.27 Settling of hornblende and plagioclase toward the base of an ocellus in a lamprophyric sill produced a lower zone of nepheline monzonite and an upper zone of nepheline syenite. The ocellus, in turn, was probably formed by coalescence of immiscible globules of felsic melt that rose and accumulated toward the top of the sill; Ste. Dorothée, Quebec.

The Cretaceous Monteregian intrusions of Quebec (Fig. 4.74) provide examples of this mafic–felsic rock association (Fig. 14.28(A). Despite the scarcity of intermediate rock types in these intrusions, regional dikes, which presumably provide samples of the magmas involved in the formation of the main intrusions, exhibit no such gap (Fig. 14.28(C)). However, all dike rocks that have compositions corresponding to the gap in the plutonic rocks all contain ocelli (open circles in Fig. 14.28(C)). This is strong evidence that ocelli do indeed represent immiscible globules and that in the more slowly cooled plutons, the immiscible liquids had time to separate from each other. This conclusion has been supported by the incompatible element partitioning between these rocks (Eby, 1980).

One distinct group of alkaline rocks that may have been generated through liquid immiscibility is carbonatites. Such

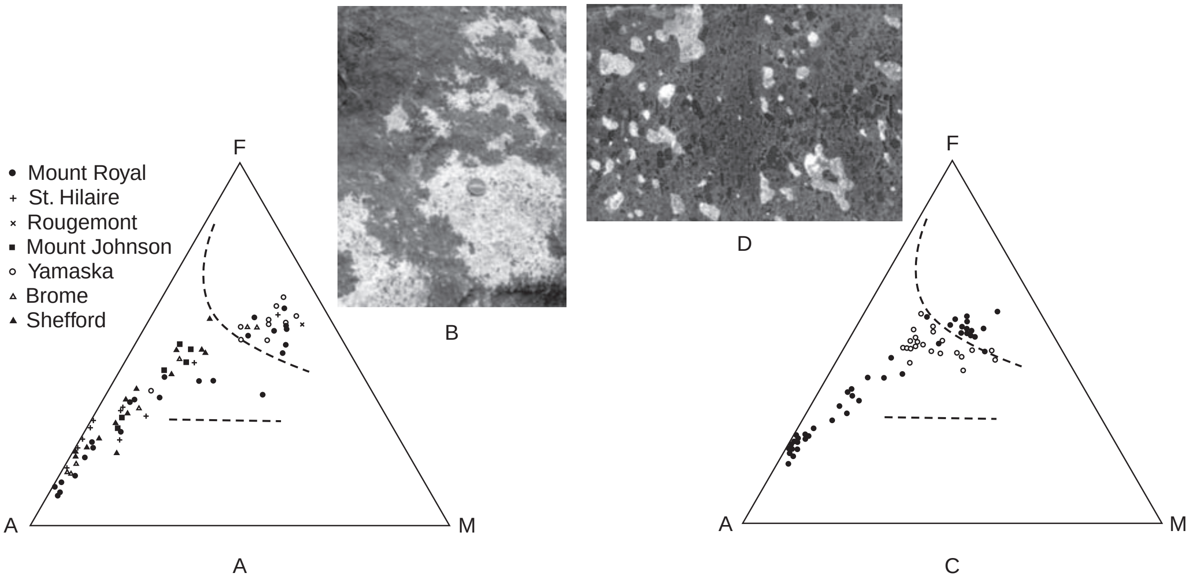


Fig. 14.28 Alkaline igneous rocks of the Monteregian Province, Quebec. (**A**) FMA plot of the plutonic rocks of the main intrusions. (**B**) Common association of mafic and felsic plutonic rocks, which in this example are hornblendite and nepheline syenite. (**C**) FMA plot of fine-grained dike rocks. Those containing ocelli are shown as open circles. The fact that the ocellar dikes have compositions that correspond to the gap in the plutonic rocks is evidence that the ocelli may have been immiscible liquid droplets that had time to separate in the plutonic bodies. Dashed line indicates probable extent of two-liquid field. (**D**) Typical ocellar camptonite dike rock containing ocelli of nepheline monzonite. Note that many ocelli contain an analcite-filled amygdale at their upper extremity, and some ocelli are deformed against an internal contact. Width of field ~15 cm. A photomicrograph of this rock is shown in Figure 12.29. (After Philpotts, 1976.)

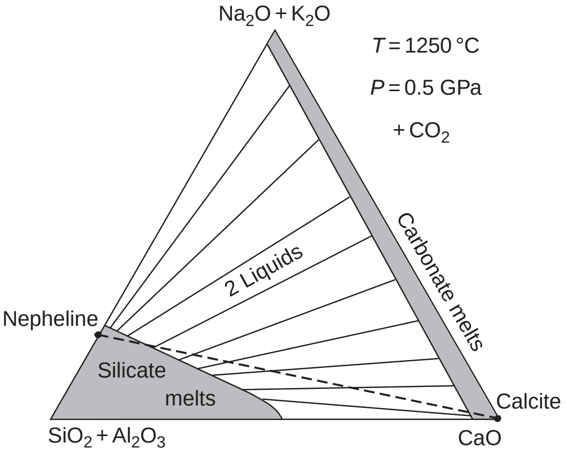


Fig. 14.29 Isothermal, isobaric ternary diagram showing extent of the two-liquid field between carbonatite and silicate liquids. (After Kjarsgaard and Hamilton, 1988.)

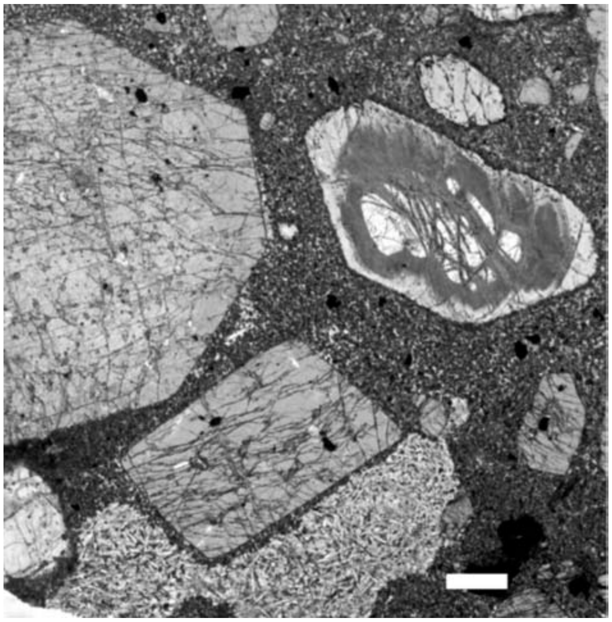


Fig. 14.30 Photomicrograph of ocellus wrapped around the lower side of an augite phenocryst in a lamprophyric sill, Mount Bruno, Quebec. The ocellus is interpreted to have been an immiscible globule of felsic liquid, which was able to deform around the sinking crystal; other ocelli were able to float and concentrate near the top of the sill. Note that the felsic material does not wet the pyroxene crystal but is separated from it by a thin sheet of the mafic matrix. Plane light; scale bar is 1 mm.

an origin has been widely accepted, in part, because of associated silicate rocks. Moreover, experiments have demonstrated a wide field of immiscibility (Fig. 14.29) separating the entire range of carbonatite compositions from silicate melts (Koster van Groos and Wyllie, 1966; Kjarsgaard and Hamilton, 1988).

It is surprising, in light of these experiments, that only a few rocks have been reported to contain immiscible droplets of carbonate liquid. Some lamprophyres contain carbonate ocelli, but these can also be interpreted as calcite amygdules. Dawson and Hawthorne (1973) interpret small diapiric structures on carbonatite sheets in kimberlite, which appear identical to those of felsic ocelli sheets in lamprophyre sills (Fig. 3.16), as evidence of liquid immiscibility. Rankin and Le Bas (1974) found immiscible carbonate droplets in silicate melt inclusions in apatite crystals from an ijolite pegmatite in the Usaki complex of West Kenya. They were able to homogenize the carbonate and silicate liquids by heating the apatite crystals; the immiscible droplets, then, reappeared on cooling.

Dasgupta *et al*. (2006) have shown that when carbonated eclogite melts at pressures of 3 GPa (~100 km depth) the first liquid formed is a carbonate melt at 1060 °C and this is joined at 1225 °C by a silica-poor silicate melt. Immiscible carbonate and silicate melts then coexist up to 1375 °C, above which the carbonate melt dissolves in the silicate melt. These experiments provide clear evidence that carbonate and silicate melts can coexist over a temperature interval of almost 200 °C at a depth where many alkaline magmas are formed. However, because the carbonate melt forms 165 °C below the first appearance of a silicate melt, a mobile carbonatite melt may have escaped the melting region before a silicate melt is formed.

The concentration of carbonate in most silicate melts is small. Crystallization of non-carbonate-bearing silicates, however, will enrich a magma in carbonate, and if the immiscibility field is encountered before solidification has proceeded too far, an immiscible carbonate liquid will be able to segregate. The low viscosity of carbonate melts will allow them to migrate rapidly and perhaps separate completely from their silicate parent.

Sulfide liquids also can form as an immiscible fraction in silicate magmas. As indicated in Section 11.4, the solubility of sulfides in silicate melts is extremely small. Thus only a slight degree of fractional crystallization of silicate minerals from a sulfide-bearing silicate melt is usually all that is necessary to produce immiscible sulfide droplets, which if free to coalesce, will sink and segregate on the floor of a magma chamber.

An important factor to keep in mind when considering how rapidly immiscible liquids of any kind might be able to segregate is that immiscible droplets are able to coalesce. If initially all immiscible droplets in a magma were to nucleate and grow to a similar size, and if they were buoyant they would, according to Stokes' law (Eq. (14.4)), rise at the same rate. However, should any two touch and coalesce, the larger droplet, so formed, would rise more rapidly, which, in turn, would cause it to catch up with smaller ones. Further coalescence would produce still larger globules that would rise still more rapidly, and so on. Furthermore, because immiscible droplets are deformable, they can travel through a crystal mush and bring about differentiation in much the same way that oil can separate and float on top of water in a porous sandstone. Figure 14.30 illustrates a felsic ocellus that was in the process of buoyantly rising around a sinking augite phenocryst in a 2-m-thick sill when solidification occurred. Other ocelli that managed to pass the sinking crystals accumulated near the top of the sill to form a continuous sheet of felsic rock. The ability of immiscible droplets to coalesce and deform may allow differentiation to take place under conditions that would not permit crystal fractionation to occur.

14.8 DIFFUSION PROCESSES: SORET EFFECT

We normally think of diffusion taking place in magmas in response to compositional gradients set up, for example, by fractional crystallization (Section 14.6) or mixing of magmas (Section 14.11). But diffusion can also be induced by thermal gradients, in which case it is known as thermal diffusion or the *Soret effect*, so named in honor of the chemist who first demonstrated the phenomenon. When a solution is placed in a thermal gradient, its components redistribute themselves by diffusing up or down the gradient in order to establish an equal distribution of internal energy. For example, in gases, where the effect can be pronounced, the translational kinetic energy of the molecules ($^1/_2 m\upsilon^2$) can be equalized by having heavier molecules move to the cooler end of the gradient (lower thermal velocity) and the lighter ones to the hotter end (higher velocity). In liquids and solids the situation is more complex because of the structure of the medium and the interactions between species. Qualitatively, however, the lighter components still migrate toward the hotter end, but the actual degree to which this occurs must be determined experimentally.

Walker and DeLong (1982) have investigated the Soret effect in a sample of mid-ocean ridge basalt placed in a thermal gradient of 330 °C cm^{-1} at 1.0 GPa. They found that Si, Al, Na, and K were enriched toward the hot end of the charge (1480 °C), and Fe, Mg, Ca, Mn, and Cr toward the cold end (1215 °C). The differentiation was great enough to produce liquids having compositions of quartz-normative andesite at the hot end and nepheline-normative picrite at the cold end.

The diffusing species in these experiments are not likely to be individual cations or oxides, otherwise the results would indicate that the mass of a species played no role in the differentiation. For example, K (atomic weight 39) and Na (23) are concentrated to a similar degree at the hot end, while Fe (56) and Mg (24) are concentrated to a similar degree at the cold end. The alkalis are probably coupled with alumina to form network-building units ($KAlO_4^{4-}$), which diffuse to the hot end along with polymerized SiO_4^{4-} groups. Walker and DeLong found that when there are insufficient alkalis to couple with alumina to form network groups, alumina diffuses to the cold end instead. The Fe, Mg, and Ca are probably coupled with silica to form chains and rings that have greater mass than the network units and so migrate to the cold end. The resulting differentiates have compositions similar to those formed through liquid immiscibility (Section 14.7). Thus a geochemical signature may serve only to distinguish fractional crystallization from liquid differentiation, not to characterize rocks as being uniquely the products of liquid immiscibility or the Soret effect.

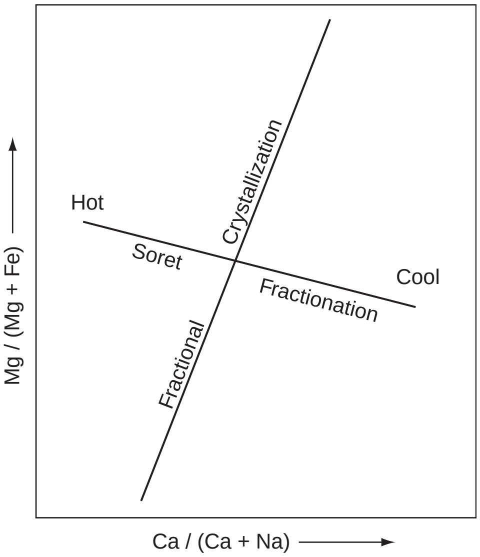


Fig. 14.31 Compositional trends of melts fractionated by crystallization and by Soret diffusion expressed in terms of Ca/Ca + Na and Mg/Mg + Fe. (After Walker and DeLong, 1982.)

The difference between Soret fractionation and crystal fractionation should be evident in the compositional variations of coexisting plagioclase and ferromagnesian minerals. With Soret diffusion, iron and magnesium both concentrate towards the cold end of the gradient, but the effect is greatest for the heavier iron. As a result, the Mg/Mg + Fe ratio is greatest at the hot end. But this is the end with the lowest Ca/Ca + Na ratio. Thus, when Soret-differentiated liquids crystallize, the resulting compositional variation in plagioclase and ferromagnesian minerals is very different from the trend formed by fractional crystallization, where Mg/Mg + Fe and Ca/Ca + Na vary sympathetically (Fig. 14.31). If fractional crystallization and Soret diffusion both affect a liquid, intermediate trends would result.

Differentiation of a magma by the Soret effect requires that a temperature gradient be maintained long enough for components to diffuse the distance over which the gradient exists. Values of thermal diffusivity in magmas are on the order of 10^{-6} m^2 s^{-1} whereas diffusion coefficients in basaltic magma are near 10^{-11} m^2 s^{-1}, and in granitic magma they are still less (Fig. 5.13). Therefore, because heat diffuses much faster than do chemical species, a thermal gradient would be dissipated long before the Soret effect could cause significant differentiation, unless some process, such as convection, were to replenish the heat and maintain the thermal gradient. If this occurs, the length of time necessary to establish a Soret gradient is given approximately by

$$\text{time}_{\text{Soret}} = \frac{d^2}{4D} \tag{14.18}$$

where d is the distance over which species must diffuse, and D is the diffusion coefficient for the mobile species. This equation is similar to Eq. (14.16) for the length of time for a nonconvecting body of magma to cool, the difference being that the diffusion coefficient is substituted for the thermal diffusivity. In Walker and DeLong's experiments, the gradient was established over a distance of 8 mm, and the diffusion coefficient of, for example, K in basaltic magma at 1300 °C is 10^{-10} m^2 s^{-1}. Potassium in their experiments would therefore have required 44 hours to establish an equilibrium gradient. The fact that they ran their experiments for 137 hours assured that there was sufficient time for the gradient to develop.

The steep thermal gradient necessary to drive Soret diffusion can be maintained at the margin of an intrusion as long as the magma convects. However, convection dramatically shortens cooling times (Eq. (14.15)), so that the Soret effect will be most pronounced in those bodies where heat is periodically replenished by the intrusion of batches of new magma from the mantle, as is thought to happen in many large granitic bodies and in magma chambers beneath mid-ocean ridges. The fraction of magma in a chamber that occupies the thermal boundary layer at any one time is very small, but convection continually renews this magma. Spera and others (1982) have calculated that the entire volume of magma in a large intrusion (6000 km^3) can be cycled through a thin (20-m) boundary layer as many as a thousand times during the life-span (10^6 years) of such a chamber. Thus, although Soret diffusion operates over only short distances, convection may allow it to affect large volumes of magma.

As long as the Soret effect produces magma near the wall of a chamber that has a density which causes it either to rise to the top or sink to the bottom of the chamber (Fig. 14.10(A) and (B), respectively), convection will not rehomogenize this magma with the main body of magma. Large volumes of differentiated magma will then be able to accumulate slowly from the thin boundary layer. Hildreth (1981), for example, believes that Soret diffusion near the walls of chambers of convecting bodies of granitic magma is responsible for developing water-rich melt that convects upward and accumulates near the roof. In addition to water, many other components are concentrated in this buoyant boundary layer. Their accumulation results in a zoned body of magma, which, if erupted, can produce ash flows that preserve, in their stratigraphy, the record of the emptying of the chamber from the roof downward. The compositional variation in these rocks cannot be accounted for simply by fractional crystallization, which is evidence that some other process, such as Soret diffusion, must be responsible for their differentiation.

Latypov (2003a, b) has invoked Soret diffusion to explain compositional gradients near the lower and upper margins of mafic sills and to explain S- and D-shaped compositional profiles through many sills. The steep temperature gradient near the contacts is supposed to produce a compositional profile with extremely evolved compositions near the contact grading inward to more primitive compositions. He invokes vigorous convection with meltback of chilled margins to maintain steep temperature gradients at the contact. Despite the inventiveness of this approach, it does not appear that the conditions in the margins of sills would be conducive to Soret diffusion (Lesher and Walker, 1991). Many of the profiles referred to by Latypov have simpler explanations involving early injection of relatively phenocryst free magma, multiple intrusions, crystal settling, and compaction. The chemical profiles through the Shiant Isles Main Sill on the northwest coast of Scotland is cited by Latypov as a classic product of Soret diffusion, but Gibb and Henderson (2006) have shown that this sill lacks any of the features that would result from Soret diffusion.

14.9 PNEUMATOLITIC ACTION

When a volatile-rich magma rises toward Earth's surface, the decreasing pressure may allow a separate fluid phase to form, which, being lighter, rises more rapidly than the magma. The streaming of the volatile-rich phase through the magma depletes the lower part of the body in those elements that preferentially enter the fluid and transports them to the upper part of the chamber, where they may accumulate beneath an impervious roof or escape to form pegmatites and hydrothermal solutions. A fluid phase can also separate from magma as a result of the crystallization of anhydrous minerals, which enriches the melt in water and other volatiles. This type of separation is known as *resurgent boiling*. In thick cumulate piles, crystallization near the base can release volatiles that will migrate upward through the pile transporting with them incompatible elements and even causing remelting of the crystal mush (Boudreau, 1999).

The fluid phase is rich in the volatile magmatic constituents, such as H_2O, CO_2, S, F, Cl, B, and P, but also contains less volatile elements that do not readily enter common rock-forming minerals but are not present in high enough concentrations to form their own minerals. These include Li, Be, Sc, Ga, Ge, As, Rb, Y, Zr, Nb, Mo, Ag, In, Sn, Sb, Cs, Ba, rare earth elements (REE), Hf, Ta, W, Re, Au, Pb, Bi, Th, U, and platinum group elements (PGE). Some of these elements on rising to the cooler upper parts of a magma chamber, or on entering fractures in the roof rocks, precipitate from the fluid to form the common minerals of pegmatites. Thus, along with quartz, feldspar, and mica, minerals such as lepidolite (Li), spodumene (Li), beryl (Be), zircon (Zr), apatite (REE), scheelite (W), and uraninite (U, Th) form. Many of the other elements remain in the fluid and precipitate at much lower temperatures from hydrothermal solutions. Of these, the group that is characteristically associated with sulfides – the *chalcophile* elements (Cu, Zn, As, Mo, Ag, Au, and Bi) – is of great economic importance.

The expulsion of fluids from compacting piles of crystal mush can lead to the formation of ore deposits. In the Stillwater Complex, Montana, Boudreau (1999) believes these fluids rose toward the top of the cumulate pile where they caused partial melting of the cumulates. Where the flow was channelized, potholes were developed in the cumulates. The fluids were enriched in Cl, and consequently were able to keep sulfur and PGE in solution. However, when the fluids reached the top of the cumulate pile and mixed with the main body of magma, which had a low Cl content, the sulfur and

PGE separated and formed an immiscible sulfide liquid, which formed the J–M (Johns–Manville) reef, an economically important deposit of palladium, platinum, and minor rhodium. Similar processes were active in the Bushveld Complex of South Africa, which led to formation of the Merensky reef, the world's richest platinum deposit. Boudreau (2004) has created a computer program to model the transport and deposition of PGE through crystallizing piles of crystal mush. This program, named Palladium, can be downloaded from his web site at www.nicholas.duke.edu/people/faculty/boudreau/DownLoads.html.

In addition to transporting incompatible elements, an immiscible fluid phase can play a role in differentiation by affecting other processes, such as convection or crystal settling. The presence of gas bubbles in a boundary layer of hydrous siliceous magma would increase the buoyant force and augment the segregation of a low-density melt that was being formed by fractional crystallization or Soret diffusion. On the other hand, gas bubbles rising through the Kilauea Iki lava lake significantly decreased the distance olivine crystals were able to sink and decreased the vigor of thermal convection.

Magmatic differentiation by volatile transfer may play a major role in determining the composition of carbonatites. Inspection of Figure 14.29 reveals that immiscible carbonate melts that coexist with most silicate liquids should contain significant amounts of alkalis. Indeed, immiscible phonolitic liquid (nepheline in diagram) can coexist only with a sodium carbonate melt. Although natrocarbonatites do occur, the majority of carbonatites are predominantly calcite-bearing rocks. Most calcite carbonatite intrusions, however, are surrounded by zones of intense alteration in which rocks are replaced by aegerine and orthoclase. This alkali metasomatism is known as *fenitization*, named after the locality where it was first identified – the Fen district of southeastern Norway. Fenitization is clear evidence that carbonatite melts give off considerable quantities of alkalis during crystallization. It seems likely, therefore, that volatile extraction is responsible for the low alkali contents of most carbonatites.

14.10 MAGMATIC ASSIMILATION AND ASSIMILATION AND FRACTIONAL CRYSTALLIZATION

The composition of a magma rising from its source can be modified by interactions with material encountered en route to the final crystallization site. Contamination can result, for example, from the incorporation and assimilation (i.e. melting) of xenoliths that come from the walls of the magma conduit or founder from the roof of a magma chamber. On the other hand, magma can encounter and mix with other bodies of magma, which may come from the same source at an earlier time or from completely different sources. In this section, we deal with the assimilation of solid rocks by magma and leave to Section 14.11 the problem of magma mixing.

Evidence of contamination of magma through the assimilation of country rocks can be found at the margins of many intrusions. It is particularly striking when minerals are produced that could not possibly form if there had been simply fractional crystallization. Quartz, for example, can be found near the margins of nepheline-bearing intrusions that were contaminated by quartz-bearing sedimentary rocks. Conversely, nepheline can be found at the margins of some tholeiitic bodies where they intrude limestone. Despite this clear evidence of contamination, assimilation of crustal rocks cannot produce large quantities of contaminated magma because of the large heat of fusion necessary to melt country rocks and the limited amount of heat available in the magma. Indeed, for magma to assimilate any significant quantity of material it must crystallize a roughly equivalent amount of material so that the heat liberated by crystallization balances the heat needed for fusion. Assimilation and fractional crystallization (AFC) consequently commonly occur together. Almost all mantle-derived magmas that crystallize in the crust show some effects of AFC, as can be documented from isotopic studies.

The lowest-melting fraction obtainable from sedimentary rocks is generally of granitic composition. It is not surprising, therefore, to find that where basaltic magmas have assimilated sedimentary rocks, contaminated magmas are intermediate in composition between the initial magma and a granitic end component. This indicates that melting of xenoliths is not wholesale but is only of that fraction which requires the least amount of heat to fuse. Partly digested xenoliths of more refractory sedimentary rocks are consequently common in such contaminated magma. Most of the Cretaceous Monteregian intrusions in Quebec, for example, are surrounded by such zones of contaminated magma (Philpotts, 1974). The early mafic phase in these bodies is nepheline normative, but where the country rocks are siltstones and shales, a zone of *rheomorphic breccia* is present that grades from a granitic composition in contact with the country rock to a progressively more silica-undersaturated rock toward the interior of the intrusions. Abundant refractory xenoliths of quartz-rich and alumina-rich sedimentary rocks are present throughout this zone of contamination (Fig. 4.76).

Where intruded sedimentary rocks contain abundant shale, the granitic fraction causing contamination is peraluminous. Such a fraction may alter the abundance of the early crystallizing minerals. Augite, for example, may react with the silica and alumina derived from the sedimentary rocks to produce anorthite and orthopyroxene according to the reaction

$$\begin{array}{ccc} \mathrm{Ca(MgFe)Si_2O_6} & + & \mathrm{Al_2O_3 + SiO_2} \\ \text{(potential augite)} & & \text{(contaminant in magma)} \\ \multicolumn{3}{l}{\rightarrow \mathrm{CaAl_2Si_2O_8} + \mathrm{(MgFe)SiO_3}} \\ \multicolumn{3}{l}{\quad \text{(anorthite)} \qquad \text{(opx)}} \end{array}$$

Thus, a magma that might initially have crystallized to gabbro, on being contaminated in this way, would form norite containing anorthite-rich plagioclase. The Haddo House Complex of Aberdeenshire, Scotland, provides a classic example of this type of contamination, where a border zone of norite contains abundant xenoliths of aluminous metamorphic rocks in all stages of digestion (Read, 1923). Longhi has argued that the early crystallization of orthopyroxene in most large layered gabbroic intrusions, such as the Stillwater and Bushveld complexes, is evidence that these mantle-derived magmas have assimilated continental crustal material. He points out that similar magmas in oceanic regions rarely crystallize orthopyroxene, and when they do, it is always preceded by calcium-rich pyroxene (Campbell, 1985). If the sedimentary rocks being assimilated are rich in graphite or other carbonaceous organic matter, the graphite can impart very low oxygen fugacities to the crystallizing magma. Granitic rocks that intruded a regionally extensive, 150–200-km-long belt of metasediments in part of the Sierra Nevada batholith of California are an example (Ague and Brimhall, 1988).

The effect of carbonate sedimentary rocks on silicate magmas is dramatic because of the reactive nature of calcite and dolomite with silica-rich melts. This can be illustrated with the reaction

$$\begin{array}{lll} \mathrm{CaMg(CO_3)_2} & + & \mathrm{MaAlSi_3O_8} \\ \text{(dolomite)} & & \text{(albite in magma)} \end{array}$$
$$\begin{array}{lllll} \rightarrow\ \mathrm{CaMgSi_2O_6} & + & \mathrm{NaAlSiO_4} & + & \mathrm{2CO_2} \\ \quad\text{(diopside)} & & \text{(nepheline)} & & \text{(gas)} \end{array}$$

In all such reactions, CO_2 is liberated, and CaO and MgO from the carbonate react with silica in the magma to produce calcium magnesium silicates. This, in turn, reduces the amount of silica available to form silica-saturated minerals, such as the albite in the reaction above. These reactions are therefore commonly referred to as *desilication* reactions. Daly (1933) proposed that all alkaline igneous rocks were formed by the desilication of subalkaline magmas through limestone assimilation. Although this can occur on a small scale at the contact of intrusions (Tilley, 1952), alkaline rocks are now known to have a primary source in the mantle.

The amount of magmatic assimilation that can take place depends on the thermal energy of the magma. Most magmas are not superheated; that is, their temperatures lie somewhere between the liquidus and solidus, and thus they would have a porphyritic texture if quenched rapidly enough. With such magmas, the heat needed to raise the temperature of the country rock to its melting point and then to supply the latent heat of fusion can come only from the crystallization of primary minerals. Thus, the amount of assimilation that takes place is normally offset by an almost equivalent amount of crystallization. Large quantities of magma containing a significant fraction of assimilated material are therefore not likely to form.

Some magma may be superheated and therefore capable of assimilating more material. Because the liquidus of dry magma increases with increasing pressure, the decompression associated with the rapid rise of magma causes resorption of crystals and may generate magma that is well above its liquidus. In addition, because Earth was hotter early in its history, mantle-derived magmas were hotter in the Archean than they are today and were therefore capable of assimilating larger amounts of crustal rocks (Sparks, 1986). The Archean ultramafic lavas known as komatiites, for example, were capable of assimilating three times as much crustal material as modern basalts.

Assimilation is greatest when the intruded rocks are near their melting point, and they have a large proportion of a low-melting fraction. The heat needed to melt this low-melting fraction of country rock is $\Delta H_m \times M_a$, where ΔH_m is the latent heat of fusion and M_a is the mass of assimilated country rock. This amount of heat withdrawn from the magma can be expressed as $C_p M_m \Delta T$, where C_p is the heat capacity of the magma, M_m the mass of magma affected, and ΔT the change in temperature of the magma. Equating these two and rearranging gives

$$\frac{M_m}{M_a} = \frac{\Delta H_m}{C_p \Delta T} \tag{14.19}$$

With typical values for ΔH_m and C_p of 400 kJ kg^{-1} and 0.8 kJ kg^{-1} °C^{-1}, respectively, the ratio of M_m/M_a is $500/\Delta\mathrm{T}$. Thus if 10% country rock were to be assimilated, the temperature of the magma would be lowered by 50 °C. If the initial magma were not superheated by 50 °C, the lowering of temperature would cause crystallization to occur, which in turn would liberate heat and maintain the magma at approximately the same temperature, provided that the heats of fusion and crystallization were similar (Problem 14.17). In magma chambers, assimilation leads to fractional crystallization (see discussion of AFC below).

The way in which magma rises through the crust may also play an important role in determining the amount of assimilation that occurs. If magma emplacement is through dikes or diapirs that force aside the country rocks, assimilation will normally be minimal. If the flow of magma in a dike is turbulent, contact temperatures are higher than when the flow is laminar, and thus assimilation is more likely. Turbulence occurs when the critical Reynolds number exceeds 2300 (Eq. (3.17)). This is possible only for magmas with low viscosities, such as those of basaltic and komatiitic composition, and then only in dikes that are more than a few meters wide (see Fig. 3.10). Many mafic dikes, however, satisfy these conditions, and zones of granophyre between quartz and feldspar grains in the wall rocks attest to the partial melting caused by the intrusion of such dikes (Fig. 3.2). The more primitive a magma is, the hotter it is, and the lower is its viscosity. Consequently, primitive magmas are more likely to assimilate crustal rocks than more evolved, cooler magmas. Huppert and Sparks (1985) estimate that komatiite magmas can be contaminated with up to 30% crustal material.

If magma rises through the crust by stoping or melting, assimilation of its roof rocks may be extensive. Near the surface of the Earth, where the roof of a magma chamber may be highly fractured, stoping of xenoliths is an efficient means of magma emplacement. If the xenoliths are denser than the magma, which is likely for all but the most mafic magmas, they will sink and possibly be assimilated, depending on their size. Deeper in the Earth, where rocks behave

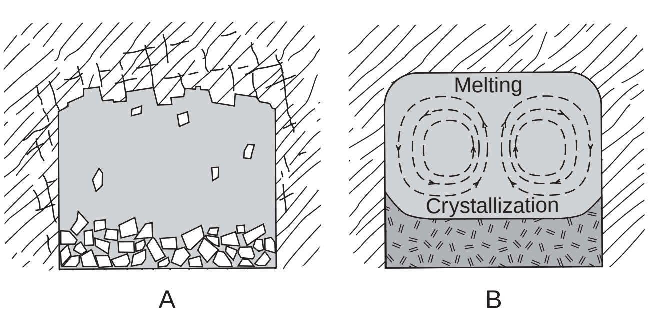


Fig. 14.32 Upward migration of magma body through crust. (**A**) At shallow depths, fractured crust provides xenoliths, which may sink through the magma, displacing it upward. Partial melting of xenoliths is accompanied by crystallization of magma. (**B**) At greater depths, where the crust is ductile, bodies of convecting magma can ascend by partly melting roof rocks while crystallizing refractory minerals on the floor. (**C**) Crust through which a convecting body of magma has passed becomes stratified by this "zone refining" process, with the most refractory minerals concentrated toward the base and the least refractory ones toward the top.

plastically, xenoliths are less likely to form. There the roof rocks may be melted *in situ*, as hot convecting magma continually removes the partially melted country rocks. These two styles of emplacement are illustrated in Figure 14.32.

Whether the magma melts the roof directly or melts sinking xenoliths, the necessary heat must come from the crystallization of the magma if there is no superheat. Convection within the magma keeps the temperature at the top and bottom of the chamber almost the same (Eq. (14.14)), so crystallization takes place on the floor of the chamber, where the pressure is greatest. Simultaneous melting of roof rocks and crystallization on the floor means that the volume of magma could remain constant if there were no loss of heat to the surroundings. The result would be a volume of magma that rises through the crust, continuously changing its composition as it goes. At first the magma might be entirely mantle-derived, but as it ascends it would steadily be contaminated by crustal material. Because the passage of this zone of melt through the crust is similar to the metallurgical process of zone refining, it is commonly referred to as *zone melting*. The distance magma can rise by assimilating and crystallizing material is of course limited by the amount of heat lost to the surroundings. Oxburgh and McRae (1984) estimate that a mantle-derived magma, on being emplaced in the crust, may have enough energy to rise by zone melting a distance equivalent to the height of the magma chamber.

For magmas that are not superheated, assimilation causes crystallization of minerals with which the magma is already saturated. These minerals would crystallize in the normal course of events, even if there were no assimilation; assimilation simply causes them to crystallize earlier. Furthermore, the composition of the low-melting contaminating fraction may be similar to that of the final fractionated liquid from the original magma. Thus, assimilation may simply bring on crystallization earlier and increase the amount of late-stage differentiate. Because the precise quantities of the various rock types in an igneous intrusion are rarely known, the effects of assimilation may not be readily detectable from major and trace element chemistries. It is here that isotopic studies are particularly useful (DePaolo, 1981b).

As shown in Section 13.4, the isotopic composition of the Earth's mantle and crust have steadily evolved along different paths throughout much of geologic time. As a result, magmas derived from the mantle have isotopic compositions that are quite distinct from those derived from the crust. Thus, the amount of crustal contamination in a mantle-derived magma can be gauged from the degree to which the mantle values are disturbed. Of particular use in this regard are the isotopic ratios $^{87}\mathrm{Sr}/^{86}\mathrm{Sr}$ and $^{143}\mathrm{Nd}/^{144}\mathrm{Nd}$. In both pairs the isotope in the numerator is the product of radioactive decay, $^{87}\mathrm{Sr}$ coming from $^{87}\mathrm{Rb}$ and $^{143}\mathrm{Nd}$ from $^{147}\mathrm{Sm}$. Because Rb is strongly fractionated into the Earth's crust, the $^{87}\mathrm{Sr}/^{86}\mathrm{Sr}$ ratio in crustal rocks is higher than that in mantle-derived rocks. Conversely, Sm is not as strongly fractionated into the crust as Nd, and therefore the $^{143}\mathrm{Nd}/^{144}\mathrm{Nd}$ ratio in crustal rocks is lower than in mantle ones.

DePaolo (1981b) showed that when a magma undergoes assimilation at the same time as it is fractionally crystallizing (AFC), the change in an isotopic ratio in the magma ($d\varepsilon_\mathrm{m}$), for example $^{87}\mathrm{Sr}/^{86}\mathrm{Sr}$ or $^{143}\mathrm{Nd}/^{144}\mathrm{Nd}$, with change in the logarithm of the fraction of magma remaining ($d\ln F$) can be expressed as follows:

$$\frac{d\varepsilon_\mathrm{m}}{d\ln F} = (a + r - 1)^{-1}\left[a\frac{C_\mathrm{a}}{C_\mathrm{m}}(\varepsilon_\mathrm{c} - \varepsilon_\mathrm{m}) + r\frac{C_\mathrm{r}}{C_\mathrm{m}}(\varepsilon_\mathrm{r} - \varepsilon_\mathrm{m})\right] \tag{14.20}$$

where a and r are the rates of change of magma mass due respectively to assimilation ($\dot{M}_\mathrm{a}$) and to replenishment with fresh primitive magma ($\dot{M}_\mathrm{r}$) relative to the rate of change of magma mass due to crystallization ($\dot{M}_\mathrm{c}$); that is, $a = \dot{M}_\mathrm{a}/\dot{M}_\mathrm{c}$, and $r = \dot{M}_\mathrm{r}/\dot{M}_\mathrm{c}$; C is the concentration of an element in the assimilated material, the magma, or the replenishing magma as indicated by the subscripts a, m, and r,

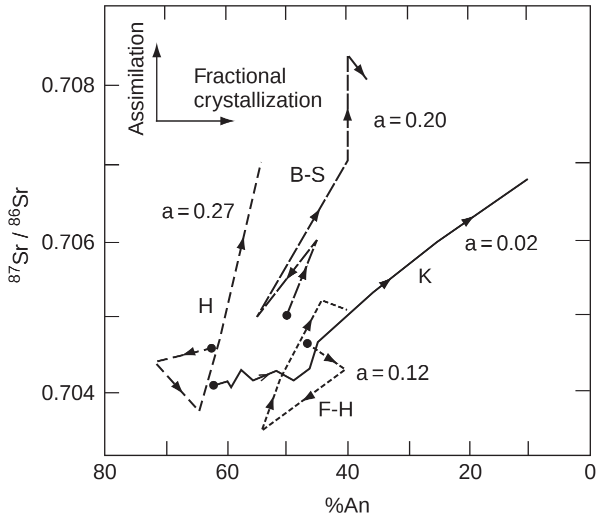


Fig. 14.33 Variation of initial ^{87}Sr/^{86}Sr ratio with anorthite content of plagioclase in rocks of the Kiglapait layered intrusion, Labrador (K) (after DePaolo, 1985), the Fongen–Hyllingen Intrusion (F–H), the Bjerkreim–Sokndal Intrusion (B-S), and Hasvik Intrusion (H), Norway (after Tegner *et al.*, 2005). Arrows on trend lines indicate up in the layered intrusion. Assimilation of crustal rocks causes the isotopic ratio to increase as plagioclase crystallizes and becomes less anorthitic, with the steeper plots indicating higher degrees of assimilation. The values of *a* are the ratios of the rate of assimilation to rate of crystallization (see text for discussion).

respectively, and ε_c , ε_m, and ε_r are the isotopic ratios in the crystals formed, the magma, and the replenishing magma.

Application of this equation is difficult because we commonly do not know exactly how much fractional crystallization has taken place; that is, F is unknown. However, in layered mafic intrusions, the layers and their stratigraphic height in the intrusion provide a good measure of F, and therefore it is possible to determine how much assimilation and fractional crystallization has taken place.

Figure 14.33 shows plots of the variation in the initial ^{87}Sr/^{86}Sr for a number of layered intrusions in terms of the anorthite content of the plagioclase, which decreases as the magmas fractionate. If there were no assimilation, the isotopic ratio would not change, and compositions would trend horizontally to the right as fractional crystallization caused the plagioclase to become more albitic. Assimilation of crustal rocks, however, causes the isotopic ratio to increase as the magmas fractionally crystallize, creating positive slopes to the trends. The greater the amount of assimilation, the steeper will be the slope.

The magma in each intrusion starts with a relatively low isotopic ratio, which is consistent with derivation from the mantle. The trends all then show minor isotopic fluctuations with changes in plagioclase composition. This part of the trends is formed during filling of the magma chambers where crystallization, assimilation, and recharge cause magma compositions to change frequently. Once the filling stage ends, the magmas then fractionally crystallize and assimilate wall rocks to produce smooth trends with positive slopes. The trend for the Kiglapait Intrusion in Labrador (DePaolo, 1985) has the shallowest slope and consequently indicates the least contamination. The three Norwegian intrusions (Tegner *et al.*, 2005), Fongen-Hyllingen, Bjerkreim-Sokndal, and Hasvik, show progressively steeper slopes, with the Hasvik Intrusion being one of the most contaminated layered intrusions known. When the isotopic ratios in these intrusions are related to the fraction of magma remaining (F), the value of a in Eq. (14.20) can be solved for. These values are shown in Figure 14.33. In the Kiglapait, $a \approx 0.02$, which indicates that the rate of crystallization was ~50 times faster than the assimilation rate. In the Hasvik Intrusion, however, $a \approx 0.27$, so the rate of crystallization was only 3.7 times the rate of assimilation. The amount of assimilation that takes place in an intrusion depends on the initial magma's temperature and how far it is above the solidus (excess heat) and on the composition and temperature of the country rocks.

Isotopic ratios can give misleading impressions of the amount of assimilation that has taken place if used carelessly. Continental flood basalts the world over vary only slightly in composition, but their initial ^{87}Sr/^{86}Sr ratios vary widely. If this variability is due to different degrees of crustal assimilation, bulk compositions might be expected to vary much more than they do. The amount of contamination in these rocks must therefore be small, but it must involve a fraction that contains a disproportionately high concentration of strontium isotopes. Of course, the ^{87}Sr/^{86}Sr ratio of the assimilated rocks also plays an important role.

14.11 MIXING OF MAGMAS

Evidence for commingling of magmas is now recognized to be common. The magmas may be successive pulses from a common source, in which case their compositions may be similar. On the other hand, they may have completely different sources and be related in time only through a common thermal event; such magmas may have very different compositions. As we have seen in Figure 14.33, most large magma bodies are not formed by a single injection of magma but are the result of numerous pulses, each of which may have slightly different compositions. In addition, magma in a chamber may cool, fractionally crystallize, or assimilate country rocks before the next pulse arrives. Successive batches may therefore encounter resident magmas that have had their compositions and physical properties modified by processes taking place in the chamber. Mantle-derived magmas bring heat into the crust, where partial melting can generate bodies of silicic magma. It is not unusual, therefore, to find mafic and silicic magmas commingling, as is clearly evidenced by the intimate mixtures of siliceous pumice and basaltic glass in some ash flows (Fig. 14.34), and the presence of pillow-like bodies of basaltic rock in some granites (Fig. 4.79(A)).

The commingling of magmas may result in a range of products. Magmas may, for instance, not mix at all but simply remain as discrete bodies or xenoliths of one in the other. At the other extreme, they may mix completely to form a

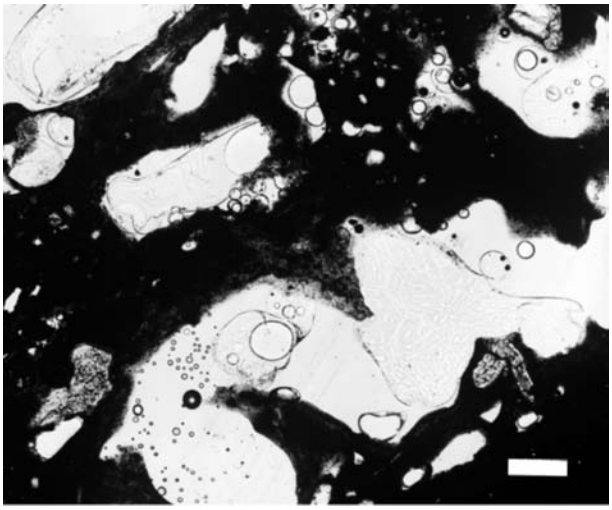


Fig. 14.34 Photomicrograph of intimately commingled basaltic and rhyolitic glasses from the Newberry volcano, central Oregon. Plane light; scale bar is 1 mm.

homogeneous hybrid magma with the only evidence of mixing being the presence of crystals that may not be in equilibrium with the melt (Martel *et al*., 2006). Whether magmas mix or not depends on how different their physical properties are. Density differences, for example, determine whether a second magma rises through or ponds at the bottom of an earlier magma. Viscosity contrasts determine how readily turbulent plumes of magma can bring about mixing. And the temperatures and volumes of the magmas and their solidification temperatures set further limits on the degrees of mixing. Studies by Huppert and Sparks (1980, 1988), Campbell and Turner (1986), and Sparks and Marshall (1986) are among a number of papers that quantify the effects these various factors have on magma mixing.

Two possible situations can result when one magma is intruded into another magma of different density. If the new magma enters at the base of a chamber in which the resident magma is denser (possibly from cooling, for example), the new magma buoyantly rises in a plume to the top of the chamber where it spreads laterally beneath the roof (Fig. 14.35(C) and (D)). If the new magma is denser than the resident magma, it may rise as a fountain into the chamber if intruded rapidly enough, but it then sinks and accumulates on the floor of the chamber (Fig. 14.35(A) and (B)). Snyder and Tait (1995) have shown through experiments that as a dense magma flows across the floor of a chamber, it develops a flow-front instability that takes the form of fingers extending in the direction of flow. This is analogous to the flow-front instability discussed in connection with laccoliths (Fig. 4.63). They believe that these fingers may bud and give rise to the pillow-like bodies of basalt found near the base of some granite bodies (Fig 14.38).

If the rate of intrusion of new magma into a resident magma is great enough, turbulence within the plume or fountain causes entrainment of the resident magma and some intimately commingled or mixed magma is generated. The bulk density of this hybrid magma is intermediate between the other two magmas, and it therefore accumulates between them (Fig. 14.35(B) and (D)).

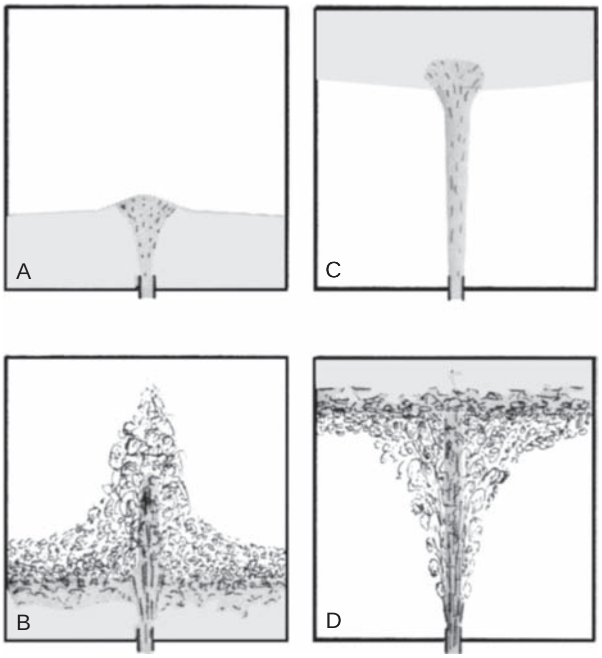


Fig. 14.35 Four possible modes of replenishment of a magma chamber. (**A**) If the new magma has a greater density than the resident one, and intrusion velocities are low (Reynolds number < 7), the new magma ponds at the bottom of the chamber and there is no mixing. (**B**) If the new magma is denser but is intruded rapidly (Re > 70), it still ponds at the bottom of the chamber, but entrainment of the resident magma during fountaining produces a layer of commingled magmas between the two. (**C**) If the new magma is less dense than the resident one and intrusion takes place slowly, the new magma accumulates near the roof. (**D**) If intrusion of a less dense magma is rapid, entrainment of the resident magma in the fountain again produces a layer of commingled magmas between the two.

Campbell and Turner (1986) have shown that the amount of mixing in a turbulent fountain depends strongly on the velocity at which magma is intruded and the viscosity of the resident magma. They define a Reynolds number based on the viscosity of the resident magma as

$$\mathrm{Re_r} \equiv \frac{2r\bar{v}\rho_\mathrm{r}}{\eta_\mathrm{r}} \tag{14.21}$$

where r is the radius of the input pipe, $\bar{v}$ the mean input velocity, ρ_r the density, and η_r the viscosity of the resident magma. If the viscosity of the resident magma is more than eight times that of the new magma, then for values of $\mathrm{Re_r} > 70$, vigorous mixing occurs along the margins of the fountain, but when $\mathrm{Re_r} < 7$, almost no entrainment occurs. If basaltic magma, for example, is injected into granitic magma, the high viscosity of the granitic magma probably will keep the Reynolds number well below 70, so very little mixing of these magmas occurs at this stage.

If no mixing accompanies the emplacement of a hot dense magma beneath a cooler, less dense one, a stably stratified

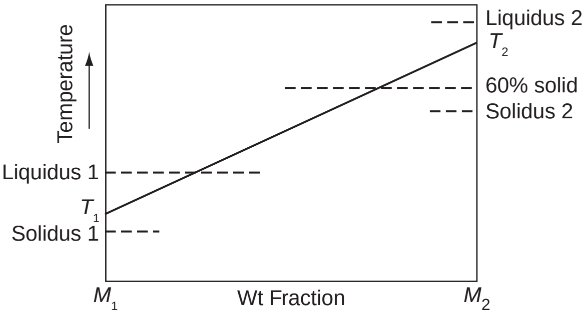


Fig. 14.36 When no crystallization or melting accompanies commingling of magmas, the final temperature of mixtures depends on the temperatures, proportions, and heat capacities of the magmas; final temperatures lie on the line between T_1 and T_2. If magmas contain crystals prior to commingling, resorption of crystals during heating of low-temperature magma results in absorption of heat, and crystallization during cooling of high-temperature magma causes liberation of heat. These additional heat effects are significant (Eq. (14.24)) and typically cause the final temperature of mixtures to plot above the line T_1–T_2.

magma chamber is established, such as that illustrated in Figure 14.10(C). Convection within the lower layer results in heat being transferred into the upper layer, but the density difference and low chemical diffusion rates allow little mixing of the two magmas to occur. This condition continues until the loss of heat from the lower layer results in crystallization; this, in turn, causes the composition and density of the residual liquid to change. If the lower layer is a picritic magma, crystallization of olivine causes the density of the residual liquid to decrease, and once it becomes less dense than the overlying magma, plumes of the fractionated lower liquid rise into the upper magma, and only then does mixing of the two layers take place.

When magmas with different temperatures commingle, heat diffuses so much more rapidly than material that thermal equilibrium is established long before significant homogenization of the magmas occurs. The fate of commingled magmas therefore depends on the physical state of the magmas after thermal equilibration. If both are mainly liquid, homogenization may occur, but if one is largely crystalline, no mixing will take place.

The equilibration temperature of commingled magmas can be represented qualitatively in the mixing diagram of Figure 14.36. Consider two magmas, M_1 and M_2, with temperatures T_1 and T_2, respectively. If neither is superheated at first, their temperatures must fall between the liquidus and solidus for the respective compositions. If the heat capacities of all phases are the same, and we ignore, for the moment, heat associated with crystallization or melting, then the final temperature, T_f, of a mixture of these magmas falls on the line T_1–T_2 at a point determined by the proportions of the two magmas. As the amount of magma 1 increases, the lower is the final temperature of the mixture. However, because the compositions of the two magmas remain essentially unchanged during this period of thermal exchange, the solidus temperatures for both magmas remain the same. Thus, as the proportion of magma 1 increases, the final temperature of the mixture approaches the solidus temperature of liquid 2, and this melt shows higher and higher degrees of solidification. Indeed, once the magma becomes more than 60% crystalline, it behaves essentially as a solid, and no mixing with magma 1 can take place.

To determine quantitatively the final temperature of commingled magmas we must take into account the heat capacities of the phases involved and the latent heat of crystallization of minerals that may form or be resorbed during mingling (Sparks and Marshall, 1986). The heat liberated by either of the magmas on commingling with the other is given by

$$\Delta H = [-C_p(T_f - T) + (L \times \Delta F)]W \qquad (14.22)$$

where C_p represents the heat capacities of all the liquids and solids, which are assumed to be equal, T_f the final temperature of the commingled mixture, T the initial temperature of the magma, L the latent heat of crystallization of minerals in the magma, F the change in the fraction of crystals present in the magma as a result of the change in its temperature, and W the weight of the magma. If T_f is greater than T, then ΔH is negative; that is, heat will be taken in by the magma. If T_f is less than T, then ΔH is positive and heat must be liberated by the magma. When two magmas commingle, the heat given up by one must equal the heat taken in by the other; that is, the sum of the ΔHs must be zero.

$$[-C_p(T_f - T_1) + (L_1\Delta F_1)]W_1 + [-C_p(T_f - T_2) + (L_2\Delta F_2)]W_2 = 0 \qquad (14.23)$$

In this equation, the subscripts 1 and 2 refer to magmas 1 and 2. If we divide both sides of the equation by $W_1 + W_2$, we can replace the actual weights of the two magmas by the weight fractions of the magmas in the mixture; that is, $X_1 = W_1/(W_1 + W_2)$ and $X_2 = W_2/(W_1 + W_2)$. Then, on rearranging Eq. (14.23), we can solve for the final temperature:

$$T_f = (X_1T_1 + X_2T_2) + \left(\frac{X_1L_1\Delta F_1}{C_p} + \frac{X_2L_2\Delta F_2}{C_p}\right) \qquad (14.24)$$

This equation consists of two groups of terms; the first gives the contribution that the temperatures of the two magmas make to the final temperature, and the second group gives the temperature changes due to crystallization or resorption. Note that ΔF can be positive (crystallization) or negative (resorption), and that $X_2 = 1 - X_1$.

When two magmas commingle, the viscosity of the lower-temperature one decreases as its temperature rises toward the final equilibrium temperature, and the viscosity of the higher temperature magma increases as its temperature falls. The final viscosities of the two magmas depend on the equilibrium temperature, which is determined by the proportions of the two magmas (Eq. (14.24)). If no crystallization takes place, the viscosities of the magmas can be expressed as a function of temperature using Eq. (2.11). If the cooling of the magma is accompanied by crystallization, the increase in viscosity can be expressed as a function of the fraction of crystals present according to Eq. (2.9). Of course, as the

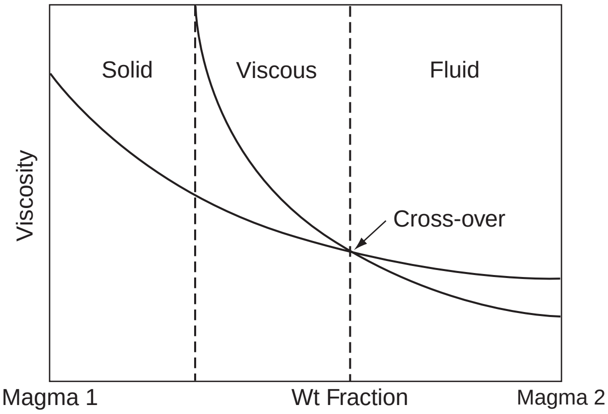


Fig. 14.37 Variation in viscosity of commingled magmas as a function of equilibrium temperature, which depends on the proportions of the two magmas. Magma 1 is a low-temperature siliceous magma, and thus it initially has a high viscosity; on commingling and heating, its viscosity decreases. The viscosity of magma 2, which is more mafic and hotter, increases as it cools and crystallizes; once it is 60% solidified it behaves essentially as a solid. (After Sparks and Marshall, 1986.)

Fig. 14.38 Mafic magma quenched to a pillow-like body in granite, Mount Desert Island, Maine. The mafic magma was more viscous than the granitic one at the equilibrium temperature, as indicated by the cuspate boundary on the pillow; sharp cusps point toward the more viscous liquid.

low-temperature magma becomes hotter, crystals in it are resorbed, and the viscosity of the melt decreases more than it would due to heating alone. Thus, depending on the amount of crystallization or resorption that takes place, the final viscosities of the magmas vary as a function of the mixing proportions in a manner such as that shown in Figure 14.37.

If the two magmas have different compositions – silicic and mafic, for example – the viscosity of the lower-temperature silicic magma is initially greater than that of the hotter mafic magma. On commingling and reaching the same temperature, however, the viscosity of the silicic magma may be greater or less than that of the mafic one, depending on the proportions of the two magmas (Fig. 14.37). Indeed, for one specific mixture, the viscosities are the same. This point is referred to by Sparks and Marshall (1986) as the *crossover* of the viscosity curves. To the silicic side of this point, the viscosity of the mafic melt is greater than that of the silicic one. In the field, evidence for this viscosity crossover is seen in the cuspate boundary developed on mafic pillows in granitic rock (Fig. 14.38); the sharp points on the cusps point toward the melt with the higher viscosity. As the proportion of the lower-temperature silicic melt increases, the viscosity of the mafic melt increases, especially as the mafic melt becomes more crystalline. From Eq. (2.9) we see that as the percentage of crystals in the mafic melt approaches 60%, the bulk viscosity approaches infinity; that is, the mafic phase behaves as a solid. Mixing cannot take place easily once the mafic melt reaches this stage. Small amounts of mafic magma in a silicic magma therefore behave as solid xenoliths.

Based on the relative viscosities following thermal equilibration, commingled magmas can be divided into three types (Fig. 14.37). The first involves mixtures containing a large percentage of the higher-temperature mafic melt, which after mingling is still more fluid than the silicic one. In the second, the proportion of mafic melt is less, and after thermal equilibration, its viscosity is higher than that of the silicic one. In the third type, the mafic magma becomes essentially solid during commingling.

Sparks and Marshall (1986) have calculated how these three types of commingling vary as a function of the composition of the higher-temperature mafic magma (Fig. 14.39(A)). The high solidus temperature of mafic magma with a high magnesium number causes the viscosity to rise rapidly when this magma mingles with a low-temperature siliceous magma. Consequently, the cross-over in the viscosity curves and the point at which the percentage of crystals in the mafic melt reaches 60 occur at relatively low contents of admixed siliceous melt. As the magnesium number of the mafic magma decreases, so does its solidus temperature, and the cross-over and 60% solidification points occur at progressively higher contents of siliceous magma. Thus, a mafic magma with a low magnesium number is more likely to remain fluid after commingling with siliceous magma than is one with a high number.

The composition of the siliceous magma can also vary, especially in water content. This is of particular importance because water both lowers the solidus temperature and decreases the viscosity. By reference to Figure 14.36, it can be seen that when the temperature of the silicic magma (M_1) is lowered by the addition of water, the point at which the mafic magma becomes essentially solid is shifted to smaller admixtures of siliceous melt. Furthermore, the lowering of viscosity caused by the presence of water results in the cross-over in the viscosity curves occurring at more mafic mixtures. The effect of water in the siliceous magma is, therefore, to increase the range of mixed magma compositions in which the mafic fraction becomes essentially solid during commingling. This range would be decreased if the mafic magma also contained water.

The combined effects of variable magnesium number in the mafic magma and water content in the silicic magma on

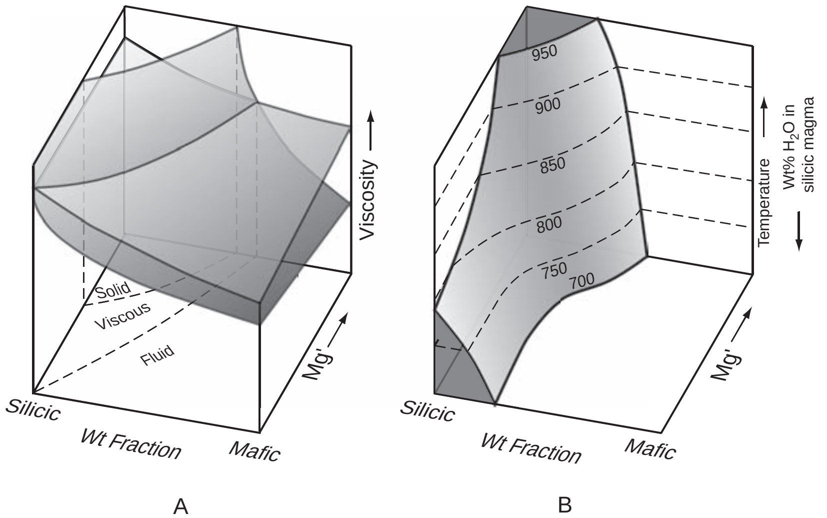


Fig. 14.39 (**A**) Variation of viscosities of commingled silicic and mafic magmas as a function of the proportions of the two magmas and the magnesium number of the mafic magma. Viscosity of the mafic magma indicated by the darker shaded surface and that of the silicic one by the lighter shaded surface. Once the mafic magma is essentially solid, no mixing of the magmas is likely to occur. (**B**) The area marked solid in (A) is the shaded area near the top of this block diagram, which shows the effect of the addition of water to the mixing ability of silicic and mafic magmas. The higher the water content of the silicic magma, the lower is its temperature and the greater quenching effect it has on the commingled mafic melt. As the magnesium number of the mafic melt decreases, so does the temperature of the mafic melt and so the less likely is it to be quenched by the silicic melt. Commingled magmas falling within the shaded volume will be cool enough to cause the mafic melt to become essentially solid and no mixing of the magmas is likely to occur. (After Sparks and Marshall, 1986.)

the commingling of magmas are illustrated in Figure 14.39-(B). It is believed that little homogenization of the magmas can take place if the mafic phase becomes essentially solid during commingling. If, however, it remains in the viscous or fluid fields, homogenization is possible if there is sufficient time. With this proviso then, we see that under all conditions, hybridization is favored by having a large proportion of mafic magma in the mix, especially if it has a low magnesium number. On the other hand, the likelihood of homogenization decreases with increasing water content in the silicic magma.

Production of a homogeneous hybrid magma first involves the intimate commingling of two magmas, but eventually diffusion of components from one to the other must take place. The driving force for diffusion comes from the difference in the chemical potentials a component has in the two melts (Section 9.3). The chemical potential of a component, of course, depends on its concentration (Eq. (9.28)), and components diffuse from regions of high concentration to ones of low concentration to equalize chemical potentials.

The rate at which diffusion can homogenize magmas can be derived from *Fick's law* (Eq. (5.51)):

$$J_z = -D_i \frac{\partial c}{\partial z}$$

where J_z is the flux of a component i diffusing in the z direction in response to a concentration gradient $\partial c/\partial z$ through a medium in which the diffusion coefficient for that component is D_i. The rate of change of concentrations of a component is given by Fick's second law (Eq. (5.52)):

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial z^2}$$

Values for D in silicate melts at magmatic temperatures are very small, being on the order of 10^{-10} to 10^{-14} m^2 s^{-1} (Fig. 5.13)). The rate at which these diffusion coefficients cause compositional gradients to change with time can be calculated using the solutions given in Chapter 5. To obtain an approximate idea of these rates, we can calculate the time necessary to homogenize a sphere of liquid. We will define the homogenization time as that necessary to raise the concentration of the diffusing component at the center of the sphere to more than 98% of its concentration in the material surrounding the sphere, which is assumed to remain constant. Under these conditions $(Dt)^{1/2} = 0.75r$, where r is the radius of the sphere. If we consider, for example, the diffusion of sodium into a sphere of basalt where the value of D_{Na} is 10^{-10} m^2 s^{-1} at 1200 °C, 1.8 years would be required to homogenize the sodium if the radius were 10 cm, and 180 years if the radius were 1 m. At 1000 °C, $D_{Na} = 10^{-11}$ m^2 s^{-1}, and then the 10-cm-radius sphere would take 18 years to homogenize, whereas the 1-m sphere would take 1800 years (Problem 14.21).

The actual calculation of the homogenization of a sphere is far more complicated than this, for as sodium diffuses into the basalt, the change in composition would probably change the diffusion coefficient. In addition, the inward diffusion of sodium would have to be coupled with the outward diffusion of other components in order to conserve mass. Nonetheless, the calculation does provide an order of magnitude estimate of the times and distances involved with diffusion. They clearly show why magmas must be intimately commingled to begin with if they are to produce homogeneous hybrids; diffusion then does not have as far to transport components.

An additional factor making homogenization difficult is that diffusion coefficients decrease exponentially with falling temperature (Eq. (5.53)). If magma mixing brings about a significant lowering of temperature of one of the magmas, the distance components can diffuse in that melt will be dramatically reduced (Problem 14.21).

According to Eq. (5.52), diffusion will be most rapid at first when the concentration gradients are steep; it will then decrease as the gradients diminish, especially if the temperature falls during this time and the diffusion coefficients decrease. It is unlikely, then, that diffusion alone can produce completely homogeneous hybrid magmas in the times available during typical magmatic episodes. Convective movements are probably necessary to help bring about complete homogenization.

A further problem with homogenizing magmas by diffusion is that even when the magmas have significantly different compositions, the chemical potentials or activities of components in the magmas may not be very different because of nonideality in the melts. This can be demonstrated by comparing the activity of SiO_2, for example, in basaltic and granitic magmas (Carmichael *et al*., 1970). If a tholeiitic basaltic magma contains crystals of olivine that are in the process of reacting to form orthopyroxene, we can write the following reaction for the magnesium components:

$$\underset{\text{(olivine)}}{Mg_2SiO_4} + \underset{\text{(magma)}}{SiO_2} = \underset{\text{(orthopyroxene)}}{Mg_2Si_2O_6} \tag{14.25}$$

The silica in this reaction is not pure quartz but is molten silica in solution in the magma. The forsterite is in solid solution in the olivine and the enstatite in the orthopyroxene. The equilibrium constant (*K*) for this reaction can be written in terms of the activities of these three components in their respective solutions; that is,

$$K = \frac{\left(a_{\text{En}}^{\text{Opx}}\right)}{\left(a_{\text{Fo}}^{\text{Ol}}\right)\left(a_{\text{SiO}_2}^{\text{L}}\right)} \tag{14.26}$$

where $a_{\text{SiO}_2}^{\text{L}}$ is the activity of silica in the magma.

In a basalt, the olivine and pyroxene are relatively magnesian, and both solid solutions are nearly ideal, so mole fractions (X_i) can be used in place of activities without introducing serious errors. Furthermore, from Eq. (9.56), $\ln K = -\Delta G/RT$, where ΔG_r is the free-energy change of the reaction. Equation (14.26) can therefore be rewritten to express the activity of silica in the magma as

$$\ln a_{\text{SiO}_2}^{\text{L}} = \frac{\Delta G}{RT} + \ln X_{\text{En}}^{\text{Opx}} - \ln X_{\text{Fo}}^{\text{Ol}} \tag{14.27}$$

The value of ΔG can be obtained from Table 7.1. If the mole fractions of forsterite and enstatite in olivine and orthopyroxene respectively are known, the activity of silica in the basaltic melt can be calculated.

In the granitic melt, the activity of silica might be determined by the presence of phenocrysts of quartz, which would buffer the silica activity through the following reaction:

$$\underset{\text{(quartz)}}{SiO_2} = \underset{\text{(magma)}}{SiO_2}$$

and the activity of silica in the magma is given by

$$\ln a_{\text{SiO}_2}^{\text{L}} = -\frac{\Delta G}{RT} \tag{14.28}$$

Using Eq. (14.27) and (14.28), Carmichael *et al*. (1970) showed that the activity of silica increases from about $10^{-0.15}$ in basaltic magma at 1150 °C to only $10^{-0.10}$ in granitic magma at 850 °C (Problem 14.22). This small increase in activity is surprising in view of the corresponding large increase in weight percent of silica, which goes from 48 to 75. This is clear evidence of the nonideal behavior of these melts. Indeed, as the mafic melt becomes more iron-rich, it approaches the immiscibility field (Fig. 14.26). Once it reaches the two-liquid field, melts with very different compositions coexist, but because the activities of components in both melts are equal, there is no driving force to cause homogenization. Therefore, although Figure 14.39(B) indicates that mafic melts with low magnesium numbers have physical properties that allow them to commingle more easily with silicic magmas than ones with high magnesium numbers, these same melts are closer to the two-liquid field, and the potential for homogenization is diminished.

Under appropriate conditions, magma mixing can produce completely homogeneous melts, in which the only evidence of mixing might come from isotopic data or assemblages of phenocrysts inherited from the parent magmas that are not in equilibrium with the hybrid magma. One of the best-studied examples of such hybrid magmas is the Paricutin Volcano, Mexico (Wilcox, 1954; McBirney *et al*., 1987). This volcano came into existence on February 20, 1943, and for nine years erupted over a cubic kilometer of basaltic andesite and andesite before the eruptive episode ended. The first 75% of the total erupted material varied little in composition, but it did contain xenoliths of siliceous basement rocks in all stages of fusion. The melts derived from these country rocks apparently migrated upward and collected near the roof of the chamber to form a density zoned and stratified magma chamber. The later-stage eruptions tapped the upper parts of this chamber to give a series of hybrid magmas that range in composition from the original basaltic andesite ($SiO_2 = 55$) to andesite ($SiO_2 = 60$). Part of the compositional variation is due to fractional crystallization of olivine and plagioclase, but McBirney *et al*. (1987) conclude that the most contaminated magma assimilated 20% siliceous material. In terms of Figure 14.39(B), this admixture of siliceous melt in a mafic magma (basaltic andesite) with a relatively low magnesium number (60) is consistent with the mafic melt remaining liquid during commingling. It is not surprising, therefore, that this contaminated magma formed a homogeneous hybrid.

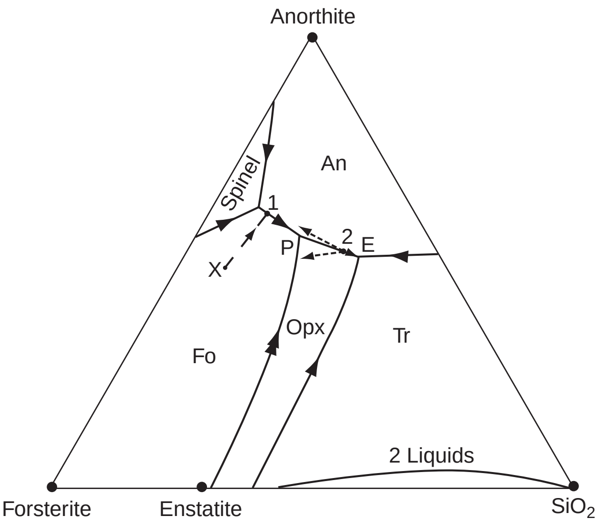


Fig. 14.40 System forsterite–anorthite–silica at atmospheric pressure (Andersen, 1915), showing the compositional trends produced by mixing a derivative magma 2 with either a less evolved magma, 1, or a primary magma, X. In the first case the hybrid magma moves into the primary field of anorthite, and a plagioclase cumulate can form. In the second case, the hybrid magma first moves into the field of orthopyroxene, and then into the field of olivine, allowing cumulates of these minerals to form.

One of the most important consequences of magma mixing is that certain phase relations develop that cannot otherwise form by fractional crystallization of a single magma. Inspection of any multicomponent phase diagram (Chapter 10) reveals that melts become saturated with an increasing number of phases on undergoing fractional crystallization. A melt that starts crystallizing in the primary field of a single mineral soon becomes saturated with a second mineral; these two minerals then crystallize together until the liquid becomes saturated in a third, and so on, until the eutectic is reached where the number of phases is a maximum. A similar increase in the number of phases accompanies the fractionation of natural magmas. It is surprising, therefore, that some magmas that are multiply saturated can abruptly start crystallizing only one mineral. Such behavior is difficult to explain without resorting to magma mixing.

Fractional crystallization of basaltic magmas can, in part, be represented in the simple ternary system forsterite–anorthite–silica (Fig. 14.40). Consider a primary magma of composition X which, because it lies in the primary field of olivine, fractionates along the line X1 away from olivine until it reaches the cotectic with anorthite, whereupon it descends to the peritectic (P). With continued fractional crystallization, the liquid descends the anorthite–orthopyroxene cotectic to the ternary eutectic (E).

If, during the fractional crystallization of magma X, a second pulse of magma mixes and homogenizes with it, the new hybrid magma will have an intermediate composition that in general will not plot on a multiply saturated boundary but will, instead, fall in the primary field of just one mineral. Consider, for example, a resident magma that had fractionated to composition 2 on the anorthite–orthopyroxene boundary when a new pulse of primary magma with composition X mixes with it. The resulting magma would plot on the straight line joining compositions 2 and X, as long as no crystal resorption disturbs the composition. If only a small amount of magma X were added, the hybrid magma would, on cooling, fall in the orthopyroxene field. Accumulation of this mineral would then form a pyroxenite on top of the norite that would initially have been forming from magma 2. If the second pulse of magma were slightly evolved, for example with a composition of melt 1, and it mixed with a magma of composition 2, the hybrid magma would fall in the primary field of anorthite, and an anorthosite could form by accumulation.

Magma mixing may produce cumulates of economic value. Irvine (1977) has shown that layers of chromitite in mafic and ultramafic intrusions may owe their origin to precipitation from evolved magmas that mix with more primitive ones. Because the olivine–chromite cotectic is concave toward chromite, mixing of different magmas on this cotectic produces hybrid melts that fall in the primary field of chromite. The strong curvature on the sulfide–silicate liquid immiscibility field can also be responsible for the accumulation of magmatic sulfide deposits. If an evolved liquid near the ternary eutectic (E) in the system FeS–FeO–SiO_2 (Fig. 11.4) is mixed with less evolved liquid, for example of composition a in Figure 11.4, the hybrid magma would enter the two-liquid field, and a dense immiscible sulfide would separate and possibly segregate.

14.12 TRACE ELEMENT FRACTIONATION BY MAGMAS

Magmatic processes and multicomponent phase relations are still too poorly understood for petrologists to explain precisely the major element variations in suites of differentiated igneous rocks. Trace element variations, on the other hand, are often simpler to account for. Unlike major elements, trace elements are not essential to the stability of the phases involved, and thus they play a relatively passive role. Nonetheless, their concentrations are affected by the magmatic processes.

When more than one phase is present, a trace element under equilibrium conditions partitions itself between the phases so that its chemical potential in each phase is the same; that is, for example, $\mu_i^\alpha = \mu_i^\beta$, where α and β are two of the coexisting phases (mineral, liquid, vapor). Making use of Eq. (9.34), we can write

$$RT \ln \frac{a_i^\beta}{a_i^\alpha} = -\left(\mu_i^{*\beta} - \mu_i^{*\alpha}\right) = -\Delta G^* \qquad (14.29)$$

where ΔG^* is the free-energy change for the transfer of element i from phase α to phase β at the particular pressure and temperature. Because trace elements are present in very small amounts, they can be expected to follow Henry's law (Eq. (9.36)), in which case Eq. (14.29) can be rewritten as

$$RT \ln \frac{K_i^\beta X_i^\beta}{K_i^\alpha X_i^\alpha} = -\Delta G^* \qquad (14.30)$$

Table 14.3 *Approximate mineral/basalt distribution coefficients at near-liquidus temperatures* Note that the coefficients change with pressure, temperature, and composition of the magma and the mineral (data from various sources).

Z	Element	Olivine	Orthopyroxene	Augite	Amphibole	Plagioclase	Garnet
12	Mg	6.1	5.5	3.5	3.8	–	–
19	K	0.007	0.015	0.03	0.60	0.17	–
21	Sc	0.17	1.2	2.7	3.2	–	4
22	Ti	–	–	0.8	–	–	–
23	V	0.06	0.6	1.3	–	–	–
24	Cr	2.1	10	34	0.15	–	0.17
25	Mn	1.2	1.4	0.9	–	–	–
26	Fe″	1.9	1.8	0.8	–	–	–
27	Co	3.8	3	1.2	1.6	–	1.1
28	Ni	13.5	5	2.3	–	–	–
37	Rb	0.01	0.025	0.04	0.25	0.1	0.042
38	Sr	0.014	0.015	0.14	0.57	1.8	0.012
55	Cs	0.0004	–	0.01	0.1	0.025	–
56	Ba	0.01	0.013	0.07	0.31	0.23	–
57	La	0.007	0.0028	0.08	0.27	0.19	0.01
58	Ce	0.006	0.0032	0.34	0.34	0.14	0.03
60	Nd	0.009	0.041	0.6	–	0.08	0.2
62	Sm	0.009	0.058	0.8	0.91	0.08	0.6
63	Eu	0.009	0.078	0.9	1.01	0.32	0.49
64	Gd	0.012	–	0.9	1.1	0.1	3
65	Tb	–	0.05	1.0	1.4	0.1	5
66	Dy	0.012	0.2	1.1	–	0.09	10
68	Er	0.013	0.022	0.8	–	0.08	24
70	Yb	0.027	0.039	0.58	0.97	0.07	30
71	Lu	0.03	0.053	0.53	0.89	0.08	35

Z, atomic number.

where K_i^β and K_i^α are the Henry's law constants for element i in phases β and α, respectively. Rearranging Eq. (14.30) gives

$$\ln\frac{X_i^\beta}{X_i^\alpha} = -\frac{\Delta G^*}{RT} - \ln\frac{K_i^\beta}{K_i^\alpha} \tag{14.31}$$

All terms on the right-hand side of this equation are constants at a given temperature and pressure, and therefore the ratio of X_i^β / X_i^α must also be a constant, as long as Henry's law is obeyed. If the solutions are dilute, the mole fractions will be proportional to the concentrations (c_i), which for trace elements, are typically given in parts per million by weight (ppm). Equation (14.31) can therefore be rewritten as

$$\ln\frac{c_i^\beta}{c_i^\alpha} = -\frac{\Delta G^*}{RT} - \ln\frac{K_i^\beta}{K_i^\alpha} \tag{14.32}$$

The ratio of the concentration of the element in the two phases (c_i^β / c_i^α) is known as the *Nernst distribution* or *partition coefficient* and is represented by K_D.

Crystal–liquid partition coefficients can be greater or less than 1 depending on the element's preference for the crystal or liquid (Table 14.3). Nickel, for example, substitutes for magnesium and iron in the structures of olivine and orthopyroxene, and so it partitions into these minerals, giving K_D values of about 15 and 5, respectively. Phosphorus, on the other hand, has no sites to occupy in early-crystallizing minerals and consequently partitions into the liquid. Elements that partition strongly into early-crystallizing minerals ($K_D^{s/l} \gg 1$) are said to be *compatible*, whereas those with a strong preference for the liquid ($K_D^{s/l} \ll 1$) are said to be *incompatible*. This distinction refers normally to basaltic melts and their early-crystallizing minerals or the mantle minerals from which basalt is extracted, but it can be extended to more siliceous magmas if it is recognized that crystallization of different minerals may change element compatibilities. In general, incompatible elements are those that cannot substitute into the structures of early-crystallizing minerals because of high ionic charge, large ionic radius, or both. They include, P, K, Ti, Rb, Sr, Zr, Cs, Ba, rare earth elements (REE), Hf, Th, and U.

In Eq. (14.32), the $\Delta G^*/RT$ and the Henry's law constants are functions of temperature and so, also, must be the partition coefficient. Indeed, trace element distributions that are strongly temperature dependent have been used as geothermometers. Hakli and Wright (1967), for example, were able to show by sampling basaltic magma from the Makaopuhi lava lake, Hawaii, that the value of K_D for nickel between olivine and liquid increases from 13.5 at 1160 °C to 16.8 at 1050 °C. Over this same interval, the augite/liquid K_D increases even more, from 2.22 to 4.40. With these data a third distribution coefficient can be defined, that of nickel between olivine and augite, which decreases from 6.10 at 1160 °C to 3.82 at 1050 °C. This latter distribution coefficient is more useful than the first two because liquids in the form of quenched glasses may not be present or well preserved in rocks, especially nonvolcanic ones. Olivine and clinopyroxene crystals, however, can preserve the magmatic partitioning of nickel long after all liquid has crystallized, as long as there is no subsolidus reequilibration.

Distribution coefficients for many trace elements do not vary significantly over the temperature intervals through which most magmatic differentiation takes place. They can therefore be used as constants to determine the equilibrium distribution of trace elements between crystals and magma during differentiation. How the concentration of the element varies during differentiation depends on whether or not crystals remain in equilibrium with the melt from which they form. There are two extreme cases. In one, surface layers grow on crystals in equilibrium with the liquid, but once formed they become isolated from the liquid, as a result of being covered by another layer through which there is insufficient time for diffusion to transfer material, or the crystal is completely removed from the liquid from which it grew. In the second case, the trace element in the crystal continuously reequilibrates with the magma as it cools and changes composition.

The first type of behavior, known as *Rayleigh fractionation*, has already been dealt with in Section 12.4 in connection with the growth of zoned garnet crystals from chlorite. If in Eq. (12.21) we substitute liquid for chlorite and solid for garnet, and let F be the fraction of residual liquid, the equation becomes

$$c_i^s = c_i^{\mathrm{ol}} K_i^{s/l} F^{(K-1)} \tag{14.33}$$

where the concentration of trace element i in the initial liquid is $c_i^{\mathrm{o}l}$. Because $c_i^s / c_i^l = K_i^{s/l}$, Eq. (14.33) can be rewritten as

$$c_i^l = c_i^{\mathrm{ol}} F^{(K-1)} \tag{14.34}$$

where c_i^l is the concentration in the fraction F of liquid remaining. When more than one mineral crystallizes, a bulk distribution coefficient must be used that weights the coefficients for the individual minerals in proportion to their abundance.

When the magma remains in equilibrium with the solids at all times the concentration of the trace element in the solids (c_i^s) is homogeneous throughout; it will be related to the concentration of the element in the remaining magma (c_i^l) by the Nernst distribution coefficient (K_i). The amount of trace element in the solids and in the residual magma must at all times by equal to the amount in the initial magma (c_i^{ol}). If the fraction of magma remaining is F, the solids formed will be $(1 - F)$. We can then write a mass balance equation for the trace element as follows:

$$c_i^l F + c_i^s (1 - F) = c_i^{\mathrm{o}l} \tag{14.35}$$

But $c_i^s = K_i c_i^l$, so Eq. (14.35) can be rewritten as

$$c_i^l = \frac{c_i^{\mathrm{o}l}}{F + K(1 - F)} \tag{14.36}$$

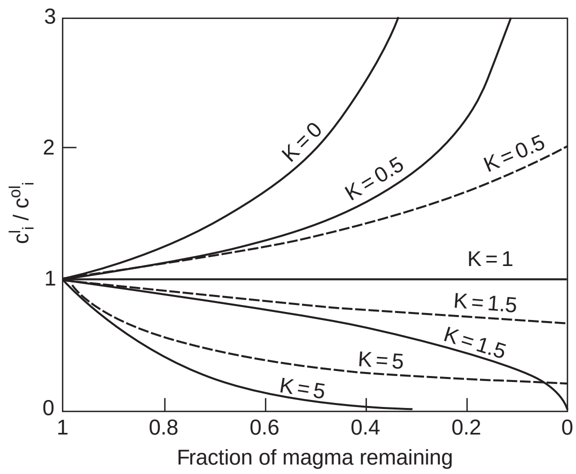


Fig. 14.41 Change in concentration of component *i* relative to its initial concentration in melts undergoing equilibrium crystallization (dashed curves) or Rayleigh fractionation (solid curves). The various curves are for elements with different crystal/liquid partition coefficients (Nernst distribution coefficient, K_D).

which is known as the *Berthelot–Nernst* equation for equilibrium crystallization. Because $c_i^s / c_i^l = K_i$, Eq. (14.36) can be rewritten for the concentration in the solid as

$$c_i^s = \frac{c_i^{ol} K_i}{F + K(1 - F)} \tag{14.37}$$

Again, if more than one mineral crystallizes, a bulk distribution coefficient must be used.

Figure 14.41 shows how the concentrations of trace elements in a magma change as a result of Rayleigh fractionation (solid lines) and equilibrium crystallization (dashed lines) when minerals with various distribution coefficients crystallize. Incompatible elements ($K < 1$) remain in the liquid and become more concentrated as the fraction of liquid decreases. For a completely incompatible element ($K = 0$), such as phosphorus in a primitive basalt, Eqs. (14.34) and (14.36) become identical, so that the rate of enrichment in the melt is independent of the crystallization process. This means that the enrichment of such an element can be used as a direct measure of the degree of fractionation of the magma. From Eq. (14.34) or (14.36), if $K = 0$, then

$$F = \frac{c_i^{\mathrm{o}l}}{c_i^l} \tag{14.38}$$

For values of K other than zero, Rayleigh fractionation leads to more rapid changes in the concentration of trace elements with crystallization than does the Berthelot–Nernst equilibrium crystallization. When the value of K is less than 1, elements become enriched in the melt, and when K is greater than 1, the melt becomes depleted in the element (Problem 14.24).

The ratios of strongly incompatible elements ($K = 0$) can be used to test for common lineage in suites of igneous rocks. The concentration of an incompatible element in a residual liquid is, according to Eqs. (14.34) and (14.36),

$$c_i^l = \frac{c_i^{\text{ol}}}{F} \tag{14.39}$$

Thus the ratio of any two incompatible elements, *a* and *b*, is

$$\frac{c_a^l}{c_b^l} = \frac{c_a^{\text{ol}}}{F}\frac{F}{c_b^{\text{ol}}} = \frac{c_a^{ol}}{c_b^{\text{ol}}} = \text{constant} \tag{14.40}$$

Clearly, the ratio of these two elements is independent of the fraction of liquid remaining and is simply equal to the initial ratio. Thus regardless of the degree of fractionation or the minerals crystallizing, the ratio of incompatible elements remains the same and serves as a signature of that lineage. It can be changed only by the assimilation of country rocks or mixing with magmas that have a different ratio.

The concentration of trace elements that have partition coefficients that differ significantly from unity can change by orders of magnitude during fractional crystallization. These elements can therefore be sensitive indicators of the degree of fractionation. In addition, because different minerals have different distribution coefficients, the appearance of a new primary phase can produce a marked change in the trace element variation. Vanadium, for example, is an incompatible element in primary basalts and therefore becomes concentrated in fractionated liquids. However, it partitions strongly into magnetite, so that once this mineral starts crystallizing the concentration of vanadium in the melt drops sharply. Similarly, strontium is an incompatible element in basaltic magma until plagioclase starts crystallizing.

The rare earth elements (REE, lanthanides La-57 to Lu-71) are particularly useful trace elements in the study of igneous petrogenesis. Each of these elements occurs in the trivalent state under normal magmatic conditions, except for europium, which occurs both as Eu^{2+} and Eu^{3+}, depending on the oxygen fugacity. They all occur in sixfold coordination, but as their atomic weight increases, the ionic radius decreases from 1.03 Å in lanthanum to 0.86 Å in lutetium. Because of their identical charge and similar ionic radii, chemical processes do not discriminate strongly between the various REE, and when there is discrimination it varies systematically through the series; europium is the only exception, because of its different valence states. The combination of high charge and relatively large ionic radius make the REE incompatible in most early-crystallizing minerals. Their K_D values, however, do vary slightly – the larger the ionic radius, the more incompatible is the element. In addition, the abundance of each even-numbered REE is greater than that of adjacent odd-numbered ones. Plots of absolute abundance versus atomic number therefore have a distinct saw-tooth appearance. This pattern is commonly smoothed by dividing the abundance of each REE by its abundance in a chondritic meteorite, which also shows similar even–odd fluctuations in abundance. The resulting chondrite-normalized plot (Fig. 14.42) provides a convenient means of viewing the overall enrichment in REE relative to a chondritic composition (presumably, a primordial composition from which the solar system formed; Section 6.1) and the degree of fractionation of the light versus heavy REE.

The REE distribution coefficients between basaltic magma and the early-crystallizing minerals are sufficiently varied that REE abundance patterns in rocks can commonly be used to determine the minerals responsible for fractionation. In Figure 14.42, the distribution coefficients in Table 14.3 have been used to construct equilibrium distribution patterns of REE between the common primary minerals and a basaltic magma that is assumed initially to have had a chondritic abundance of REE; that is, $c^l/c_{\text{chondrite}} = 1$. Two sets of curves are given for each mineral. The lines marked with squares indicate the chondrite-normalized abundance of each REE in the basaltic magma following 75% crystallization ($F = 0.25$ in Eq. (14.36)). The lines marked with Xs indicate the abundance of each REE in the coexisting solid (Eq. (14.37)).

Olivine, which is by far the most common primary mineral in basalts, almost completely excludes the REE, with the heavier (smaller ionic radii) ones being only slightly more compatible than the light ones. Early crystallization of olivine therefore invariably results in strong enrichment of REE in the melt, but the slope in the plot of chondrite-normalized abundance versus atomic number remains almost horizontal.

The REE are slightly less incompatible with respect to orthopyroxene, especially those toward the heavy end. Fractionation of orthopyroxene therefore results in the development of a prominent negative slope throughout the chondrite-normalized plot. Only the lightest three REE are incompatible with respect to augite; the remainder have distribution coefficients very near unity. As a result, fractionation of augite causes no change in the abundance of the REE except for La, Ce, and Nd, which become concentrated in the liquid. Separation of augite from a basaltic magma causes a distinct steepening of the REE pattern at the light end of the chondrite-normalized plot. Amphibole behaves in a similar way to augite.

All the REE exhibit approximately the same degree of incompatibility with respect to plagioclase, except for europium, which is significantly less incompatible than the rest. Eu^{3+} is probably as incompatible as the other trivalent REE, but Eu^{2+} is far less incompatible, substituting for Ca^{2+} in plagioclase. The more reducing the conditions, the greater will be the proportion of Eu^{2+} and the greater will be the difference in the europium distribution coefficient from those of the other REE. Because plagioclase always has a positive europium anomaly, that is, europium is relatively more abundant than the other REE in plagioclase, magmas from which plagioclase separates develop a negative europium anomaly. Such a negative anomaly exists in all lunar mare basalts, indicating that plagioclase fractionation was involved in their genesis.

Finally, garnet preferentially takes in REE with masses greater than europium and rejects those that are less. Consequently, basalt that has undergone garnet fractionation develops a strong negative slope to its REE pattern in the chondrite-normalized plot.

Basalt fractionation typically involves several minerals, making it necessary to calculate bulk distribution coefficients

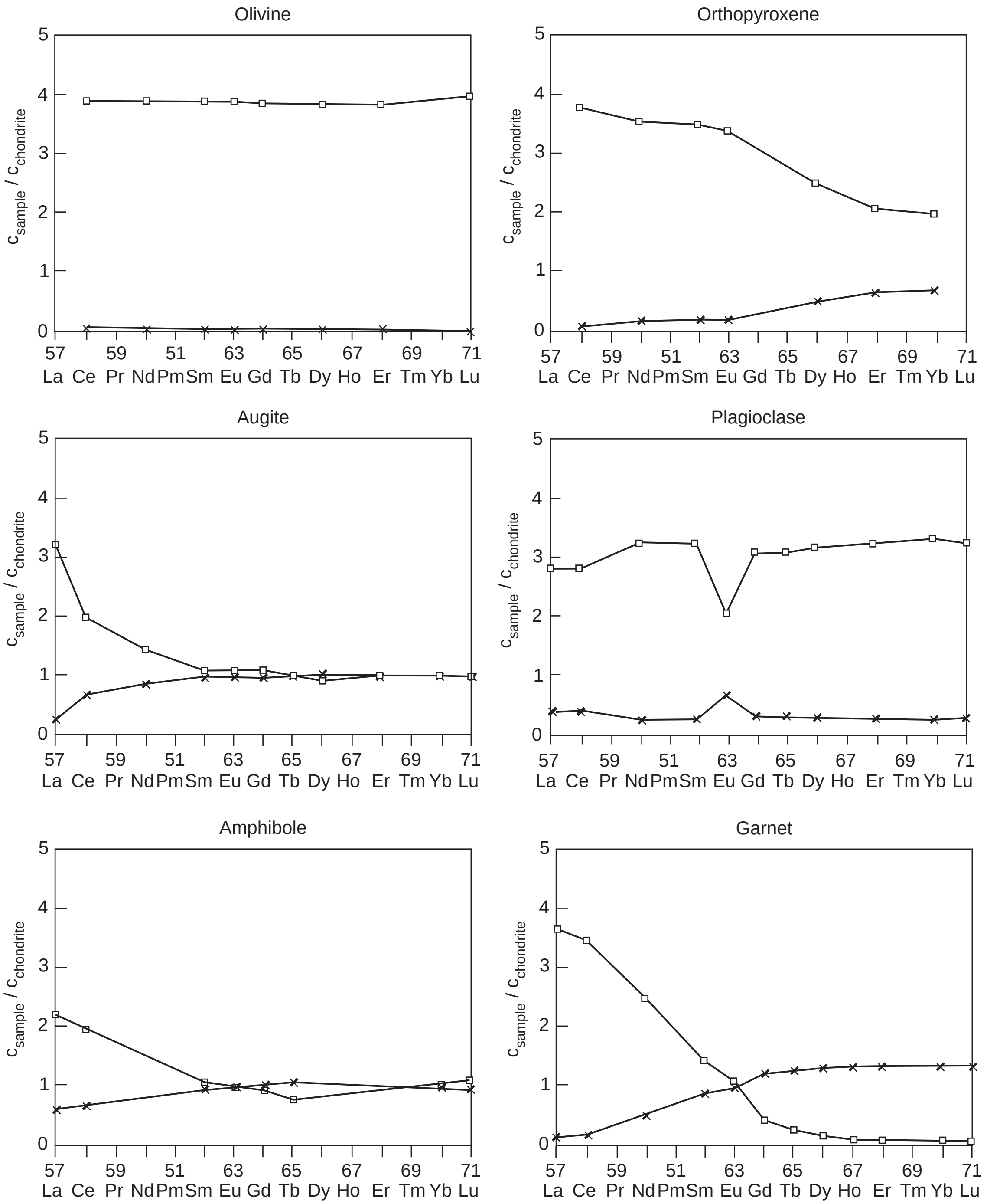


Fig. 14.42 Chondrite-normalized abundances of the rare earth elements partitioned between common minerals (crosses) and basaltic melt (squares) following 75% equilibrium crystallization. The melt initially had a chondritic abundance of rare earth elements; that is, $c_{sample}/c_{chondrite} = 1$. See text for discussion.

for the REE. Depending on the weight proportions of these minerals, various patterns develop in the chondrite-normalized plot (Problem 14.25 and 14.26). Some of these patterns are sufficiently characteristic that they are diagnostic of the minerals causing the fractionation. In Figure 14.43, typical REE patterns are given for five major types of basalt (BVSP, 1981). Each of these types shows varying degrees of REE enrichment relative to a chondritic composition, which is assumed to be similar to the primordial composition of the Earth. Such enrichment is consistent with the general incompatibility of REE. The pattern for each basalt, however, is distinct, indicating differences in the minerals involved in their fractionation or generation.

The simplest pattern is that of the island arc basalt (IAB), which shows a remarkably constant enrichment relative to chondrites of about ten times over the entire range of REE.

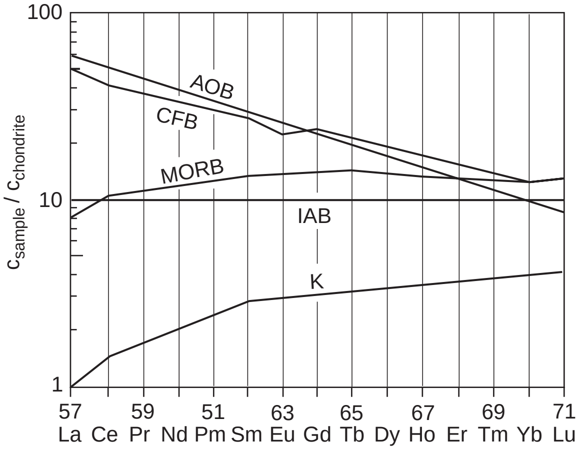


Fig. 14.43 Chondrite-normalized plot of rare earth abundances in typical island arc basalts (IAB), alkali olivine basalts (AOB), tholeiitic continental flood basalts (CFB), mid-ocean ridge basalt (MORB), and Archean komatiites (K). See text for discussion.

Reference to Figure 14.42 indicates that only olivine fractionation could give this flat REE pattern. Alkali olivine basalts (AOB) also show a simple linear pattern but with strong enrichment in light REE. This enrichment could result from high degrees of fractional crystallization, but the high magnesium number and remarkably consistent REE pattern of these basalts argues against such an origin. Instead, the high enrichment is interpreted to result from very small degrees of partial melting in the source region, as discussed in Chapter 22. Also, because the light REE are strongly enriched relative to the heavier ones, garnet is likely to be a stable phase in that source region. Tholeiitic continental flood basalts (CFB) also show enrichment in the light REE relative to the heavy ones, but the effect is not as pronounced as in the alkali basalts. Moreover, many continental flood basalts exhibit a distinct negative europium anomaly, indicating that plagioclase has been removed. It is significant that none of the other basalt types has a negative europium anomaly, even though most contain plagioclase phenocrysts. Clearly, no significant amount of plagioclase can have been fractionated from these other basalts.

Two of the most peculiar, yet characteristic, REE patterns are those of mid-ocean ridge basalts (MORBs) and Archean komatiites (K). Both are less enriched in light REE than heavy ones. Inspection of Figure 14.42 reveals that none of the common early-crystallizing minerals causes such enrichment. Fractional crystallization of magma derived from a source of chondritic composition cannot, therefore, be responsible for this pattern. This problem, however, has a simple solution. Consider a source containing augite that initially has a chondritic REE abundance. Partial melting of this rock gives rise to magma enriched in light REE (Fig. 14.42 augite) and leaves residual augite depleted in light REE. Subsequent melts derived from this depleted source will, themselves, be depleted in light REE relative to heavy ones. Mid-ocean ridge basalts and komatiites are therefore interpreted to come from a mantle source that has been depleted in incompatible elements by earlier periods of partial melting.

In closing this chapter, it is important to emphasize that chemical variations in suites of igneous rocks are not due only to fractional crystallization. Another important factor is the partial melting and magma extraction process that takes place in the source region. This topic is dealt with in Chapter 23, where we will see that the Nernst distribution coefficients and Rayleigh and equilibrium fractionation models (Eqs. (14.34) and (14.36)) are as applicable to partial melting as they are to fractional crystallization.

14.13 PROBLEMS

14.1 Using the data in Table 14.1, construct a triangular variation diagram for the weight percentages of CaO, Na_2O, and K_2O in the Kilauean lavas. Discuss the trend obtained in light of the variations found in the diagram involving the magnesium number (Fig. 14.1).

14.2 **(a)** Using the MELTS program, determine the liquidus temperatures for the rocks represented by analyses 2, 5, and 7. How do these temperatures compare with the observed or experimentally determined eruption temperatures given in Figure 14.2? Explain any significant differences.

(b) If the lava represented by analysis 5 was seen to erupt at 1200 °C, what percentage of phenocrysts would the lava be expected to have contained at the time of eruption according to the MELTS program?

14.3 Magma rising beneath Kilauea is thought to pond in a magma chamber at depths between 6.5 and 5.7 km before erupting at the summit or intruding into the east rift zone. Use the MELTS program to calculate the crystallization sequence in magma with the composition of lava 7 in Table 14.1 if it crystallized near the base of this chamber where the pressure is 0.2 GPa (2 kb). Compare the crystallization sequence with that at 1 bar in Problem 14.2.

14.4 If the compositional trend in the Kilauean basalts from analysis 8 to 15 in Table 14.1 is due to assimilation of more siliceous rocks, use a Harker variation diagram to determine the composition and amount of the most siliceous material that could be added to analysis 8 to produce 15. Calculate the CIPW norm of the assimilated material, and comment on the likelihood of this assimilation mechanism being responsible for the compositional variation of lavas between number 8 and 15. In extending the trends to higher silica contents in the Harker diagram, certain oxides decrease, and where the first one reaches zero must be the most siliceous material that can be assimilated, because negative values of components are not possible.

14.5 The Kilauean lava whose analysis is given as number 10 in Table 14.1 contains only 1.5% olivine phenocrysts; its analysis can therefore be treated essentially as that of a liquid. Using a K_D of 0.33, what forsterite content would you expect these phenocrysts to have? Does this composition agree with the composition of the first crystallizing olivine indicated by the MELTS program?

14.6 From the MELTS *out file* for analysis 7 in Table 14.1 calculated at both 1 bar (Problem 14.2) and 2 kb (Problem 14.3) what were the densities of the first-formed plagioclase crystals and the liquid from which they formed? Comment on the potential for plagioclase to separate from its host liquid on the basis of density.

14.7 Calculate the sinking velocities of olivine crystals with a density of 3500 kg m^{-3} and diameter of 2 mm and chromite crystals with a density of 5010 kg m^{-3} and diameter of 1 mm in magma with a density of 2600 kg m^{-3} and Newtonian viscosity of 300 Pa s. Express your answer in cm day^{-1}. Check that Stokes' law is applicable by calculating the Reynolds number in both cases. Comment on the relative importance of a mineral's density and grain size to rate of sinking. (Be careful with units and determining the radius of crystals.)

14.8 Stubby augite phenocrysts, with a density of 3410 kg m^{-3}, occur in the upper chilled margin of a 3.41-m-thick mafic alkaline sill and then are completely absent down to a distance 0.85 m above the base where they form a cumulate layer. Augite phenocrysts in the upper chilled margin have diameters of 1 mm, whereas those in the cumulate layer have diameters of 2 mm. Calculations show that the magma took 7.22×10^5 s to solidify up to the 0.85-m level. This, then, was the length of time the crystals had to sink. The density of the magma was 2540 kg m^{-3}. Calculate the Newtonian viscosity of the magma assuming (**a**) that the crystal growth rate was controlled by a phase boundary reaction, and (**b**) that the crystal growth rate was controlled by a surface nucleation process. (*Hint*: First evaluate the constants in the growth rate equations (12.8) and (12.9).)

14.9 **(a)** Derive an expression for the distance through which a spherical crystal of radius r will sink in magma if the crystal growth rate is diffusion-controlled and the viscosity and density of the magma remain constant. Use Eq. (12.6) for the crystal growth rate. Plot your equation in a graph of depth versus time.

(b) How do you think the velocity of the crystal will affect the growth rate of the crystal? How might the real sinking velocity of the crystal differ from that calculated in part (a)?

14.10 If the magma in Problem 14.7 has a yield strength of 50 Pa, what minimum sizes of olivine and chromite grains would be required to cause crystal settling?

14.11 If magma in which crystal sinking has occurred has a yield strength, the minimum size of cumulate crystal on the floor of the chamber is determined by the yield strength of the magma and the density contrast between the crystal and the magma (Eq. (14.8)). Because we can measure the size and density of mineral grains in a sample of cumulate, we are left with two unknowns, the density and yield strength of the magma. If two or more minerals have accumulated together (Fig. 14.24(A)), we can determine the density and yield strength of the magma, assuming, of course, that the yield strength played a role in determining the grain size. Later resorption of, or addition to, grains would disturb the relations. Olivine and chromite grains such as those illustrated in Figure 14.24(A) have diameters of 2.54 and 0.88 mm, respectively, and their densities are 3500 and 5010 kg m^{-3}, respectively. With the assumptions above, calculate the density and yield strength of the magma. Do the results indicate that the proposed mechanism determining grain size has any validity?

14.12 A flood basalt is erupted rapidly and forms an extensive 50-m-thick lava lake. A 5-m-thick crust develops rapidly due to radiation cooling and water percolating into surface cracks. A 1-m-thick chill zone forms at the base of the lake, but the low thermal diffusivity of underlying rocks does not allow it to continue thickening rapidly. Following some cooling, the magma in the lake has an average viscosity of 10^3 Pa s, a density of 2700 kg m^{-3}, a thermal diffusivity of 5×10^{-7} m^2 s^{-1}, and a coefficient of thermal expansion of 5×10^{-5} K^{-1}. The surface of the lava lake is kept at the ambient temperature of 25 °C, which in turn keeps the base of the crust (top of magma) at 1065 °C. The top of the lower chill zone remains almost constant at 1150 °C; heat lost into the underlying rocks is replaced by heat liberated by crystallization.

(a) Assuming that these conditions hold relatively constant for some period of time, will the magma convect?

(b) If convection does occur, what would the maximum convective velocity be?

(c) Approximately how long would the lava lake take to crystallize if there is convection, and how long if there is not?

(d) Assuming that the convective velocity calculated in part (b) is valid for the first quarter of the crystallization period, how many revolutions might a neutrally buoyant plagioclase phenocryst make around a convecting cell in this time?

14.13 **(a)** If the heat loss from the convecting magma in the Palisades Sill was 4 W m^{-2} and the latent heat of crystallization was 400 kJ kg^{-1}, what was the rate of crystallization in kg m^{-2} day^{-1}? (Assume that crystallization takes place isothermally.)

(b) If crystallization takes place on the floor of the sill, and the cumulate minerals, which have a density of 3300 kg m^{-3}, form a layer with 40% porosity, what would the daily rate of rise of this layer be?

(c) At the rate determined in part (b), how long would it take the floor of the sill to rise to the olivine layer 15 m above the base?

(d) If the Rayleigh number for the convecting Palisades magma was 10^{11}, what approximately would the Nusselt number have been?

(e) Using the Nusselt number, calculate what the rate of rise of the cumulate layer would have been had the magma not convected. How long would it take in the nonconvecting magma to accumulate the same thickness as formed in one day in the convecting magma?

14.14 Draw a schematic velocity profile through the margin of a convecting basaltic magma, where partial melting of the country rocks has developed a boundary layer of buoyant granitic magma.

14.15 Using the method of Bottinga *et al*. (1982) (Section 2.3), or the MELTS program, determine the densities of the following Kilauean lavas (Table 14.1) at the given temperatures, assuming that analyses are representative of liquids: no. 4, 1270 °C; no. 5, 1265 °C; no. 8, 1200 °C; no. 9, 1160 °C; no. 11, 1100 °C; no. 13, 1075 °C; no. 14, 1040 °C; and no. 15, 1020 °C. Plot the density of the lavas as a function of the magnesium number and comment on the shape of the graph.

14.16 From the book's web site (www.cambridge.org/philpotts), download the grayscale image of the layered gabbro from Stavanger, Norway, which is illustrated in Figure 14.17. The image has been rotated so that top is to the right. Open this image in NIH Image. The scale can be set by drawing a line from the left to the right side of the image; then go to *Set scale* under *Analyze* in the top menu bar and enter 75 cm for the known length of the line. A profile of the abundance of feldspar (light-colored pixels) through the graded layers can be obtained by first drawing a line from left to right (bottom to top layer) and then selecting **Plot profile** from the draw-down menu under *Analyze*. The resulting plot contains a lot of noise due to the grain size of the rock. A better result can be obtained by dragging a rectangle from left to right and repeating the same command to create a profile. This averages the values along vertical lines and creates a much better plot. The actual percentage of feldspar (or conversely dark minerals) can be determined if a small area of feldspar is selected and then defined as pure feldspar under *calibrate* in the *Analyze* menu. Similarly, a small area selected on a dark mineral can be defined as 100% dark mineral. For our purposes, the relative scale is adequate. Scroll over the profile and **Save as** a Jpeg file. Next, download the grayscale image of the graded layering in the Palisades Sill (Fig. 14.19(B)). By following the same procedure as above, create a profile across this layering, and again save the image as a Jpeg file. Finally, open both Jpeg files in any drawing program. Note that in each of the images top is to the right. If the Palisades' profile is flipped, so that top is to the left, the pattern created by the graded layering will then match the normal graded layering exhibited by the Stavanger gabbro.

14.17 Two spherical 1-m-diameter xenoliths, one of quartzite and the other of granite and both at 500 °C, are incorporated into a basaltic magma. Both xenoliths absorb heat from the surrounding magma, but the granitic one, unlike the refractory quartz one, undergoes melting and thus absorbs additional heat. By the time the xenoliths reach 1000 °C the granitic one is totally melted.

(a) If both xenoliths have a density of 2650 kg m^{-3}, their heat capacities (for solid and liquid) are 0.8 kJ kg^{-1} °C^{-1}, and the latent heat of fusion of the granite is 400 kJ kg^{-1}, calculate the total amount of heat withdrawn from the surrounding magma by each xenolith in raising its temperature to 1000 °C.

(b) If the heat withdrawn from the surrounding magma by the xenoliths results in crystallization of the basalt, what would the relative thicknesses of crystal mush be around the two xenoliths?

14.18 A basaltic magma at 1200 °C and containing 10 wt% phenocrysts commingles with a 900 °C rhyolitic magma containing 10 wt% phenocrysts of cristobalite. The weight proportions of basaltic to rhyolitic magma are 80 to 20. After commingling, all cristobalite phenocrysts are resorbed, whereas the percentage of phenocrysts in the basaltic magma increases to 30 wt%, which is made up of 60 wt% pyroxene and 40 wt% plagioclase. If the heat capacities of all phases are taken to be 0.8 kJ kg^{-1} °C^{-1} and the latent heats of fusion of cristobalite, pyroxene, and plagioclase are 135.8, 587, and 490 kJ kg^{-1}, respectively, what is the final temperature of the commingled magmas?

14.19 A basaltic and a rhyolitic magma commingle. Their initial temperatures are 1200 and 900 °C, respectively, and neither contain any crystals at first. Their viscosities obey an Arrhenius relation (Eq. (2.11)), where η_0 is 5×10^{-9} Pa s for the basalt and 10^{-13} Pa s for the rhyolite; their activation energies are 300 and 500 kJ mol^{-1}, respectively. On commingling, the basaltic magma cools and begins to crystallize, whereas the rhyolitic one becomes superheated. The mass fraction of crystals formed in the basaltic melt, F, is a linear function of temperature: $F = 4 - T$ °C/300. The latent heat of crystallization is 500 kJ kg^{-1}, and the heat capacity of the magma is 0.8 kJ kg^{-1} K^{-1}. First, determine the temperature of the basaltic magma as a function of degree of crystallization, and then determine what admixture of rhyolitic magma would give that temperature. Second, plot the $\log_{10}$ of the viscosities of both magmas as a function of the weight fraction of rhyolite in the commingled mixture, making use of Eq. (2.9) to account for the crystallization of the basalt. Finally, at what fraction of rhyolite does the crossover in the viscosity curves occur?

14.20 **(a)** A granitic magma containing 5 ppm Rb and 5 ppm Cs is intruded by a second granitic magma containing 105 ppm Rb and 105 ppm Cs. Diffusion coefficients for Rb and Cs at 900 °C, the temperature of both magmas, are $10^{-12.3}$ and $10^{-15.3}$ m^2 s^{-1}, respectively. Calculate how long it would take for the concentration of Rb in the Rb-poor melt to rise to 25 ppm at a distance of 1 cm from a planar interface between the two melts (see Eq. (5.57)).

(b) What is the concentration of Cs in this melt at this time?

(c) How near to the contact do you have to go to find a Cs concentration of 25 ppm?

(d) In light of the difference in the diffusion coefficients for Rb and Cs, which element would be best to use if you were looking for evidence that an apparently homogeneous magma had been formed by mixing of magmas with different trace element contents?

14.21 If, at 1200 °C, one year is required for the sodium concentration in the core of a spherical globule of molten basalt immersed in granitic magma to rise to 98% of that in the surrounding granite, how long would it take to achieve the same result at 1100 °C if the preexponential factor (D_0) in the Arrhenius relation describing the temperature dependence of the diffusion coefficient (Eq. (5.53)) is 9.6×10^{-5} m^2 s^{-1}, and the activation energy (E) is 163 kJ mol^{-1}?

14.22 Basaltic magma containing crystals of olivine (Fo_{80}) and orthopyroxene (En_{77}) commingles with rhyolitic magma containing quartz crystals. The free energies of formation of pure forsterite, enstatite, quartz, and silica liquid are –2 372 478, −3 368 760, −993 236, and –710 402 J mol^{-1}, respectively, under these conditions with the formulas as given in Table 7.1. Assuming ideality in the olivine and orthopyroxene solid solutions, calculate the activity of silica in both melts, and comment on the potential for silica to homogenize by diffusion.

14.23 If a magma at the ternary eutectic in Figure 10.21 mixes with a more primitive one of composition l_1, what mineral would crystallize first from the hybrid magma, and what would happen to crystals of the other minerals that had been forming at the eutectic?

14.24 **(a)** Plot a graph similar to the one in Figure 14.41 showing the variation in concentration caused by Rayleigh fractionation of elements having distribution coefficients of 0.45 and 2.0, and the variation in concentration caused by equilibrium fractionation of elements having distribution coefficients of 0.35 and 3.0.

(b) Phosphorus in Kilauean magmas is completely incompatible, so that the ratio $c^{l}_{P_2O_5}/c^{ol}_{P_2O_5}$ in any magma is a measure of the degree of crystallization involved in its derivation from the parental magma. Assuming that the concentrations of P_2O_5, TiO_2, and MgO in the parental Kilauean magma are 0.18%, 2.3%, and 12.0%, respectively, plot in the graph prepared for part (a) the relative concentrations (c^{l}/c^{ol}) of TiO_2 and MgO as a function of the fraction of the parental magma crystallized for analyses 7 to 15 in Table 14.1.

(c) Is Ti as incompatible as P in these rocks? Explain why analysis 15 plots at such a different position from the other analyses.

(d) Do the trends for TiO_2 and MgO fit a Rayleigh or equilibrium fractionation model best?

(e) The early-crystallizing ferromagnesian minerals have a bulk MgO mineral/liquid distribution coefficient of approximately 5; plagioclase, on the other hand, completely excludes Mg ($K = 0$). Based on the fit of the data in part (b), what proportion of plagioclase to ferromagnesian minerals would provide the best bulk distribution coefficient?

14.25 Calculate the chondrite-normalized abundances of REE in a basaltic magma that undergoes 60% equilibrium crystallization to form solids consisting of 80% olivine and 20% augite. Use the distribution coefficients given in Table 14.3 and assume that the basalt initially had a chondritic abundance of REE. Plot the results versus the atomic number of the REE.

14.26 Repeat Problem 14.25, but for 80% equilibrium crystallization to form solids consisting of 50% olivine, 20% augite, and 30% plagioclase.

15 Igneous rock associations

15.1 INTRODUCTION

Early in the development of petrology, it was recognized that certain rock types are commonly associated, whereas others never occur together. Moreover, the common associations were seen to correlate with certain geologic settings. Today, with the insight provided by plate tectonic theory, most igneous rocks can be assigned to particular plate tectonic environments, each of which has its own distinctive thermal regime, magma source region, and crustal stress pattern. But not all rock associations can be explained through plate tectonics. Some magmatism in the Archean and even the Proterozoic was different from that of Phanerozoic time, and distinctive rock associations were formed that were never again repeated in later times.

Seismic evidence indicates that the lithosphere and upper mantle are essentially solid, although a small amount of liquid may exist in the low velocity layer. The formation of large magma chambers and volcanic edifices is therefore a rare occurrence that requires special conditions. Yet, the majority of crustal rocks are of igneous origin, and thus these conditions must, on occasion, be met. The steady-state geotherm beneath a continent or ancient ocean floor (Section 1.6) does not come near the dry beginning of melting curve for mantle peridotite, at least not at the depths at which we believe magmas are generated. Therefore, either the geotherm must be raised or the beginning of melting curve lowered if magmas are to form. Both of these perturbations appear to occur most often as a result of convective movements in the mantle, which at the same time cause motion of the lithospheric plates. For this reason, magmatism and plate motion are directly related.

The greatest production of igneous rocks is at mid-ocean ridges where steepened geothermal gradients resulting from upward-convecting mantle and decompression cause melting at shallow depths. The degree of partial melting under these conditions is moderately high, so that incompatible elements that enter the first-formed liquid are diluted by the additional melt, and magmas of tholeiitic composition are formed. Despite the enormous production of igneous rocks in this environment, most are subducted and do not enter the geologic record preserved on continents. The relatively few instances where ocean floor has been obducted onto continents to form the ophiolite suite provide rare opportunities to study these ocean floor rocks.

The second most productive region of magmas is at convergent plate boundaries. Here, subduction of cold lithosphere lowers geothermal gradients, but melting is caused by the fluxing effect of water liberated from subducted oceanic crust. The magmas are of calcalkali type and have compositions that are determined by the depth at which the fluxing water is released, which is controlled by the rate and angle of subduction.

Where convecting mantle rises beneath a continent prior to its rifting apart to form ocean floor, geothermal gradients are not as steep as those beneath mid-ocean ridges, and melting occurs at greater depth and involves a smaller degree of partial melting. This produces alkaline magmas with higher concentrations of incompatible elements. These magmas and their differentiation products therefore characterize rift valleys on continents. Similar magmas can form in mantle plumes that rise at locations not necessarily related to plate boundaries. Movement of the lithosphere over these “hot spots” produces trails of igneous rocks. If the plume is particularly hot, higher degrees of partial melting can produce transitional or tholeiitic rocks.

15.2 IGNEOUS ROCKS OF OCEANIC REGIONS

Beneath a thin veneer of sediment, the ocean floors, which comprise approximately 65% of the Earth’s surface, are composed almost entirely of basaltic lavas. This is the most extensive occurrence of igneous rocks on Earth. While ocean floor is continually being created along oceanic ridges, equivalent volumes of ocean floor are being destroyed at subduction zones (Fig. 15.1). This creation and destruction of ocean floor is an extremely active process, as indicated by the age of ocean floors, which are Mesozoic or younger. More than 10 km^3 are erupted per year from the 65 000-km-long mid-ocean ridge system. This is an order of magnitude more than is erupted subaerially. The enormous production of igneous rocks within oceans accounts for approximately 60% of all heat lost from the Earth.

Although most oceanic rocks form at mid-ocean ridges, some form within plates above long-lived “hot spots” – regions of thermal upwelling – whose positions appear to have remained relatively constant for long periods of time, despite the movement of overlying lithospheric plates (Wilson, 1963, 1973; Sleep, 1992; see Tarduno (2007) for evidence of movement). Morgan (1971) proposed that plumes rose from the core mantle boundary. Although seismic evidence supporting this interpretation has been elusive,

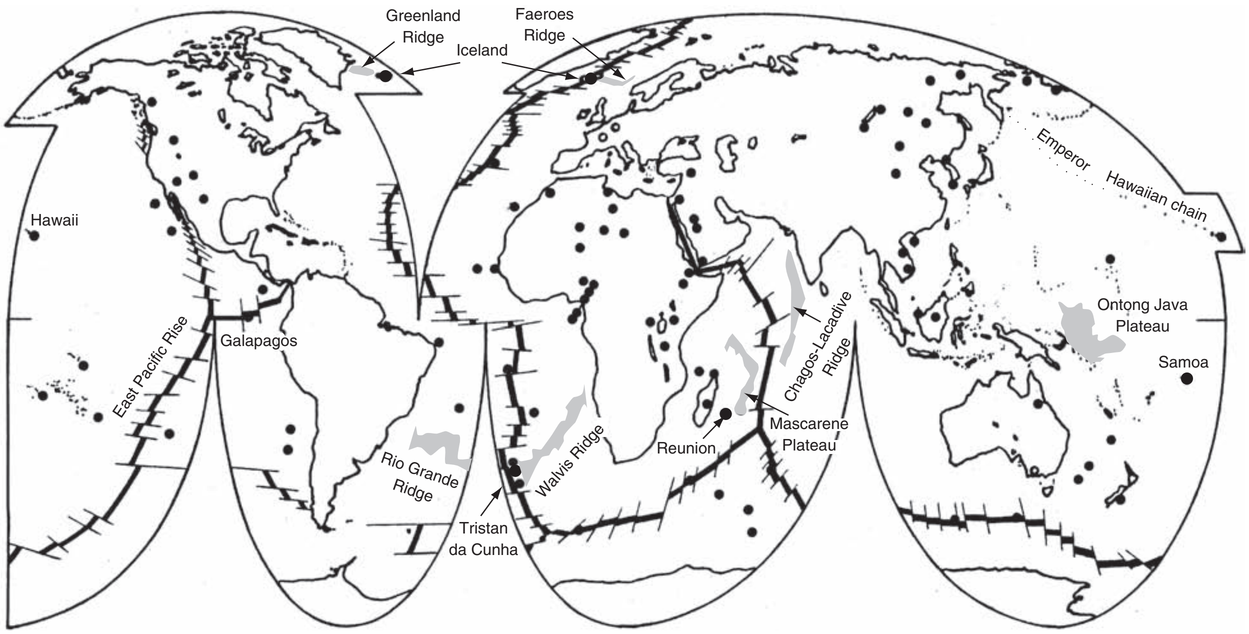


Fig. 15.1 Mid-ocean ridge system where tholeiitic basalts (MORB) are being generated at a rate of more than 10 km^{3}a^{-1}. Other centers of non-plate-margin volcanism (hot spots) are shown with black dots. Aseismic ridges and oceanic plateaus are shown in gray. (Data mainly from BVSP, 1981.)

some recent findings do indicate a deep source (Montelli *et al*., 2004; Nolet *et al*., 2007). A low-velocity anomaly can be traced beneath Iceland down to the transition zone, but no deeper. Beneath Hawaii, however, a low-velocity anomaly can be traced to the base of the mantle. However, other geologists claim that these thermal anomalies have a very shallow source (e.g. Anderson, 2005; Foulger *et al*., 2005). This debate will undoubtedly continue for some years to come (Great Plume Debate, 2007; see also www.mantleplumes.org). Some ocean floor rocks are generated on aseismic ridges, some of which may be related to hot spots on spreading axes.

15.2.1 Mid-ocean ridge basalts (MORBs)

The crest of most of the mid-ocean ridge system is marked by a ~35-km-wide rift valley, along the center of which is a prominent fracture zone (Hekinian, 1982). Lava erupting from this fracture either floods the rift valley or builds small volcanic edifices along the fracture (Fig. 4.3). Some lava is erupted from faults that bound the rift valley. Along most of their length, oceanic ridges are segmented by transform faults. Within individual segments, the surface character of lava flows commonly changes from relatively smooth near the center, through lobate, to pillowed near the end of segments. This is thought to correspond to decreases in temperature and magma flux away from the center of segments.

Almost all rocks created at mid-ocean ridges are basalts that have a small but significant range of composition. They are classified under the general name of *mid-ocean ridge basalt* (MORB) and are olivine tholeiites. They are either aphyric, or contain phenocrysts of olivine (± Mg–Cr spinel) or plagioclase, or both; rarely do they contain augite phenocrysts. Olivine ranges in composition from Fo$_{91}$ to Fo$_{65}$ and plagioclase from An$_{88}$ to An$_{40}$. The olivine–phyric lavas tend to erupt from the axial fracture zone, whereas the plagioclase–phyric ones occur more toward the edge of the rift valleys, suggesting the existence of zoned magma chambers beneath the rifts (Fig. 15.4). The olivine-bearing variety has a higher Mg number than does the plagioclase one, indicating that differentiation and not simply crystal accumulation is involved in their genesis.

MORBs are thought to form by the passive upwelling of mantle beneath ocean spreading axes (McKenzie, 1967). Melting occurs because of decompression with no need for addition of heat (isenthalpic), but once melting occurs, the rise of magma into the crust steepens geothermal gradients. In Section 10.23, we saw that McKenzie and Bickle (1988) introduced the concept of potential temperature to describe the temperature at which the extrapolated mantle adiabat would intersect the Earth's surface if there were no melting (Fig. 10.39). The mantle rising beneath a spreading axis would follow this adiabat until partial melting started to form MORB magma, at which point the temperature would fall below the adiabat due to the latent heat of fusion. McKenzie and Bickle estimated that the potential temperature for magma beneath spreading axes away from hot spots is ~1280 °C. More recent studies indicate that the potential temperature is at least 100 °C higher than this, possibly ~1450 °C. Putirka (2005) has used the maximum forsterite content of olivine phenocrysts (Fo$_{91.5}$) and the composition of coexisting MORB to obtain a global mean potential temperature of 1454 ± 78 °C beneath mid-ocean ridges.

Because MORBs have a restricted compositional range, their source must be relatively constant, and their degree of fractionation must be similar. Most have a Mg number of ~60, indicating that they have differentiated only slightly from a primitive magma, which would have a Mg number of 70. Also, the absence of a europium anomaly in the chondrite-normalized rare earth elements (REE) pattern of most MORBs indicates that no significant amount of plagioclase could have been fractionated (Fig. 14.42). When unaltered, MORBs have low contents of alkalis and other incompatible elements, and the light REE are strongly depleted (Fig. 14.43). Concentrations of compatible elements such as Cr and Ni, by contrast, are high. As indicated in Section 14.12, the source from which MORBs are derived must have been depleted in incompatible elements and enriched in compatible ones by repeated partial melting episodes. The lack of a negative slope to the chondrite-normalized REE pattern (Fig. 14.43) has been interpreted to indicate that garnet was not present in the source, which means that magmas came from depths of less than 65 km (Chapter 23). However, the detailed seismic study of the East Pacific Rise (MELT Seismic Team, 1998) shows that incipient melting occurs at depths of ~140 km, but most of the melt occurs at depths between 70 and 20 km, which probably explains the absence of a garnet signature in the REE plots of MORB. Toward topographic highs on ocean ridges, especially those capped by an oceanic island, MORBs are less depleted in incompatible elements, have a flat REE pattern, and resemble more the tholeiites on ocean islands. In these regions, magma may be tapping a deeper, less depleted source. The largest variation in the composition of MORBs is due to alteration by circulating ocean water, which increases the alkali content, in particular sodium, and converts plagioclase to albite and ferromagnesian minerals to chlorite. The altered rock, which is known as *spilite*, is discussed under ophiolites below.

Another minor rock type that occurs among these basalts is a plagiogranite or rhyodacite. This is believed to be formed as a final differentiation product of MORB magma. In coarsely crystallized pillows of MORB, the residual mesostasis is seen to consist of immiscible globules of iron-rich glass in silica-rich glass (Philpotts, 1982). The silica-rich glass matches closely the composition of plagiogranite. Dixon and Rutherford (1979) have shown through experiments that this immiscible silica-rich liquid invariably develops after MORB is crystallized to about 95%.

Variations in the composition of MORB can be explained broadly in terms of fractionation of olivine and, in some cases, plagioclase, but in detail it breaks down (BVSP, 1981). Because the compositions of olivine and plagioclase are so different from those of MORB, fractionation of these minerals brings about significant changes in the concentration of major elements in the fractionated liquids. The compositional range for major elements in MORB, however, is quite small compared with the range that would be predicted for the degree of fractionation indicated by changes in the concentration of incompatible elements. But even the degree of fractionation based on incompatible elements often cannot be made to match that based on the decrease in the concentration of highly compatible elements. Explaining the differentiation in terms of a single parental magma from which olivine, spinel, and plagioclase (the normal phenocrystic phases) have separated is, consequently, very difficult, if not impossible, to do. The problem is alleviated somewhat by having several parental magmas, perhaps in separate magma chambers, that mix at different stages of fractionation (Rhodes *et al*., 1979). As pointed out in Section 13.4, isotopic studies reveal that MORBs can indeed have several different sources. Normal MORB (N-type MORB) is derived from a depleted mantle that can be divided into two possible sources, one having a slightly higher $^{206}Pb/^{204}Pb$ than the other. MORB that is erupted from mid-ocean ridges near hot spots shows involvement of sources that are less depleted in incompatible elements and have higher $^{207}Pb/^{204}Pb$ and $^{208}Pb/^{204}Pb$ and lower $^{143}Nd/^{144}Nd$ and relatively higher $^{87}Sr/^{86}Sr$. However, involvement of these different sources does not account for the fractionation trends that MORBs follow.

Only when augite is added as a fractionating phase can the MORB compositional trends be satisfactorily explained. Because augite's composition is much closer to that of MORB, its fractionation can cause enrichment in incompatible elements while having only minor effect on the other elements. There is little doubt that augite must play a role in fractionating MORB, despite its rarity as a phenocrystic phase. Experiments indicate that with increasing pressure, augite becomes a primary phase on the liquidus of MORB (see Problem 14.3).

Rhodes *et al*. (1979) have pointed out that magma mixing could also be responsible for removing augite as a primary phase. In the discussion of the system forsterite–anorthite–SiO_2 (Fig. 14.40), mixing of an evolved, multiply saturated magma of composition 2 with a primary magma of composition X was shown to cause resorption of plagioclase. The phase diagram for the system forsterite–augite–SiO_2 is almost identical; the augite field can simply be substituted for the anorthite field. Mixing of magma 2 with magma X would therefore bring about resorption of augite. Primary magma ($Mg' = 70$) is rarely erupted along ridges, but some may be intruded into the underlying magma chambers, where it would mix with evolved magma. The composition of the evolved magma may be determined by fractionation of augite, but when the primary magma mixes with it, augite disappears from the liquidus of the hybrid magma, which could then be erupted as MORB. This continual replenishing with primary magma could also explain why MORB does not become highly evolved. In the section on ophiolites we will see that beneath ocean ridges, there are indeed magma chambers in which clinopyroxene forms cumulate rocks.

Although the MELTS experiment has identified the presence of liquid beneath the East Pacific Rise extending to depths of more than 100 km, the amount of melt is only a few percent, at the most. Large magma chambers have not been identified. The melt may occur on grain boundaries or in small fractures, which could contribute to seismic anisotropy.

A thin ~100-m-thick by <1-km-wide axial magma chamber has been identified at depths of 1–2 km beneath the ocean floor along the East Pacific Rise (Hussenoeder *et al*., 1996). A similar lens has been found beneath the Galapagos spreading center at depths of 1 to 4.5 km below the ocean floor (Blacic *et al*., 2004). These lenses, which contain a high percentage of crystals, appear to be the magma chambers from which MORBs are erupted (Problem 3.1).

15.2.2 Intraplate oceanic islands

Ocean floors are low areas on the surface of the Earth because they are underlain by oceanic crust, which, on average, is denser than continental crust. An oceanic island, especially one with considerable topographic elevation, is therefore a remarkable feature that normally requires active volcanism to survive. Some measure of the enormous amount of construction required to build one of these islands can be obtained by comparing the heights of Mauna Loa on the Island of Hawaii, which is the largest volcanic edifice on Earth, and Mount Everest, which is the highest tectonically constructed feature on Earth. Mauna Loa towers 10 km above the ocean floor, whereas Mount Everest rises only 8.9 km above sea level.

Some oceanic islands are formed by volcanoes developed at convergent plate boundaries (Section 15.3), but most of the remainder are formed at "hot spots", which many people believe are formed above rising mantle plumes (Campbell, 2007) whose positions have remained relatively fixed over extended periods of time. The adiabatic rise of the mantle within a plume would produce a partial melt, which would buoyantly rise through the lithosphere to form a volcano on the ocean floor. If the plume is sufficiently active, the volcano grows and eventually becomes an island; otherwise, it remains a seamount. The thermal input from the mantle plume lowers densities and allows the ocean floor to stand higher than it otherwise would. Because the lithosphere is continually moving relative to plumes, volcanoes (and islands) tend to develop in chains or ridges whose orientation defines the direction of plate motion. Once plate motion moves a volcano off a hot spot, the volcano becomes extinct. With cooling and sinking of the lithosphere, the island is eventually planed off at wave base to form a truncated seamount or atoll if the growth of fringing coral reefs is able to keep pace with the rate of subsidence. The Hawaiian Islands–Emperor Seamount chain provides the best example of such a linear array of hot-spot-generated volcanoes. Hot spots are not restricted to oceanic regions, but most occur there, and many of these are located at divergent plate boundaries (Fig. 15.1).

Mantle plumes provide an attractive explanation for many features associated with ocean-island volcanism, but their acceptance is far from universal (Great Plume Debate, 2007). Familiar examples of plumes are thunderheads or smoke rising from tall smoke stacks on a calm day. As a plume rises, it forms a mushroom-shaped head. Clouds of ash rising from volcanoes take on this form (Fig. 4.20), as described by Pliny the Younger when he likened the shape of the ash plume rising from Vesuvius in AD 79 to the pines of Rome (Fig. 4.36(A)). Much is known about the fluid dynamics of plumes from theoretical and experimental studies. Modeling of mantle plumes indicates that they too will mushroom on reaching the top of the mantle. Although a plume rising through the mantle has a small diameter (this is probably why they are difficult to resolve seismically at depth), on reaching the top of the mantle, it mushrooms to a diameter of >1000 km. A plume head's diameter (D) is given by

$$D = Q^{\frac{1}{5}}(v/g\alpha\Delta T)^{1/5}K^{2/5}z^{3/5} \qquad (15.1)$$

where Q is the buoyancy flux, v the kinematic viscosity of the lower mantle, α the coefficient of thermal expansion, ΔT the temperature difference between the plume and the adjacent mantle (commonly referred to as the *excess temperature*), K the thermal conductivity, and z the distance through which the plume travels (Campbell, 2007). The dominant term in this equation is the distance z, because it is raised to the power $^3/_5$. If $\Delta T = 300$ °C, and for reasonable values of the other variables, a plume rising from the core–mantle boundary should have a diameter of ~1000 km on reaching the top of the mantle, and this would flatten to a disk with a diameter of ~2000 km. A plume can therefore be expected to affect a large area. Based on the occurrence of flood basalts in Greenland and the northwest coast of the British Isles, the postulated Icelandic plume would have affected such an area. Large plume heads have been hypothesized to produce large igneous provinces (LIP), which are discussed in Section 15.4.

Some oceanic islands appear to have formed from less vigorous thermal events, forming smaller islands, and shorter island chains that may not show the same rate of age progression as chains formed over major hot spots. The source of these magmas appears not to have remained stationary but to have been deflected slightly by plate motion. Still other oceanic islands and seamounts are connected to no chain of islands and appear to be isolated volcanic features. Such islands have led to hypotheses of *thermochemical* plumes and sinking of dense metamorphosed basalt. A thermochemical plume is visualized to rise from the core–mantle boundary, bringing with it dense material (possibly recycled ocean floor from the D″ layer) which causes it to stall at the 670 km discontinuity, where it spawns a secondary plume with less dense components rising through the upper mantle (Farnetani and Samuel, 2005). The thermal effect of the secondary plume is less than that of a plume that penetrates both the lower and upper mantle, and consequently, a smaller island chain would be produced. Individual volcanic islands might be formed by dense metamorphosed basalt sinking into the mantle and being remelted. When basalt is metamorphosed at high pressure, it is converted to an extremely dense rock known as eclogite, which consists of omphacite (Na- and Al-bearing pyroxene) and pyrope garnet (Section 16.4). Because of its density, eclogite could sink into the mantle, but because of its low melting point, it would soon melt to form basalt.

The volcanic rocks of ocean islands differ from MORB in having a wider range of composition, which in part is due to differentiation within near-surface magma chambers, but also to the composition of primary magmas, which contain higher concentrations of incompatible elements (Section 14.12). This difference is thought to be due to the magmas of oceanic islands having deeper sources than MORB; that is, the mantle rising in the plumes is not as depleted as the shallow mantle beneath mid-ocean ridges. A deep source is also required to account for the fixed position of hot spots beneath the moving lithospheric plates.

The rocks of oceanic islands are commonly both tholeiitic and alkaline (mainly Na), with tholeiitic rocks typically being early and alkaline ones late in the development of an island. Small oceanic islands tend to have only alkaline rocks, some of which are ultra-alkaline and silica-poor; in this respect, they resemble the ultra-alkaline rocks associated with rifting and mantle plumes on continents (Section 15.7). Despite the greater differentiation of oceanic-island rocks compared with MORB, basaltic compositions still predominate. Table 14.1, for example, gives the typical range of tholeiitic rocks erupted from Kilauea. On Iceland, tholeiitic magmas differentiate all the way to rhyolite, which constitutes as much as 8% of the rocks. The most common differentiate of the alkaline magmas is trachyte, and in most differentiated suites intermediate rocks between basalt and trachyte are scarce.

The earliest crystallizing mineral in the tholeiitic lavas is a Cr–Al spinel. This is followed by olivine, which occurs only as phenocrysts and has a reaction relation with liquid to produce a Ca-poor pyroxene, either orthopyroxene or pigeonite. Olivine is consequently restricted to relatively magnesian compositions (Fo_{88-80}). Plagioclase and augite crystallize next and both can form small phenocrysts, but these are not as common as in the alkaline rocks. Because augite coexists with a Ca-poor pyroxene, it has the subcalcic composition ($Ca_{40-35}Mg_{50-42}Fe_{10.23}$) required of an augite on the pyroxene solvus (Fig. 10.32).

The tholeiitic rocks of ocean islands have higher concentrations of incompatible elements than do MORBs. Their chondrite-normalized REE pattern does not show the depletion in light REE. Indeed, the REE pattern of most ocean-island tholeiites is similar to that of alkali olivine basalts of ocean islands (Fig. 14.43) except that the overall concentration is slightly less in tholeiites. The steep negative slope to the REE pattern indicates that ocean-island tholeiites are derived from mantle that is far less depleted than that from which MORBs are derived, and that it probably contains garnet. This would indicate a depth of origin greater than 65 km (Chapter 23). Many of the volcanic rocks from large hot spots, such as Hawaii and Iceland, exhibit elevated $^3He/^4He$. This has been interpreted to indicate the presence of larger amounts of the primordial ^3He, which is thought to have come from a deep source and thus be consistent with the plume hypothesis. Anderson (1998), however, claims that the high $^3He/^4He$ could result from low amounts of ^4He, which is produced by radioactive decay of U and Th and thus would indicate a source depleted in these elements rather than enriched in ^3He.

Alkali olivine basalt is normally the most abundant and most primitive member of the alkaline rocks. It differs from tholeiitic olivine basalt in having a lower activity of silica, which causes it to be nepheline normative. Some actually contain modal nepheline, in which case the rock is called a *basanite*. Spinel and olivine form the earliest phenocrysts, followed by plagioclase and augite. Olivine covers a much wider range of composition than in the tholeiites (Fo_{87-34}). This is because olivine in alkaline rocks does not react to form a Ca-poor pyroxene, and thus it can continue crystallizing to late stages. Ca-poor pyroxenes never form in alkaline rocks. These rocks only ever contain a single pyroxene, augite, which commonly contains a significant amount of Ti ($CaTiAl_2O_6$), which gives it a distinct pink or violet color in thin section. Augite may also contain significant amounts of Calcium-Tschermak's molecule [$CaAl(AlSi)O_6$]. Both of these substitutions result from the low silica activities of these melts, as does the lack of a Ca-poor pyroxene (Ca-poor Pyx = Ol + SiO_2).

Stemming from alkali basalt is a series of differentiated rocks, in which, with decreasing color index, the plagioclase becomes progressively more sodic (Fig. 6.13(A)). The members of this series and their plagioclase compositions are: alkali basalt (labradorite), hawaiite (andesine), mugearite (oligoclase), benmoreite (anorthoclase), and trachyte (anorthoclase). The series is critically undersaturated in silica at its mafic end and may remain so or become oversaturated toward its evolved end. Trachyte is normally the most differentiated member of the series, but in more extreme cases of differentiation phonolite (nepheline-bearing), pantellerite (Na-rhyolite), or comendite (K-rhyolite) may form. The development of both critically undersaturated and oversaturated melts from a common parental magma cannot be satisfactorily explained by fractional crystallization (Section 10.6 and Fig. 10.5). If both melts are indeed developed from a common parent, some process of differentiation affecting only the liquid must be involved, such as immiscibility, Soret diffusion, or vapor transport.

In most alkaline series, intermediate rocks are scarce, and on many ocean islands, basalt and trachyte are the only members of the series present. This lack of intermediate volcanic rocks is an example of the Daly Gap discussed in Section 14.7. It has been suggested that trachyte is not even derived from the basalt but is a separate product of partial melting in the mantle.

The trace elements of the alkaline rocks are very different from those of MORB. Alkaline rocks are much richer in incompatible elements and show no signs of having been derived from the strongly depleted mantle from which MORB has come. Their steep chondrite-normalized REE pattern indicates that they are derived from a source in which garnet is stable (>65 km). Alkaline rocks contain higher concentrations of incompatible elements than do the tholeiites of oceanic islands. This difference cannot be accounted for by differentiation of the alkaline magmas, because the contents of compatible elements, such as Cr and Ni, are comparable to those in the tholeiites. Fractional crystallization of olivine or pyroxene would rapidly decrease the concentration of these elements. The high concentration of incompatible elements is therefore interpreted to result from small degrees of partial melting of a relatively undepleted mantle (Problem 15.1 and 15.2). The smaller degree of partial melting would also explain why alkaline rocks are less abundant than tholeiitic ones and why they tend to form only small oceanic islands.

On some islands, extremely undersaturated alkaline rocks occur, in which the silica activity is so low that nepheline and melilite form instead of plagioclase. These are the *nephelinites*. They are typically younger than associated tholeiites or alkali basalts and commonly erupt explosively to form small cinder cones. They have very much higher concentrations of incompatible elements and a slightly steeper chondrite-normalized REE pattern than have alkali basalts. They are consequently probably formed by still smaller degrees of partial melting. Nephelinites commonly contain mantle nodules whose mineralogy indicates they were derived from depths of at least 100 km. These rocks are more common in alkaline complexes associated with rifts on continents and are discussed further in Section 15.6.

If rocks of oceanic islands are formed from partial melting in mantle plumes, their sources should have higher potential temperatures than normal mantle; that is, they should exhibit excess temperatures. Putirka (2005) used the compositions of the most forsteritic olivines and coexisting liquids and estimates of the degree of partial melting to obtain potential temperatures beneath Hawaii and Samoa of 1722 °C and beneath Iceland of 1616 °C. These temperatures are 200 to 300 °C above the mean potential temperature of 1454 °C for mantle beneath spreading axes, which is consistent with the mantle plume hypothesis. However, Falloon *et al.* (2007) have used the composition of olivine in MORB and ocean island basalts to conclude that the potential temperatures of their sources were similar (see the Great Plume Debate, 2007).

15.2.3 Aseismic ridges

A small percentage of ocean floor rocks are formed along aseismic ridges, such as the Greenland–Iceland and Iceland–Faeroe ridges to the west and east of the Iceland hot spot respectively (Hekinian, 1982) or the Rio Grande and Walvis ridges extending west and east respectively from the Tristan da Cunha hot spot in the South Atlantic (White and McKenzie, 1989) (Fig. 15.1). It is not clear whether these ridges formed over hot spots or developed along fracture zones–leaky transforms. According to the plume hypothesis they were formed as the plume head spread laterally.

The rocks from these ridges resemble those of oceanic islands rather than MORB. They have a wide range of composition and have undergone significant differentiation. Like the rocks of oceanic islands, their trace element content indicates derivation of magmas from a mantle that has not been strongly depleted.

Finally, large areas of the ocean floor are underlain by thicker than normal oceanic crust (up to 30 km) that forms *oceanic plateaus*, which stand 2 to 3 km above the surrounding ocean floor and have areas >200 000 km^2, the largest being the Ontong Java plateau in the western Pacific (Fig. 15.1). The origin of these plateaus remains a puzzle (Kerr and Mahoney, 2007). Despite their enormous size, volcanism never built structures that extended above sea level. Once volcanism ceased, cooling would be expected to have caused subsidence, but they remain elevated above the surrounding sea floor. Geochemically, they are distinct from MORB and oceanic island basalts. Many of the basalts have flat REE patterns similar to those of island arc basalts, but their isotopic ratios of Nd, Pb, Hf, and Sr are similar to ocean island basalts. Some rocks show isotopic evidence of contamination with continental crustal rocks (elevated ^{87}Sr / ^{86}Sr and negative $\varepsilon_{\mathrm{Nd}}$). In the Caribbean Plateau, some rocks have exceptionally high MgO and approach komatiites in composition. Much more needs to be known about oceanic plateaus before reasonable models for their origin can be constructed.

15.2.4 Ophiolite suites

Despite the large number of studies in recent years of ocean floor rocks, the sampling of such a large fraction of Earth's surface (65%) is still extremely small. Furthermore, the sampling, by necessity, is strongly biased toward material within the first few meters of the ocean floor; it certainly is not representative of the entire oceanic crust. Deep drilling and exposures of rock in fracture zones where there is significant relief has shown that the pillowed basalts on the ocean floor are underlain by intrusive rocks, which like the basalts, have undergone varying degrees of hydrothermal alteration (Brunelli *et al.*, 2006).

Seismic velocities reveal that oceanic crust has a relatively simple three-tiered structure. Layer 1, which has the lowest velocity and averages 0.3 km thick, corresponds to the sedimentary veneer covering the lavas. Layer 2, with a compressional wave velocity of ~5 km s^{-1} and average thickness of 1.4 km, has been shown by drilling to be pillowed basalt. Layer 3, with a compressional wave velocity of 6.7 km s^{-1}, is ~5 km thick, and overlies the Mohorovičić (M) discontinuity, below which the seismic velocity abruptly increases to 8.1 km s^{-1}. Layer 3 is probably composed of intrusive rocks. The mantle immediately below the M discontinuity exhibits seismic anisotropy, with the fast shear wave component being oriented perpendicular to the spreading axis, which is attributed to a preferred crystallographic alignment of the *a*-axis of olivine crystals (Wolfe and Solomon, 1998).

In rare cases, the oceanic lithosphere is exposed along transform faults. For example the Vema transform, which offsets the Mid-Atlantic Ridge at 11° N by 310 km, exposes a section extending from ocean floor pillow basalts down through sheeted dikes and layered gabbros into peridotite (Brunelli *et al.*, 2006). Such sections allow the three-tiered seismic division of the oceanic crust to be correlated with actual rock types.

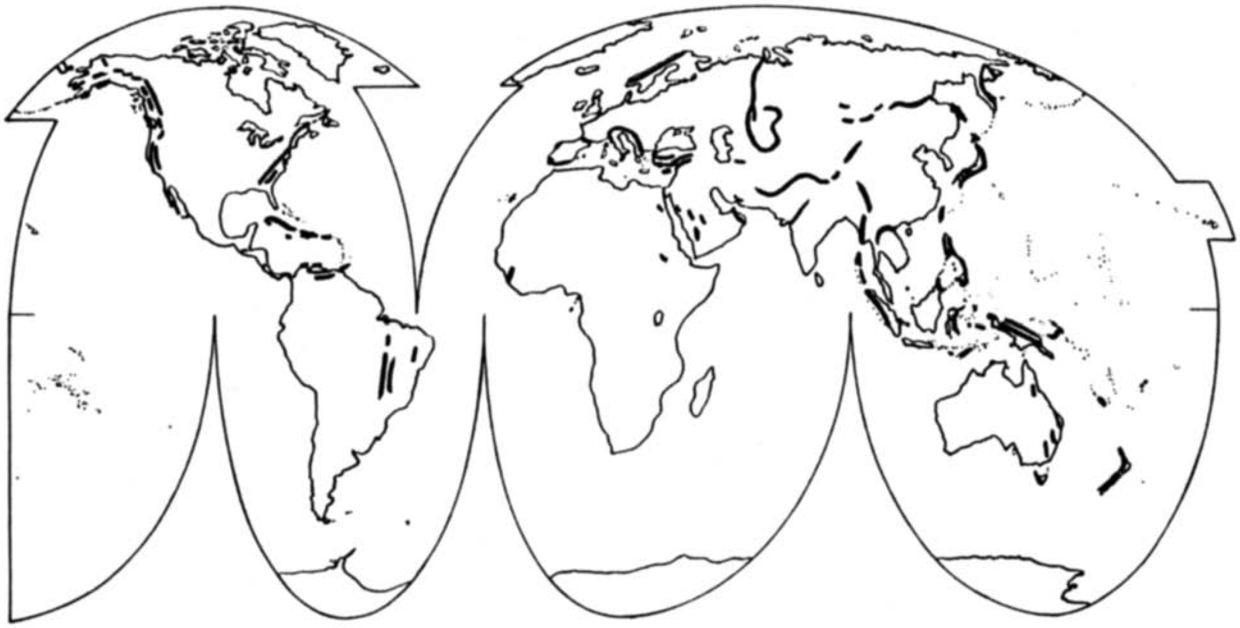


Fig. 15.2 Distribution of ophiolite suites. (Data from Coleman, 1977.)

Although most oceanic crust is destroyed by subduction at convergent plate boundaries, some is preserved by being thrust onto continents (*obduction*) to form what is known as the *ophiolite* suite (Coleman, 1977). This suite has three distinct members: serpentinized ultramafic rocks, hydrothermally altered pillowed basalts (spilites) and diabases, and pelagic sediment, which is characterized by radiolarian ribbon cherts. The pelagic sediments and pillow lavas have long been recognized as indicating a deep-water origin. Initially, however, ophiolites were thought to be autochthonous (formed *in situ*) eugeosynclinal rocks. The association of the three types of rock is so common in most Phanerozoic mountain belts (Fig. 15.2) that it became known as the "Steinmann trinity," named after the petrologist who first drew attention to the association. Ophiolites, however, exhibit puzzling field relations, such as the lack of contact metamorphism around the ultramafic rocks, which were clearly formed at high temperature. These ultramafic bodies were called *alpine peridotites* to distinguish them from peridotites that do have metamorphic aureoles. Not until the advent of plate tectonics was the ophiolite suite recognized to be allochthonous (not formed *in situ*) and probably composed of rocks that were formed at mid-ocean ridges.

Because ophiolites are tectonically emplaced, many of the contacts between its members are faulted, and much of the suite may be missing. Nonetheless, a number of well-preserved examples, such as those of Oman, Newfoundland, Papua, and the Troodos Massif in Cyprus (Fig. 15.4), show that the ophiolite suite has a characteristic stratigraphy (Fig. 15.3), which consists of a lower relatively homogeneous ultramafic unit, overlain by a sequence of layered ultramafic and mafic rocks, which in turn may be overlain by a sheeted dike complex, which feeds overlying pillow lavas; these rocks are in turn overlain by the sedimentary rocks.

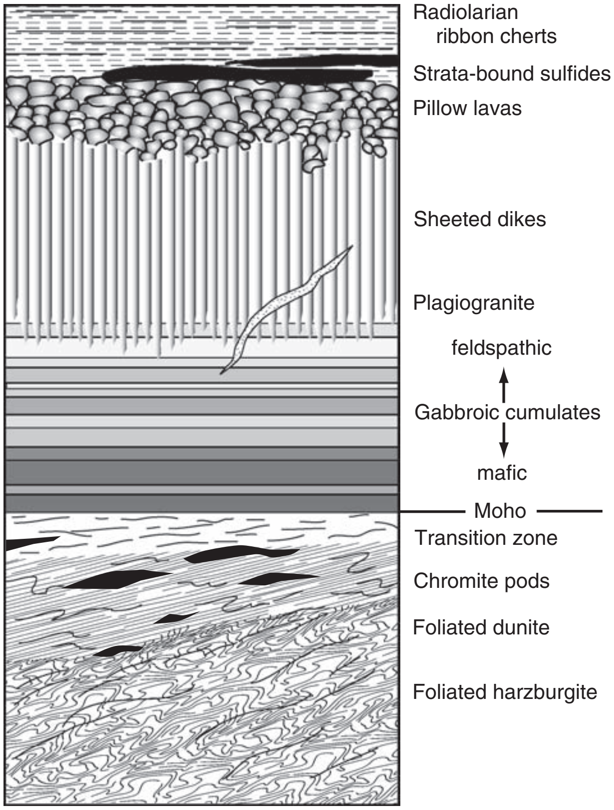


Fig. 15.3 Schematic section through an ophiolite suite. Chilled margins on sheeted dikes are indicated by darker shading; spreading axis is to left. See text for discussion.

The lowest member of the ophiolite suite is invariably a strongly deformed, serpentinized harzburgite (ol + opx). Primary layering, formed by variations in the abundance of olivine or spinel, is folded and cut by a deformational

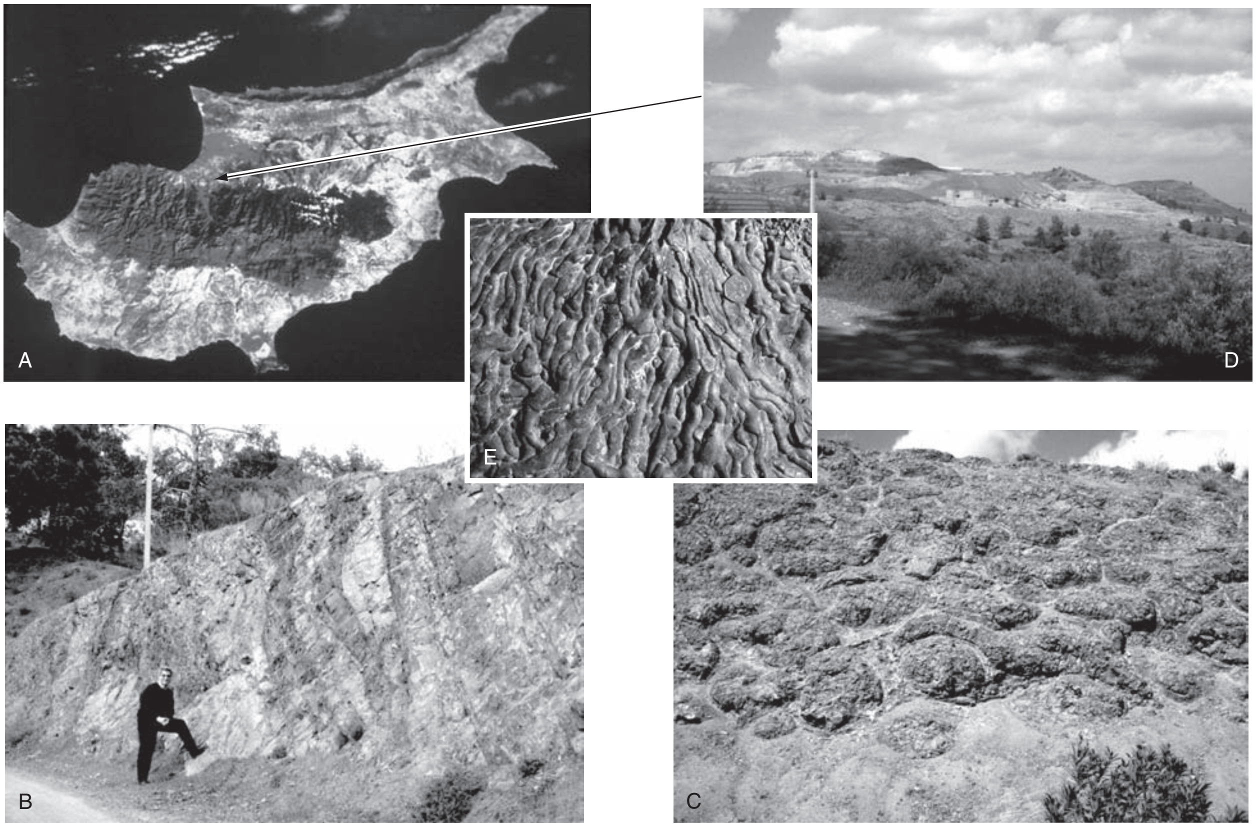


Fig. 15.4 (**A**) A section through the oceanic crust and upper mantle is exposed in the Troodos ophiolite suite of Cyprus, which is the dark area in this NASA satellite image. The ophiolite suite is surrounded by overlying light-colored marine sediments. (**B**) Sheeted dikes in the Troodos; author Anthony Philpotts for scale. Many of the dikes are hydrothermally altered. (**C**) Pillow lava on the north side of the Trodoos. (**D**) Copper mine at Skouriotissa (translates to *place of slag*). Copper has been mined here since 4000 BC. The deposits are located at the boundary between the pillow lavas and overlying sediments. (**E**) Roman slag at Skouriotissa (coin for scale).

foliation. Toward the top, the foliation becomes parallel to the layering in the overlying rocks, which is thought to have originally been horizontal. The harzburgite also becomes richer in olivine toward the top, typically grading into a dunite. In addition, the harzburgite can be cut by veins of dunite (Fig. 3.12), which can be up to tens of meters wide (Braun and Kelemen, 2002). Irregular pod-like bodies of chromite up to 1 km in length are common in the dunite (Thayer, 1964). These rocks have been strongly deformed and recrystallized at high temperature. The olivine, which is unzoned and has a composition of Fo_{90-92}, has deformation kink bands and a strong preferred crystallographic orientation. The orthopyroxene is similarly unzoned and has a composition of En_{90}. Orthopyroxene never bears a reaction relation to olivine, indicating that these rocks must have equilibrated at pressures in excess of 0.5 GPa (Fig. 10.8). Spinel is variable in composition, generally becoming more chromium-rich toward the top. Clinopyroxene, which is a chrome diopside, never exceeds 5% and commonly is totally absent.

A thin transitional zone ranging from tens of meters to several hundred meters thick separates the lower ultramafic rocks from the overlying layered rocks (Nicolas and Prinzhofer, 1983). The deformation, which is characteristic of the underlying harzburgite, dies out in this transition zone. The amount of clinopyroxene and plagioclase increases upward through the zone. Thus dunite (ol) at the base passes up through wehrlite (ol + cpx), to clinopyroxenite and troctolite (plag + ol). Pods and lenses of chromitite are common. Near the base of the transition zone olivine grains exhibit kink bands, but the chrome diopside and plagioclase that grow poikilitically or interstitially show no signs of deformation. With the decrease in amount of olivine, the degree of deformation decreases.

The transition zone terminates abruptly against the overlying layered gabbroic rocks, in which all signs of penetrative deformation are lacking. These cumulates are generally mafic at the base and grade upward to more feldspathic rocks. Olivine becomes progressively more iron-rich, reaching Fo_{70} in the uppermost gabbros. Orthopyroxene is present but is far less abundant than in the underlying deformed ultramafic rocks. Diopsidic augite is the principal pyroxene. Toward the top of the layered gabbros, irregular intrusive bodies of plagiogranite occur. These are thought to be the final differentiation product of the gabbroic magma, which gave rise to the layered rocks and would be equivalent to the granophyres formed in differentiated tholeiitic sills on continents (Section 15.4).

In many but not all ophiolite suites, the layered gabbroic rocks are overlain by a sheeted dike complex (Fig. 15.4(B)). This unit consists of 100% dikes; that is, dikes intruding dikes, with no intervening screens of other rocks. Most dikes are less than 5 m wide. Many appear to have only one chilled margin. This is a consequence of repeated intrusions in the center of a single opening fissure. By studying statistically the asymmetry of chilled margins, the position relative to the main axis of spreading can be determined. The dikes preserve an ophitic texture, but most of their minerals have been hydrothermally altered. They were initially aphyric subalkaline diabases composed of plagioclase and augite; olivine and orthopyroxene are not reported. Although probably related to the underlying gabbros, they are not derived directly from them. The dikes cut the underlying gabbros but pinch out downward (Fig 15.3). The overlying lavas appear to have been fed by the dikes.

The pillow lavas are generally basaltic in composition but have been pervasively altered to mineral assemblages in the zeolite and greenschist metamorphic facies (Section 16.4). Thus the plagioclase has been converted to albite, and the mafic minerals have been transformed mainly to chlorite. These basaltic pillow lavas with albitic plagioclase were originally thought to have formed from a special type of magma and were given the name *spilite*. Today, however, their peculiar composition is recognized to be the product of hydrothermal alteration (Amstutz, 1974). Like the dikes, they initially were composed of plagioclase and augite; olivine is not common. The REE, which are less prone to change than most other elements during hydrothermal alteration and weathering, indicate that spilites are similar to MORB. Toward the top of the pillowed succession, which typically is about 1 km thick, nonpillowed keratophyre and quartz keratophyre (Chapter 6) lavas occur. Like the basalts, these too are pervasively altered to low-grade metamorphic rocks. The keratophyres are undoubtedly the volcanic equivalents of the plagiogranites.

Despite the pervasive low-grade metamorphism of the lavas and dikes, these rocks are not penetratively deformed. For this reason, the metamorphism is interpreted to result from hydrothermal alteration by circulating seawater near mid-ocean ridges. Exploration of mid-ocean ridges with submersibles has provided firsthand evidence of the effectiveness of the hydrothermal systems along spreading axes. Much of the cooling of the lavas and dikes near ridges is by circulating seawater. The entire volume of the oceans is estimated to cycle through these hydrothermal systems once every 8 million years. The circulating hot waters are certainly capable of converting the original basalt to spilite. Metals leached from basalts are known to precipitate from the hydrothermal solutions when they issue forth on the ocean floor as “black smokers” (Von Damm, 1990). Meter-high chimneys of iron, zinc, and copper sulfides are built up around the hyrothermal vents. Pyrite and pyrrhotite are the most common sulfides, with lesser amounts of sphalerite and wurtzite, and variable amounts of isocubanite, bornite and chalcopyrite. Strata-bound sulfide deposits in the pillow lavas and overlying sedimentary rocks of the ophiolite suite were probably formed by similar hydrothermal systems along ancient spreading axes. These sulfide deposits are of economic importance. Indeed, those associated with the ophiolites of the Troodos Massif in Cyprus played a pivotal role in the development of the Chalcolithic Period and later Bronze Age (Fig. 15.4(D)); the name Cyprus is derived from the ancient Greek word for copper (Kypros).

The ophiolite suite is important because it provides a section through the oceanic crust. Whether these sections correspond to normal oceanic crust developed at mid-ocean ridges or is crust developed in back-arc basins is not totally clear, but they do appear to be similar to exposures of oceanic crust in transforms, such as the Vema transform (Brunelli *et al*., 2006). MORB, however, typically crystallizes olivine phenocrysts followed by plagioclase phenocrysts; augite is a rare phenocrystic phase. Spilites and sheeted dike complexes, in contrast, commonly have augite phenocrysts, and olivine is rare. Nonetheless, the ophiolite suite is undoubtedly a slice of oceanic crust.

The boundary between the undeformed cumulate rocks and the underlying highly deformed ultramafic rocks is interpreted to have been the Mohorovičić discontinuity at the time these rocks were formed. Seismic velocities of the harzburgite match those of the mantle immediately below the M discontinuity, and the preferred orientation of olivine crystals in the harzburgite would certainly account for the seismic anisotropy of this mantle. The harzburgite is strongly depleted in incompatible elements and is interpreted to be the mantle from which MORB has been extracted; that is, it is a *restite*. A partial melt would initially be in equilibrium with olivine and orthopyroxene, but as it rises, the decreasing pressure expands the olivine field at the expense of orthopyroxene (Fig. 10.8), and consequently pyroxene dissolves as olivine precipitates. The dunite veins in harzburgite, consequently, are interpreted to be reaction products marking the channels along which basaltic melt exited the harzburgite (Braun and Kelemen, 2002). The transition zone between harzburgite and overlying gabbroic rocks is interpreted to be mantle from which extraction of melt was incomplete; the clinopyroxene and plagioclase, which do not exhibit the deformation shown by the coexisting olivine in this zone, formed from this melt (Nicolas and Prinzhofer, 1983).

Magma extracted from the harzburgite is believed to have collected at the base of the crust, where it formed the layered gabbroic rocks. The magma chamber may never have been very wide or very deep and was probably repeatedly replenished. Where ophiolite suites contain sheeted dike complexes, the magma chamber may have had a continuous existence. Because the dikes have no intervening screens of other rocks (Fig. 15.4(B)), they represent 100% extension. Structures in the underlying layered rocks, however, show no evidence of significant extension. The magma chamber may therefore have been stretched laterally as new magma was emplaced. A zoned magma chamber, such as that shown in Figure 14.10, could produce the layered rocks, with the more mafic cumulates forming near the base and the more feldspathic ones near the top. As the magma chamber was replenished, the

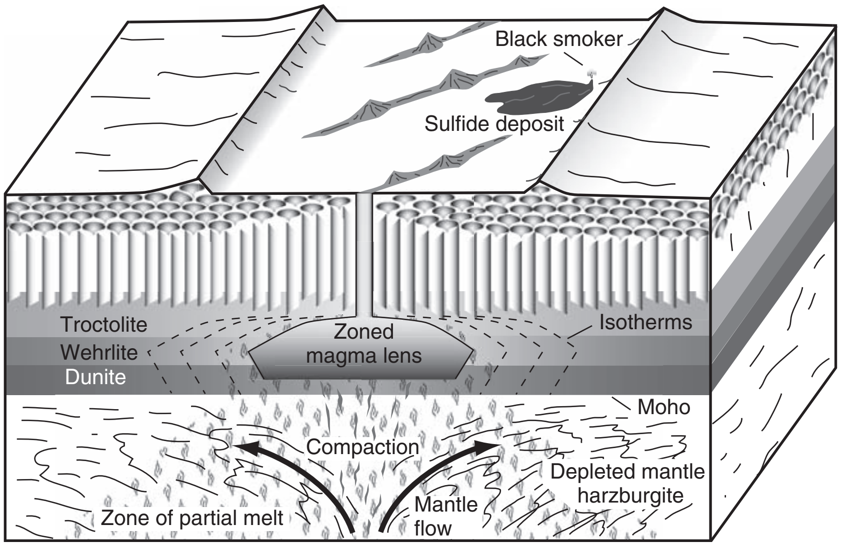


Fig. 15.5 Lens-like body of magma beneath a mid-ocean ridge. The lens is surrounded by crystal mush, which extends to depths of more than 100 km. Compaction of this crystal mush expels liquid upward to form the magma lens. The shape and volume of the lens may remain relatively constant; as divergent plate motion stretches the lens, its extremities cool and solidify, while new magma expelled from the underlying mush by compaction rises into the axial zone of the lens. Magma may escape from the lens through dikes, which commonly form en echelon sets. The resulting fissure eruptions may be limited to the area directly above the magma lens, but lateral flow in dikes may extend this activity. If the magma lens is zoned from more mafic at the bottom to more feldspathic at the top, layers of different types of cumulate can form simultaneously and continuously as the plates diverge and fresh magma is fed into the lens. Laterally extensive layers of cumulate (troctolite, wehrlite, dunite) could form in this way from a magma body that was never more than ~1 km wide. Plate divergence causes repeated splitting and injection of magma into axial dikes to form sheeted dikes with chilled margins that are symmetrically disposed on either side of the spreading axis.

sidewalls of the chamber would move outward, and magma on these walls would solidify to give the layered rocks (Fig. 15.5). At the same time, roof rocks would be pulled apart, providing the extension necessary for the emplacement of the sheeted dikes. A small, zoned magma chamber expanding and cooling laterally and being filled axially could eventually produce layered rocks of great lateral extent. Such an interpretation is consistent with the axial magma chambers that have been identified beneath the East Pacific Rise and Galapagos spreading axis (Hussenoeder *et al*., 1996; Blacic *et al*., 2004). These axial chambers are less than 1 km wide and ~100 m thick. They are lenses of liquid that occur 1 to 4 km below the ocean floor near the top of a pile of crystal mush that can extend to depths of more than 100 km. They are presumably replenished by liquid expelled from the underlying crystal mush by compaction. En echelon fissure systems on the ocean floor above active spreading axes (Fig. 15.5) have also been interpreted to indicate that the axial magma chambers may be no more than a couple of kilometers wide (Nelson, 1981; BVSP, 1981).

15.3 IGNEOUS ROCKS ASSOCIATED WITH CONVERGENT PLATE BOUNDARIES

Convergent plate boundaries rank next to mid-ocean ridges in their rate of production of igneous rocks. Unlike MORBs, most of which are doomed to a fate of subduction and disappearance, igneous rocks created at convergent boundaries survive by being accreted to continents. Indeed, the igneous processes at convergent plate boundaries can be thought of as those primarily responsible for the building of continents.

Igneous activity does not occur right at convergent boundaries, but is located in the overriding plate some distance behind the boundary. The distance from this boundary to the igneous rocks depends on the angle of subduction of the downgoing plate. The top of this cold, brittle slab is the source of many earthquakes, which defines the so-called *Benioff seismic zone*. The descent of the slab into the mantle can be traced by mapping the depth of earthquake foci. Igneous activity begins once the slab descends to a depth of ~100 km. On the surface, the onset of igneous activity is marked by arcuate chains of regularly spaced volcanoes, which in oceanic regions give rise to *island arcs*, and in orogenic regions to mountain chains topped by large strato volcanoes (Fig. 15.6).

The fact that igneous activity begins when the subducting plate goes below a depth of 100 km, regardless of the dip on the Benioff zone, indicates that the mechanism causing melting must be pressure sensitive. As was seen in Section 13.4, isotopic studies indicate that the primary magmas must be derived largely from the mantle wedge overlying the Benioff zone, rather than from the subducted oceanic slab itself. Why, then, is the 100-km depth to the Benioff zone so critical to generating magmas? The most likely explanation is that melting is triggered by the release of water from the downgoing slab (Ulmer, 2001). Ocean floor rocks are highly altered and hydrated. As they are subducted, metamorphic reactions release this water, which then rises into the overlying mantle to act as a flux in generating melts (Section 23.3).

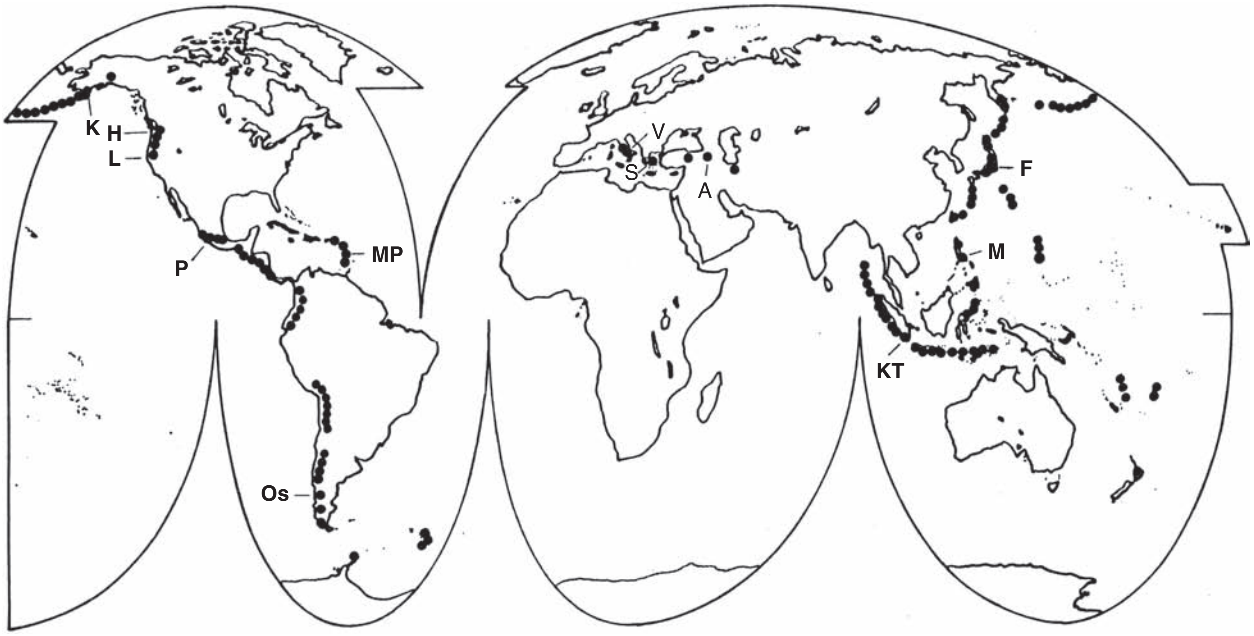


Fig. 15.6 Distribution of active and recently active volcanoes formed above Benioff zones (individual dots may represent more than one volcano). Locations of the following well-known volcanoes are indicated: Ararat (A), Fujiyama (F), Katmai (K), Krakatoa (KT), Lassen (L), Mayon (M), Osorno (Os), Paricutin (P), Pelée (MP), St. Helens (H), Santorin (S), Vesuvius (V).

In general, the igneous rocks produced at convergent plate boundaries are enriched in large ion lithophile (LIL) elements (e.g. K, Rb, Ba, La, Pb, U) relative to high field strength (HFS) elements (P, Ti, Nb, Ta, Zr, Hf). Major elements and isotopic compositions indicate that the magmas originate by partial melting of the mantle wedge above the Benioff zone, as a result of fluids liberated from the subducting slab. Subducted sediments may contribute components through the fluids released during subduction, and in the case of subduction of young ocean floor or oceanic ridges there may actually be partial melting of the subducting plate (Pearce and Peate, 1995).

A wide range of igneous rocks is associated with convergent plate boundaries. Andesite is normally the most abundant volcanic rock, but many others, ranging from basalt (gabbro) to rhyolite (granite), are present and can predominate (Thorpe, 1982). The proportions of various rock types depend on a number of factors, some of the most important of which are the types of plates involved in the convergence, the angle of subduction, and the rates of convergence. Oceanic plates tend to produce higher proportions of basaltic rocks, whereas continental ones produce more silicic rocks. Both the angle and rate of subduction can affect the depth at which the water-releasing metamorphic reactions take place, and this, in turn, can affect the types of magmas formed.

It is commonly possible to recognize three distinct series of rocks, each of which covers most of the compositional range between basalt and rhyolite but which differ in their potassium contents. The series with the lowest K contents is the *island-arc tholeiite*, which as the name implies, is found in oceanic regions but also occurs at continental margins, where rates of subduction exceed 7 cm per year. The *calcalkali series*, which is characteristic of convergent boundaries involving continental crust (orogenic belts), has intermediate K contents. Finally, a *high-K series* occurs on continental crust. An extremely potassic series, the *shoshonites*, occurs in some belts.

The suites of rocks formed above Benioff zones commonly have distinct chronologies, as seen in the volcanic stratigraphy of dissected volcanoes. In general, the sequence goes from early mafic rocks to late silicic ones, and includes boninite, island-arc tholeiite, calcalkali basalt, shoshonite, basaltic andesite, andesite, dacite, and rhyolite. Above any one subduction zone, only part of the sequence may be present, and the proportions and compositions of individual rocks may differ depending on the type of convergent boundary (ocean–ocean or continent–continent, for example).

A continuous chemical range of rock types occurs between basalt and rhyolite (BADR in Fig. 15.7), and describing individual members of this series can create the false impression that distinct rock types exist. Terms such as “basaltic andesite” and “rhyodacite” (Chapter 6) have been introduced to bridge gaps between some of the main rock units. Although these terms increase the number of available choices, they belie the fact that the rocks form a continuous series. The descriptions below are therefore of typical samples from the midrange of the various rock types.

Normally, the earliest igneous rocks associated with convergent plate boundaries are basaltic in composition (BVSP, 1981). Where the convergent plates involve oceanic crust, the basalts are island-arc tholeiites, whereas when continental crust is involved (orogenic belts), they are either calcalkali basalt or high-K basalt, shoshonite. These basalts can be distinguished by their K O contents: <0.5% in the island-arc

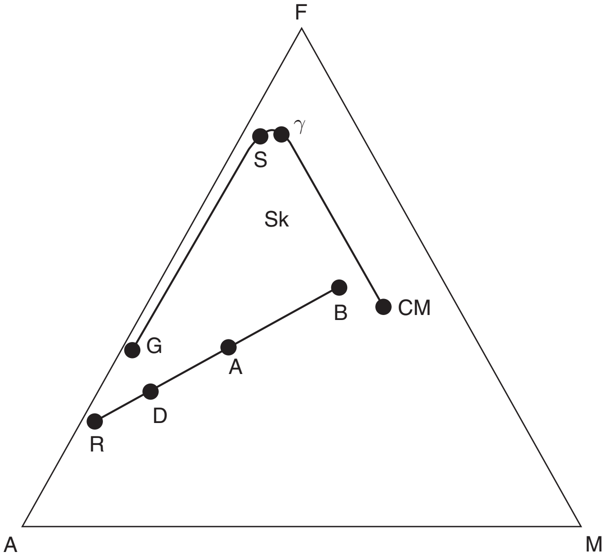


Fig. 15.7 *FMA* plot of average calcalkali rocks: basalt (B), andesite (A), dacite (D), rhyolite (R). In contrast, the differentiation trend of the tholeiitic Skaergaard intrusion (Sk), which consists of chilled margin (CM), Upper Border Group (γ), Sandwich horizon (S), and transgressive granophyre (G), shows strong early enrichment in iron.

tholeiite, ~1% in the calcalkali basalt, and >2% in the high-K basalt. Island-arc tholeiites may carry phenocrysts of plagioclase and pyroxene, but they are not strongly porphyrytic rocks. They are distinct from tholeiitic continental flood basalts and MORB, as indicated by their chondrite-normalized REE patterns (Fig. 14.43), which point to derivation from a relatively undepleted mantle. In addition to occurring in island arcs, they are found on continental margins, where the rate of convergence exceeds 7 cm per year. Calcalkali basalts have high alumina contents (about 18%), and consequently are also referred to as *high-alumina basalt* (Table 6.7). They are subalkaline and, typically, quartz normative. Plagioclase is the dominant phenocryst phase, which can be as calcic as An_{90}. Orthopyroxene and augite can also occur as phenocrysts. These basalts are strongly porphyritic, with phenocrysts forming as much as 25% of the rock. The high-K basalt, shoshonite, is the K equivalent of alkali olivine basalt, and some actually contain leucite in their groundmass. They have phenocrysts of olivine, labradorite, and augite. Because basaltic rocks tend to form early, they are likely to be concealed by later eruptions, and in the case of island arcs, these early rocks may all be submarine. Basaltic rocks may therefore be more abundant than presently recognized.

In some island arcs, such as those of Bonin, Mariana, and Papua (Fig. 15.6), there are highly magnesium rocks that range from basalt (MgO = 12%), through basaltic andesite, to andesite in composition. High magnesian contents normally indicate the presence of abundant olivine, but most of these rocks are quartz normative and typically contain two pyroxenes. They are further characterized by very low contents of Ti, Zr, Y, and REE; in this respect they resemble MORB, but their silica contents are all in excess of 54%. This peculiar group of rocks is known as the *boninites*. They may be present in other island arcs but concealed by later rocks.

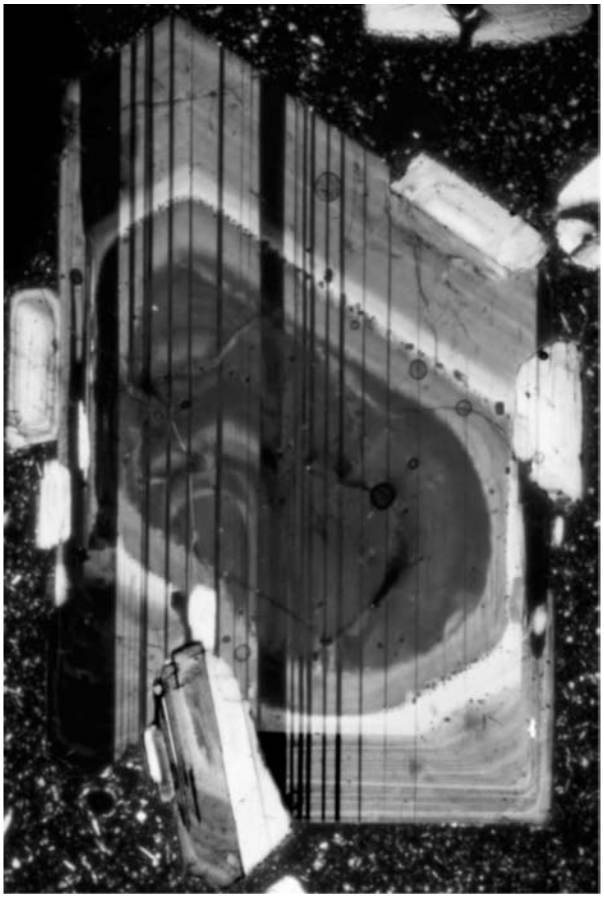


Fig. 15.8 Complexly zoned plagioclase phenocryst in andesite from Mount St. Helens. Such phenocrysts may record magma mixing events, or changes in the depth of the magma.

Andesite is the major volcanic rock formed above Benioff zones (Ewart, 1976). Much andesitic volcanism is explosive, so that many andesites form tuffs and agglomerates. Most andesites are markedly porphyritic or vitrophyric, containing large, complexly zoned plagioclase phenocrysts (Fig. 15.8), which can be as calcic as anorthite (Problem 15.3). The explosive activity and porphyritic nature of andesite may in fact be related. Explosions result in a loss of volatiles from near-surface magma chambers, which can bring about rapid crystallization (e.g. Brophy *et al*., 1999). Augite also commonly forms phenocrysts, as does orthopyroxene. Olivine phenocrysts may occur, but not in rocks containing orthopyroxene phenocrysts. Orthopyroxene also occurs in the groundmass of orogenic andesites, where it may be rimmed by pigeonite. Andesites associated with island-arc tholeiites typically contain pigeonite in the groundmass rather than orthopyroxene. This is a consequence of the stronger iron enrichment in the tholeiitic rocks. Resorbed phenocrysts of hornblende can also occur. Titanomagnetite is also a common phenocrystic phase in orogenic andesites but not in island-arc andesites. High-K andesites may also contain phenocrysts of sanidine. The groundmass of most andesites consists of a pale

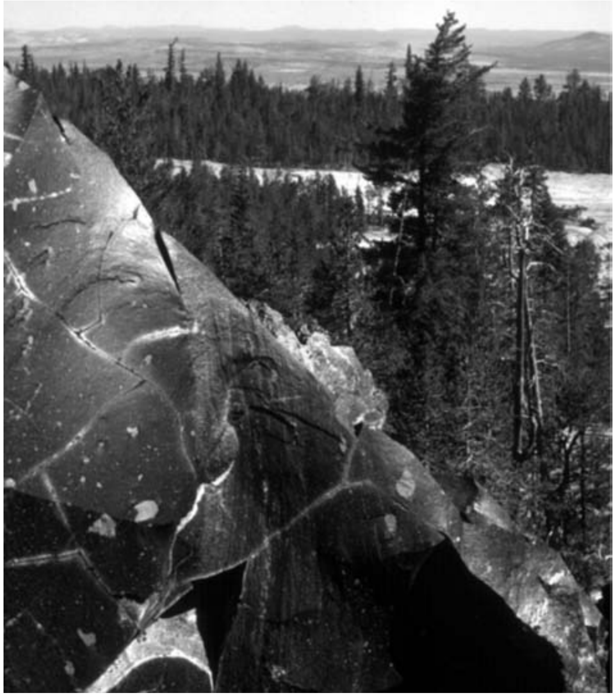


Fig. 15.9 Fine-grained vesicular basalt inclusions in obsidian, Glass Mountain, Medicine Lake Highlands, California (see Fig. 4.33). Eruption of obsidian was probably caused by the influx of basalt at depth (Grove *et al.*, 1997).

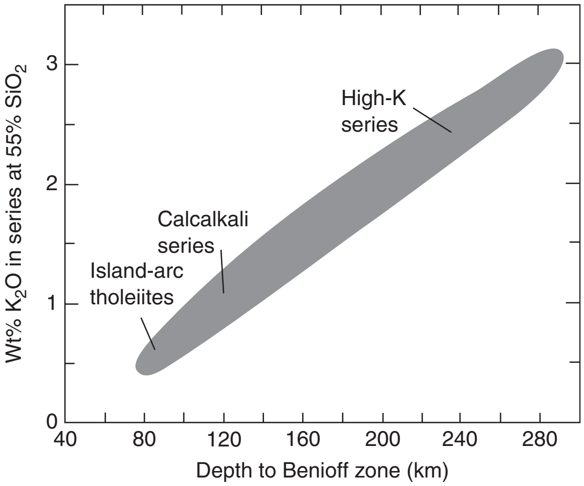


Fig. 15.10 Plot of K_2O content of rocks containing 55% SiO_2 versus the depth to the Benioff zone. (After Hatherton and Dickinson, 1969, and others.)

brown glass or cryptocrystalline mesostasis. Many pyroclastic andesites are crystal tuffs, with phenocrysts surrounded by glass shards.

The bulk composition of andesites can vary considerably due to the presence of inclusions and xenoliths. Some of these have glomeroporphyritic texture, and may represent phases that accumulated in the magma at depth. Others are true xenoliths. Many phenocrysts in andesites are not in equilibrium with the melt in which they occur. This is most clearly seen in the case of plagioclase (Fig. 15.8) but is also true of other phenocrystic phases. Magma mixing is the most satisfactory explanation for these complexities, but failure of phenocrysts to reequilibrate with magma as it rises through the lithosphere must also be a contributing factor, especially if these magmas rise rapidly (Demouchy *et al.*, 2006; Problem 5.17).

With increasing silica content, andesites grade into dacites and then rhyolites. These rocks tend to be glassier than andesites and many are obsidians. With increasing silica, the magmas also erupt more explosively, and the percentage of pyroclastic rocks increases until ash flows and ash falls constitute most of the rocks at the rhyolitic end of the series. Dacites and rhyolites are generally porphyritic or vitrophyric, and those that are not pyroclastic are commonly flow-banded. Plagioclase phenocrysts, as calcic as bytownite, occur in most dacites except those of the high-K series, where the phenocrysts are no more calcic than andesine. Phenocrysts of augite, hypersthene, hornblende, biotite, Fe–Ti oxides, and quartz may also be present. Sanidine is a common phenocryst in high-K dacites. In the rhyolites, plagioclase is still a common phenocrystic phase, but its composition is normally more sodic than andesine. It is accompanied by phenocrysts of sanidine, and quartz, which commonly have embayed outlines.

Like andesites, dacites and rhyolites tend to be inhomogeneous, containing xenoliths or schlieren of more mafic volcanic rocks (Fig. 15.9). In many cases, the mafic rocks were liquid when incorporated by the silicic magma (Fig. 14.34). This is clear evidence, then, of magma mixing.

A strong correlation exists between the potassium content of volcanic rocks and the depth to the Benioff zone (Fig. 15.10), as first demonstrated by Hatherton and Dickinson (1969). To separate spatial variations in K_2O from those due to differentiation, Hatherton and Dickinson restricted their comparisons to rocks containing 55 wt% SiO_2. At this concentration, which corresponds to a low-silica andesite (basaltic andesite), the 0.5% K_2O typical of the island-arc tholeiite series corresponds to a depth to the Benioff zone of about 80 km, whereas the approximate 1.0% K_2O of the calcalkali series corresponds to a depth of 120 km, and the >2.0% K_2O in the high-K series corresponds to a depth in excess of 200 km. Other incompatible elements show this same correlation.

The most likely cause for this variation is the amount of melt formed in the source region, but changes in the composition of partial melts with depth, and differences in the thickness through which melts rise and assimilate low melting fractions containing incompatible elements could also play a role. Incompatible elements enter the first formed melt, but they become diluted if additional melting takes place. This type of geochemical evidence has been used to argue that the degree of melting decreases with increasing thickness of the lithosphere (Plank and Langmuir, 1988; Pearce and Parkinson, 1993). High-K magmas in southern and northern Tibet have been attributed to very low degrees of melting (<2% and <4%, respectively) as a result of convective removal of subcontinental lithospheric mantle. The hot asthenosphere that flows in to replace the lithosphere raises the temperature just enough to develop small amounts of melt (Williams *et al*., 2004).

In addition to compositional variations across arcs, there can be lateral variations. Rhyolites (and granite), for example, occur at the eastern end of the Aleutian arc where the overriding plate is of continental material, but they are almost totally lacking from the western end where only oceanic plates are involved. Assimilation of crustal material would seem a likely explanation for this variation, but as shown below, magma has longer residence times in chambers in continental crust than in oceanic crust, and this could allow more time for differentiation. In the Lesser Antilles, however, the variation occurs in the mafic members of the series, which change from tholeiitic in the north to alkaline in the south. Because these differences affect the most primitive members of the series, the differences must be generated in the mantle source regions.

The correlation of intrusive rocks with convergent boundaries is not as easily made as with volcanic rocks, because it is only in older belts that erosion has exposed the plutonic rocks. Nonetheless, these rocks constitute an important fraction of the material created at convergent boundaries. Plutonic equivalents of all the volcanic rocks can be found, ranging from gabbro, through diorite and granodiorite, to granite. Representatives of the low-, intermediate-, and high-K series can be recognized, but members of the low-K series, occurring mainly in island arcs, are less likely to be exposed by erosion than those occurring in orogenic belts.

Although andesite is volumetrically the most abundant volcanic rock, its plutonic equivalent, diorite, is normally a minor constituent of the plutonic suite. Instead, granodiorite, or even granite, is the most abundant plutonic rock. These rocks form batholiths extending the entire lengths of arcs, but as pointed out in Section 4.13, these huge bodies are a composite of many smaller intrusions. Compositional zoning, schlieren, xenoliths, or even pillows of one rock type in another are common and produce considerable inhomogeneity in all rock types. One feature that is prominent because of its absence is large magma chambers in which differentiation of basaltic magma has given rise to silicic rocks. Moreover, there is no evidence of the bodies of ultramafic cumulates that would have to be formed if the silicic magmas are products of differentiation of basaltic magma. Nor is there geophysical evidence to indicate that these dense rocks are hidden at depth in the crust.

Inclusions of mafic rocks in most granites and granodiorites signal an input of basaltic magma into the crust at depth (Fig. 4.79). If basaltic magma were to rise just to the base of the crust, it could differentiate there, leaving dense cumulates in the underlying mantle. Once sufficient low-density silicic magma had formed, it would rise into the crust. Such a magma would bear a mantle isotopic signature. On the other hand, a basaltic magma emplaced at the base of the crust could raise temperatures to the point that crustal rocks would partially fuse, and, again, once sufficient low-density silicic melt had formed, it would rise (Annen *et al*., 2006). In this case, however, the silicic magma would bear the imprint of the crustal rocks from which it formed.

White and Chappell (1983) have shown that granitic rocks can be divided, on the basis of chemistry, into two groups that may reflect these two extremes in origin: the *I-* and *S-type granites* (Section 6.10). I-type granites are relatively rich in Na and Ca and so have hornblende as the major mafic mineral. Initial ^{87}Sr/^{86}Sr ratios (Section 13.4) are less than 0.708, indicating a large input of mantle-derived magma. Porphyry copper deposits are associated with this type of granite. S-type granites, in contrast, are Na-poor, Al-rich rocks that contain normative corundum, which appears modally as muscovite. They contain biotite instead of hornblende. Xenoliths and schlieren of metamorphosed sedimentary rocks are common. Their initial ^{87}Sr/^{86}Sr ratios are greater than 0.710, which indicates a large component of continental crust. These granites are believed to have formed through partial fusion of sedimentary rocks. Tin deposits are associated with S-type granites.

I-type granites occur nearer convergent boundaries than do S-type granites, suggesting that thicker crust may be necessary to cause partial melting of crustal rocks. S-type granites are water-rich, and as they rise toward the surface and the pressure decreases, they solidify (Section 11.6). This type of magma is therefore not as widely represented as the I-type among volcanic rocks. Other types of granites have been identified on the basis of geochemical traits, but these simply serve to emphasize how varied the conditions of formation of these rocks are.

Igneous rocks associated with convergent plate boundaries clearly form under a wide range of conditions. No single model of their genesis is therefore possible to make. The one common thread they all share is that the igneous activity is triggered by the subduction of an oceanic plate. If the overriding plate is also oceanic, magmas can be derived only from oceanic crust or underlying mantle. As will be shown in Section 23.3, these magmas must be derived largely from the mantle wedge overlying the Benioff zone. The subducted slab contributes mainly water to the magmas, which have a tholeiitic character. Because no continental crust is present, silicic magmas must be differentiates of mafic magmas. Their presence in island arcs is therefore proof that differentiation is capable of producing such an extreme rock type. The quantities of silicic rock, however, are small. In the South Sandwich Islands, for example, dacite and rhyolite constitute only 4% of the rocks (Baker, 1968). This may mean that primary magmas are capable of producing only this small amount, but it may also indicate that without a thick continental crust through which to rise, magmas have insufficient time to differentiate (see residence time at the end of this section).

When the overriding plate is continental crust, the diversity of rock types increases, as does the proportion of silicic rocks. Based on their occurrence in island arcs, silicic magmas can clearly be derived by differentiation from basaltic magma. Isotopic data, however, indicate that a large proportion of the silicic rocks in orogenic zones is formed by partial melting of crustal rocks. Intermediate members of the series could therefore be the products of either differentiation or magma mixing.

The series formed by basalt, andesite, dacite, and rhyolite in the calcalkali series of orogenic belts (BADR in Fig. 15.7) defines a relatively narrow trend in FMA diagrams, as would be expected of a differentiated series. This trend, however, differs from that commonly observed in known differentiated magma bodies in that intermediate members do not exhibit iron enrichment. If the series is the product of differentiation, the lack of iron enrichment can be accounted for by early crystallization of magnetite, which is a result of the higher Fe^{3+}/Fe^{2+} ratio in these rocks compared with that in tholeiites. In addition, if crystal fractionation is responsible for the differentiation of this series, it may have to take place at depth where cooling rates are slow. When these magmas rise into near-surface magma chambers and volatiles are lost through explosive activity, crystallization may be too rapid to allow for separation of crystals (Brophy *et al.*, 1999), and this effect increases with the increasing silica content and viscosity of the magmas (Marsh, 1981). The evidence for magma mixing in the series, however, is strong and includes inhomogeneities in volcanic and plutonic rocks, isotopic data, complex zoning of phenocrysts, and the presence of phenocrysts that are out of equilibrium with their surrounding melt. The straight-line variation of BADR in the FMA diagram could therefore simply reflect mixing of basaltic and silicic end members. In all likelihood, the series owes its compositional variation to some combination of fractional crystallization and assimilation of crustal material (e.g. Grove *et al.*, 1997).

Attempts have been made to determine the time necessary for magmas either to differentiate or to assimilate material by using short-lived radioisotopes (Hawkesworth *et al.*, 2000). In Section 13.2, U and Th were shown to decay in a long chain of reactions to produce Pb isotopes. These overall decay schemes are slow (half-lives ~Ga) and hence are useful for dating old rocks, but some of the intermediate steps have shorter half-lives, which are on the timescale of magmatic processes.

In Chapter 13, we saw that the rate of radioactive decay is proportional to the number of atoms of the radionuclide present (Eq. (13.2)). In a chain of reactions, however, the number of atoms of any member of the series depends on the decay rate of the preceding member of the series. Consequently, given sufficient time, the decay rates of the parent and the daughter become the same (Dickin, 1995). If they did not, the values of N in Eq. (13.2) would change and this, in turn, would change the decay rates until they did become equal. ^{238}U, for example, decays after a number of steps to ^{230}Th, which in turn decays by α emission to ^{226}Ra, with a half-life of 75.4 ka. After ~5 half-lives (>380 ka), the rate of production and decay of ^{230}Th reaches a steady equilibrium state. If the rapidly decaying ^{230}Th is measured with respect to the slowly decaying ^{232}Th (half-life 13.9 Ga), samples that initially contained different amounts of ^{238}U will plot in an isochron diagram along a line with a slope of 1:1 (Fig. 15.11). The adherence of points to this line, which is known as the *equiline*, carries no age information, but it does show that the reactions had time to establish equilibrium. If a magma in which the reactions had established equilibrium were intruded into the crust where it differentiated or assimilated material, the equilibrium could be disturbed, if the magmatic processes fractionate U from Th. Take, for example, a magma in which the equilibrium ratio of $^{230}Th/^{232}Th$ was 1.12 (Fig. 15.11). If this magma were enriched in U relative to Th by either fractionation or assimilation, it would shift to the right of the initial point in the isochron diagram. These U-enriched samples would then start decaying to reestablish the equilibrium chain reaction. If samples had undergone different degrees of enrichment in U (Fig. 15.11), an isochron develops whose slope records the age since the initial system was displaced from equilibrium. The individual samples can be comagmatic rocks or different minerals separated from the same rock. If the magma had been enriched in Th relative to U, the points would have shifted to the left of the equiline, and decay to reestablish equilibrium in the chain reaction would again have produced an isochron (x in Fig. 15.11).

Reagan *et al.* (2003), for example, have shown that rhyodacite from Crater Lake, Oregon, plots almost on the equiline, but low-Sr andesite is enriched in ^{238}U over ^{230}Th and has relatively high $^{230}Th/^{232}Th$. This andesite, consequently, plots to the right of the equiline (similar to the gray circles in Fig. 15.11) and defines an isochron of ~57 ka. In contrast, high-Sr andesite is enriched in ^{230}Th over ^{238}U and has relatively low $^{230}Th/^{232}Th$. This andesite plots to the left of the equiline (e.g. x in Fig. 15.11) and defines an isochron of 7.7 ka.

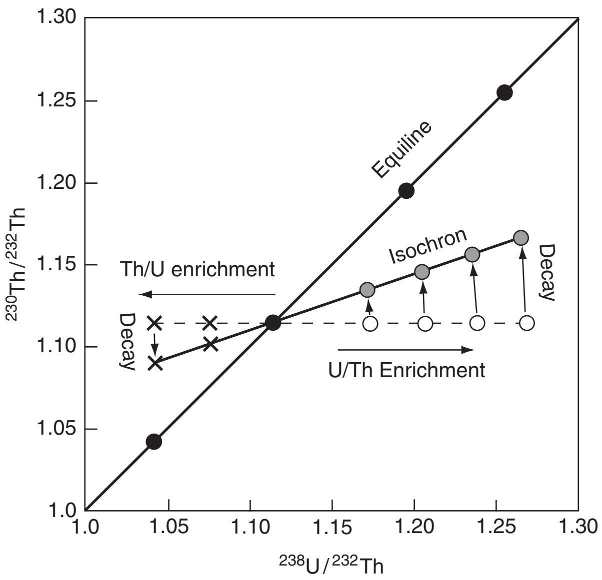


Fig. 15.11 Plot of $^{230}Th/^{232}Th$ versus $^{238}U/^{232}Th$. Samples on the equiline (black dots) have had time for the various decay steps between ^{238}U and ^{206}Pb to achieve equilibrium. Open circles represent samples that have recently been enriched in U relative to Th; these samples then start decaying so as to reestablish the equiline, and while they are doing this they create an isochron that indicates how long the reequilibration has been progressing. The crosses represent samples that have been enriched in Th, which also starts decaying to reestablish the equiline. For the points to return to the equiline, ^{230}Th must decay to ^{226}Ra, which takes almost 400 ka; in the meantime, the isochrons provide a convenient means of timing igneous processes. See text for discussion.

The Crater Lake rhyodacite has trace element and isotopic compositions that indicate mixing of crystal-fractionation products from both high- and low-Sr andesitic magma and minor assimilation of granitic wall rocks. The fact that the rhyodacite plots very near the equiline in the ^{230}Th–^{238}U isochron plot and those samples that deviate from the equiline define a very steep 145-ka isochron indicates that the rhyodacite resided in a crustal magma chamber for a considerable period of time (>100 ka). In contrast, the andesites show that their U–Th compositions were disturbed more recently and, thus, their residence time in the magma chamber must have been much shorter (57 ka and 7.7 ka). The fact that the low-Sr andesite has lower concentrations of incompatible elements but higher ratios of Ba/La, U/Th, and ^{87}Sr/^{86}Sr compared with the high-Sr andesite is interpreted to indicate that the low-Sr andesite differentiated from a basaltic parent that involved a larger influx of fluids released from subducted sediments. Because the high-Sr andesite defines a 7.7-ka isochron, which is the age of the climactic eruption that generated the present Crater Lake, its intrusion may well have played a role in triggering that eruption.

Reagan *et al*. (2003) show that rhyolite and dacite from Glass Mountain in the Medicine Lake area of California (Figs. 15.9 and 4.33) and from Katmai, Alaska, are similar to the rhyodacite from Crater Lake in having long crustal residence times in contrast to much shorter times for associated andesite and basalt. Similar studies of the Soufrière volcano on St. Vincent in the Lesser Antilles island arc indicate a long-lived magma chamber, in which little crystallization has occurred in 60 ka (Heath *et al*., 1998). For magma to remain molten for such a length of time requires either excellent insulation or repeated intrusions of basaltic magma to replenish heat loss. Because the isochrons have not been disturbed for 60 ka, replenishing-basaltic magma must either have the same isotopic composition as the magma in the chamber or old and new magmas do not mix. Basaltic magma could pool at the base of a silicic magma chamber and transfer just heat to the overlying silicic magma (see Fig. 14.10).

Silicic magma chambers above subduction zones may therefore have long, complex histories, involving repeated pulses of basaltic or andesitic magma into their base. These pulses supply the heat necessary to maintain magmatic temperatures and, through mixing, produce a range of intermediate rock types. The residence time in these chambers can vary considerably. Tholeiitic magmas may pass through rapidly (~1 ka) in chambers developed above subduction zones involving only oceanic plates, whereas andesitic magma and more silicic magma may reside for longer times (~100 ka) in chambers developed in continental crust above subduction zones.

15.4 CONTINENTAL FLOOD BASALTS AND LARGE IGNEOUS PROVINCES

Continental flood basalts are one of the most consistent rock types in the geologic record. Comparison, for example, of the Proterozoic Zig-Zag Dal basalts of eastern North Greenland (Kalsbeek and Jepsen, 1984) and the Tertiary basalts of County Antrim, Northern Ireland (Patterson, 1951), reveal similarities in the following: tectonic setting, associated sedimentary rocks; thicknesses and lateral extent of flows; internal structures of flows; and compositions of flows, including major and trace elements, and isotopes. These similarities are important, because they indicate that the source from which these rocks have been derived has not changed significantly over this period; also, similar quantities of heat (volumes of magma) have been released, and mechanisms of emplacement and differentiation have not changed.

A description of the morphology of flood basalts and their internal structures is given in Section 4.2 and will not be repeated here. One of their most remarkable features is the enormous cumulative volume of the flows in any one province, which can be in the hundreds of thousands of cubic kilometers. Individual flows can also have enormous volumes, some measuring hundreds of cubic kilometers. Eruptions of flood basalt take place rapidly from fissures or strings of vents along fissures. The feature, then, that sets continental flood basalt volcanism apart from other types is the very large volumes of magma that must be able to collect and erupt in short periods of time (<1 Ma).

Flood basalts are formed in association with continental rifting (Fig. 15.12) and thus occur along passive continental margins that were previously successful branches of rift systems. They may also extend into continents along paleorifts that follow failed arms. Flood basalts are not restricted to rift systems, as are alkaline igneous rocks (Section 15.6), but can extend into surrounding regions. The lavas are interlayered with sedimentary rocks, which are typically of the terrestrial red-bed variety. Sedimentation precedes the volcanism, indicating that subsidence and crustal thinning are precursors to igneous activity. Subsidence continues during and after volcanism, with the result that the lowest stratigraphic members can eventually be well below sea level.

The enormous volumes of magma that are erupted in flood-basalt provinces has led many geologists to believe that large quantities of heat are required for their formation. This has led to the idea that flood basalts form when a mantle plume first impinges on the lithosphere (White and McKenzie, 1989). The fact that many flood-basalt provinces on continental margins can be traced via oceanic ridges to hot-spot oceanic islands whose chain of islands and seamounts bear testimony to long-lived igneous activity supports such an interpretation (Fig. 15.12). Coffin and Eldholm (1994) coined the term large igneous province (LIP) to include flood-basalt provinces, oceanic plateaus, and oceanic ridges, all of which are characterized by large volumes of erupted material (Saunders, 2005). Even in ancient terranes where volcanic rocks may have been removed by erosion, extensive dike sets, many forming a radial pattern with diameters of several thousand kilometers, can indicate a LIP (Ernst and Buchan, 2003; Ernst *et al*., 2001).

One of the oldest flood-basalt provinces contains the 1.1-Ga-old Keweenawan basalts of Lake Superior (Fig. 15.12). These basalts are believed to have formed along a failed rift

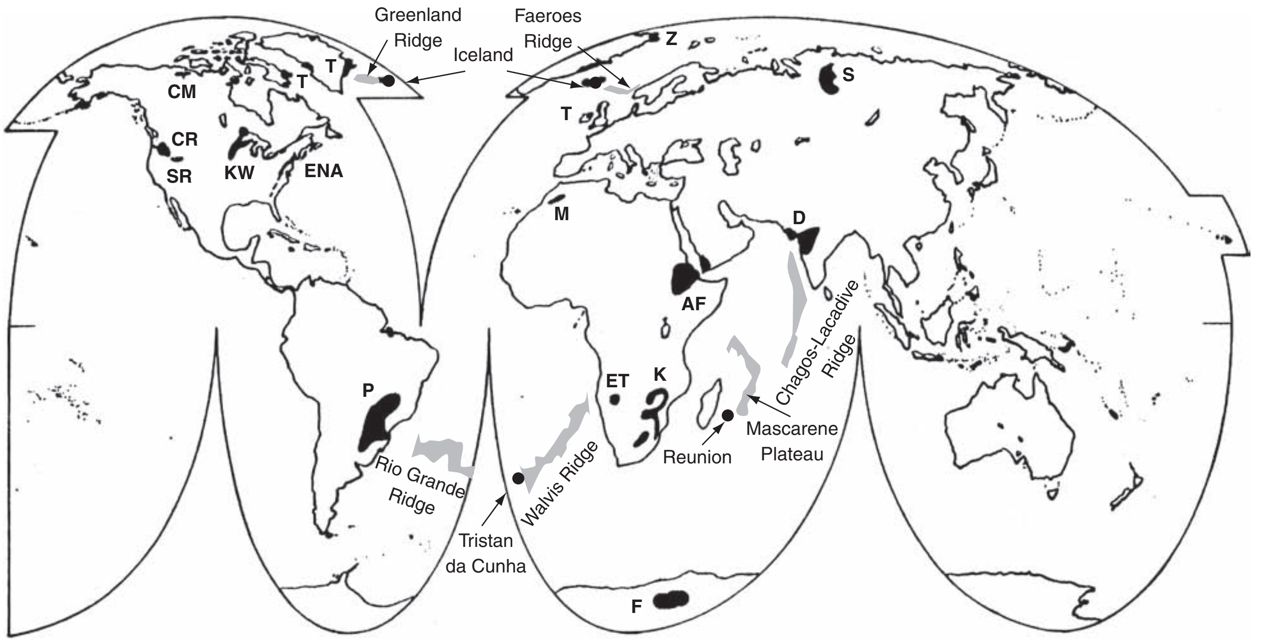


Fig. 15.12 Distribution of continental flood basalts. Afar (AF), Columbia River (CR), Coppermine River (CM), Deccan (D), Eastern North America (ENA), Etendeka (ET), Ferrar (F), Karoo (K), Keweenawan (KW), Morocco (M), Parana (P), Siberia (S), Snake River (SR), Thulean (T), Zig-Zag Dal (Z). Also shown are oceanic ridges that connect flood-basalt provinces with present oceanic hot spots.

that is now marked by an enormous geophysical anomaly known as the *midcontinent gravity high* (BVSP, 1981). A similar paleorift, but of early Mesozoic age, occurs within the Siberian Platform. Flood basalts were erupted in many regions during the Mesozoic because of the breakup of Pangea. First, with the opening of the central Atlantic, basalts erupted in many basins along the entire length of what is now the eastern seaboard of North America and the coast of Morocco, and then with the opening of the South Atlantic, the Parana basalts formed in Brazil and Etendeka and Karoo basalts in South Africa. At the same time, the Ferrar basalts erupted in Antarctica and Tasmania. During the Cretaceous, the Deccan traps formed in northwestern India as it broke away from Gondwanaland. Still later (~50 Ma), during the opening of the North Atlantic, flood basalts of the Thulean Province formed in northwestern Scotland, Antrim (Northern Ireland), and the southeast coast of Greenland. When the Arabian and African plates started to separate, the vast flood-basalt province (350 000 km^3) of the Afar triangle was formed between 30 and 20 Ma. The most recent (Miocene 17–6 Ma), extensive continental flood basalts were erupted in the Columbia and Snake River plains of the northwestern United States (Fig. 4.2).

The Deccan traps of India can be connected with the Reunion hot spot via the Chagos–Lacadive ridge and Mascarene–Seychelles ridges. Similarly, the flood-basalt provinces of Parana in South America and Etendeka in South Africa can be related to the Tristan da Cunha hot spot through the Rio Grande and Walvis ridges. Finally, the British Tertiary flood basalts and those of Greenland can be related to the Iceland hot spot by ocean ridges (Fig. 15.12). In the case of these flood-basalt provinces, there is no doubt that they were initiated over hot spots. However, in the case of the central Atlantic magmatic province (CAMP), which produced the flood basalts and dikes of eastern North America and Morocco, there is no evidence of an obvious hot spot or plume. Here the igneous activity seems to be equally developed along the entire length of the divergent boundary between North America and Africa.

The most common type of flood basalt is a quartz tholeiite, which contains about 52% SiO_2 and has a magnesium number of 55. Olivine tholeiites also occur, and in some provinces, transitional and even alkali basalts are present. These rocks are either aphyric or only slightly porphyritic, carrying phenocrysts of olivine, plagioclase, and augite. The augite is commonly rounded and resorbed, as are rare phenocrysts of orthopyroxene. Regardless of the composition of the basalt, individual flows are remarkably homogeneous, and they rarely contain xenoliths of country rocks or cognate inclusions of cumulate material. Although vesicles occur near the base and top of flows, flood basalts do not appear to have formed from highly gas-charged magmas. Indeed, pyroclastic rocks are extremely rare. Most flows are massive, unless they are pillowed from having flowed into lakes that formed in the sedimentary basins. A few provinces have small volumes of dacitic or rhyolitic flows. These silicic rocks occur late in the history of a province and are more common where central volcanic complexes develop.

Irregular sheet-like intrusions of diabase intrude the sedimentary rocks beneath the lavas. They can normally be chemically matched with specific volcanic units. These sheets are particularly abundant near the unconformity at the base of the associated sedimentary sequence. The Palisades Sill, for example, is emplaced at or near the base of the Triassic sedimentary rocks in the Newark Basin of New Jersey. These sheets rarely intrude the older rocks beneath the sediments, especially if the older rocks are deformed and metamorphosed. The sheets are normally composed of massive, medium-grained diabase. They rarely exhibit prominent cumulate layering but, nonetheless, do undergo considerable differentiation, which is marked by strong iron enrichment. Toward the top of most thick diabase sheets, lenses of coarse-grained granophyre are formed from the residual liquids (Shirley, 1987).

Dikes and dike swarms are also associated with flood basalts. Some dikes were undoubtedly feeders to the flows, but exposures of actual connections between dikes and flows are rare. Most dikes are less than 5 m wide, but their cumulative thickness in a swarm can be considerable. The orientation of dike swarms can therefore be used to determine directions of crustal extension during ancient episodes of rifting. Dikes, like the flows, show little variation in composition within individual dikes, but separate dikes can have compositions as varied as the flows. In contrast to the flows, the dikes are generally more porphyritic and may also contain xenoliths of country rock. Regional variations in dike composition have been found. For example, dikes associated with the Mesozoic basins of eastern North America are all quartz tholeiites north of Pennsylvania, but to the south, olivine tholeiites also occur (Weigand and Ragland, 1970).

In most flood basalt provinces igneous activity is episodic. Each episode is marked by lavas that, although perhaps covering a wide range of composition, are nonetheless characterized by a geochemical signature that distinguishes them from lavas of other episodes. Periods of sedimentation or erosion may separate the episodes, but in some cases, episodes may overlap slightly in time. Although the compositions of successive episodes may not form a clear evolutionary progression, lavas within an episode may follow an apparent differentiation trend. In many flood basalt provinces, three distinct episodes can be recognized, which commonly result in the lavas being grouped into lower, middle, and upper, as for example in the Zig-Zag Dal, Keweenawan, Newark, Deccan, and Antrim basalts.

The trace element and isotopic data for tholeiitic flood basalts indicate small but variable amounts of crustal contamination of magmas that are initially derived from a mantle that is only slightly depleted in incompatible elements relative to the bulk Earth (Section 13.4). This initial magma is probably similar to the tholeiitic magma of ocean islands (Thompson *et al*., 1983). The chondrite-normalized REE pattern of flood basalts (Fig. 14.43) has the same steep negative slope that characterizes the ocean island rocks, and indicates initial derivation from mantle containing garnet. The isotopic ratios of $^{87}Sr/^{86}Sr$ and $^{143}Nd/^{144}Nd$ are variable, but their average is near that for the bulk Earth (Section 13.4). This initially led DePaolo and Wasserburg (1979) to invoke a relatively primitive undifferentiated source for these rocks, but subsequent work has shown that the source is somewhat depleted, and that mixing with crustal material gives compositions near that for the bulk Earth (BVSP, 1981, p. 87). Isotopic ratios within a given province may show a progression toward more crustal values as flows become younger, but even within an individual unit there can be considerable variation. A positive correlation between the $^{87}Sr/^{86}Sr$ ratio and the SiO_2 content is also found, as might be expected if the higher ratios are associated with crustal contamination.

Experiments indicate that at low pressure tholeiitic flood basalts are multiply saturated, or nearly so, with olivine and plagioclase (BVSP, 1981). This suggests that fractionation of olivine or plagioclase, or both, has taken place; that is, a magma is not likely to have a composition on the plagioclase–olivine cotectic unless it has been fractionated there by crystallizing the phase in excess of the cotectic. Geochemical mass balance calculations (Section 14.2 and Problem 14.2) on these basalts commonly require that augite be included as a primary crystallizing phase if solutions with small residuals are to be obtained. Although augite is commonly not an early crystallizing phase in 1-atm experiments, it is at higher pressures (BVSP, 1981). The amount of plagioclase fractionated from these basalts cannot be great; otherwise, they would have larger negative europium anomalies (Fig. 14.43).

If multiple saturation is due to low-pressure fractionation, it is natural to wonder what the composition of the primary magma was, where the fractionation took place, and what processes were involved. In answering these questions, it is important to keep in mind that the source of these lavas must be capable of repeatedly supplying enormous volumes of homogeneous magma through laterally extensive fissures.

There is no reason to suppose that the primary magma in all flood-basalt provinces must be identical; inhomogeneities in the mantle, for example, could provide different source materials. But the compositional uniformity of flood basalts, in particular the quartz tholeiite, has led petrologists to seek a single primary magma, and they have narrowed the search to two likely candidates: quartz tholeiite and picritic basalt (Cox, 1980). The most abundant, and commonly the most primitive, lava in flood-basalt provinces is the quartz tholeiite. It is, therefore, a good candidate for a *parental* magma. However, its magnesium number (<60) is too low to be in equilibrium with olivine of the composition thought to exist in the upper mantle (Fo_{90}). It must therefore be a differentiate of a more primitive magma, or the mantle beneath flood-basalt provinces contains more iron-rich olivine than is currently believed. Some provinces contain picritic basalts, and these do have magnesium numbers (~70) that are compatible with mantle olivine compositions; they could therefore be *primary* magmas. The objection to them being *parental* to the more evolved flood basalts is that they are absent from many provinces and even where present, are normally not abundant. Advocates for a primary picritic magma must therefore invoke plumbing systems that minimize the chances of this magma reaching the surface.

Regardless of which magma is primary, the compositional variation among flood basalts appears to be controlled by low-pressure fractionation involving olivine, augite, plagioclase, and, in some, orthopyroxene. These phases are fractionated in the associated sheet-like intrusions. But with the iron enrichment that normally characterizes this differentiation, residual liquids would have difficulty extruding rapidly in large volumes because their densities would be greater than that of the overlying sedimentary rocks. More important, dikes having the compositions of the differentiated lavas can be found cutting through these sills and associated sedimentary rocks. The source of these magmas must therefore be still deeper. Evidence of large magma chambers lying immediately beneath the sedimentary rocks is rare. The Keweenawan province may be an exception; there, the mid-continent gravity high points to large volumes of mafic rocks at depth. In other provinces where this evidence is lacking, magma reservoirs must be hypothesized to exist in the lower crust.

Cox (1980) proposed that these magma chambers form at the crust–mantle boundary in much the same way as those associated with ophiolite complexes (Fig. 15.5), the only difference being that they are overlain by continental crust rather than oceanic crust. Picritic magma, on reaching the base of the crust, spreads laterally, floating the less dense crust above it. In regions of rapid crustal extension and fracturing some picritic magma may reach the surface, but most is filtered out by the low-density crust. Some contamination of the magma with crustal material may take place at this stage. With cooling, the differentiating picritic magma becomes less dense and eventually is able to intrude the overlying rocks; at this point flood basalt volcanism is initiated. Reference to Figure 3.1 or your answer to Problem 14.15 will reveal that the lowest density achieved by a differentiating tholeiitic magma occurs when the magnesium number is about 60. This, then, could explain the ubiquity of the quartz tholeiite lavas.

Further differentiation of the magma in this lower crustal chamber leads to iron enrichment and greater densities (Fig. 3.1). As the density of the magma approaches that of the overlying crust, the driving force for extrusion diminishes. Nonetheless, voluminous, high-iron flood basalts are common, even though their magmatic density would have been greater than that of the rocks they intrude. It may therefore be necessary with this model to provide some additional driving force for intrusion.

A point of considerable significance is the aphyric nature of most flood basalts. The mantle source and lower crustal magma chamber cannot be overheated – extra heat would simply result in more melting in the mantle or slower crystallization and differentiation in the chamber. The lack or scarcity of phenocrysts must therefore be due to resorption of crystals during the rise and decompression of the magma. Providing a melt is relatively dry, which is indicated by the scarcity of hydrous minerals and pyroclastic rocks with flood basalts, rapid decompression is accompanied by melting. Flood basalts are certainly emplaced rapidly (see Chapter 3, and Shaw and Swanson, 1970), so no significant loss of heat can occur en route to the surface. Thus, as the magma rises (Fig. 15.13), its temperature approaches, and perhaps exceeds, the liquidus. While crystals are melting, the latent heat of fusion causes the temperature of the magma to fall, but once all crystals have been resorbed, the temperature remains essentially constant until cooling starts in the lava flow. The temperature of the magma would therefore reach its highest point above the solidus on the surface of the Earth. Up to this point, resorption of crystals and assimilation of xenoliths would take place.

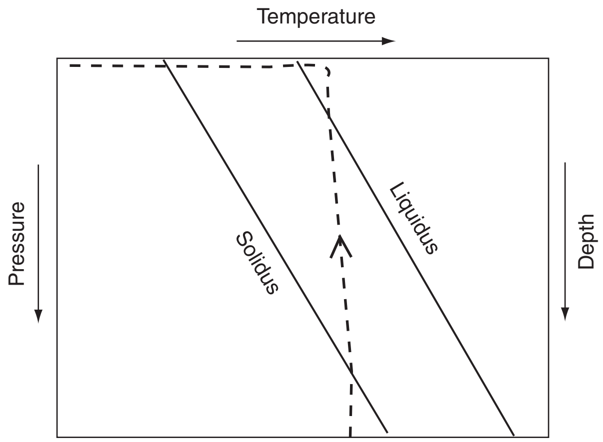


Fig. 15.13 Possible path of anhydrous continental flood-basalt magma as it rises to the surface. Decreasing pressure causes melting, and the magma may become totally liquid before reaching the surface. Because of rapid extrusion rates, magma reaches its highest point above the solidus on reaching the surface.

In light of the relations illustrated in Figure 15.13, it is not surprising that flows are commonly less porphyritic than their feeder dikes. This is well illustrated by a 250-km-long diabase dike that fed the first lava in the Mesozoic Hartford basin of Connecticut (Fig. 4.49). Normal faulting associated with the subsidence of the basin has exposed the dike at levels ranging from an initial depth of 10 km to the point where the dike actually connects with the surface flow. The chilled margin of the dike in the deepest section contains zoned phenocrysts of euhedral plagioclase and orthopyroxene and rounded augite. These phenocrysts constitute 27% of the rock. Where the dike connects with the flow, chilled margins contain only 15% phenocrysts, including euhedral plagioclase, rounded orthopyroxene and augite crystals, and small euhedral crystals of olivine. The olivine apparently is a low-pressure phase in this dike, taking the place of orthopyroxene, which was clearly stable in the dike at the deeper section. In the lava flow itself, phenocrysts constitute only 5% of the rock, and these are euhedral plagioclase and olivine with minor rounded augite crystals and extremely rare, highly embayed crystals of orthopyroxene (Philpotts, 1998).

If a flood basalt magma does contain crystals that are resorbed on rising toward the surface, a number of interesting effects ensue. First, as the percentage of crystals decreases, the effective bulk viscosity of the magma decreases (Eq. (2.9)). This then allows the magma to rise more rapidly, which, in turn, increases the rate of resorption. Second, as crystals melt, the bulk volume of the magma increases, due to the approximate 10% volume expansion on melting. The increased volume would result in greater rates of intrusion, which, in vertical dikes, would cause more melting and thus more expansion. Both of these processes would have feedback effects that could contribute to the high rates of extrusion of flood basalts.

A further consequence of resorption is that the magma continually changes its composition to be in equilibrium with the minerals that are dissolving; that is, the magma is multiply saturated. Thus, instead of invoking a magma chamber in which fractional crystallization brings about multiple saturation, a rising, melting crystal mush could produce the same result. Moreover, should there be any separation of crystals from liquid on the way to the surface, the resulting compositional variations would look similar to those formed by fractional crystallization in a magma chamber. Finally, turbulence in a superheated dike or lava flow would provide ideal conditions for the fusion and assimilation of crustal xenoliths (Fig. 3.2).

The volumes of magma erupted in flood-basalt provinces are so large that the gaseous emanations may have had significant effects on global climate and possibly caused mass extinctions (Wignall, 2005). When the Lakagigar fissure eruption occurred in Iceland in 1783, it killed 75% of Iceland's livestock, and yet this eruption was only 14 km^3. Individual flood-basalt flows can have volumes of >1000 km^3 and occur in thick sequences with little if any sediment between them (eruptions closely spaced in time). These large eruptions could therefore have had a far more serious environmental effect than the Lakagigar eruption. The idea that these eruptions may have caused mass extinctions became much more attractive when improved absolute age determinations showed that the Deccan traps had an age of 65 Ma, which corresponds to the Cretaceous–Tertiary boundary and the extinction of dinosaurs. Other mass extinctions at the end of the Permian (250 Ma), the end of the Triassic (200 Ma), and in the Toarcian Stage (180 Ma) correspond closely in age with the flood basalts of Siberia, CAMP, and Karoo–Ferrar, respectively. Despite these correlations, other equally large flood-basalt provinces have no associated extinction. Self *et al.* (2005) have argued that the release of SO_2 could have produced aerosols which could have caused global cooling. The release of CO_2, which could cause warming through the greenhouse effect, is not thought to have been great enough to cause significant effect and was probably less than the current anthropogenic production.

15.5 LARGE LAYERED IGNEOUS COMPLEXES

Grouped together under this category are rocks that may have very different origins, but they are all products of the differentiation of tholeiitic magma. At a number of times in Earth history, particularly during the Precambrian, conditions as varied as continental rifting and meteorite impact led to the development of enormous volumes of tholeiitic magma, which, on slowly cooling, differentiated into remarkable suites of cumulates, some of which are of great economic importance. Although size is not an essential criterion in this category, it is mostly in large intrusions that magmatic processes were able to produce the strongly differentiated rocks. Detailed descriptions of most of these occurrences are given by Wager and Brown (1967).

The form and structures of these bodies have been described in Section 4.11 and will not be repeated here. Suffice it to say that most of these intrusions are lopoliths. Some are circular to elliptical or lobate in plan, such as the Bushveld Complex of South Africa (Fig. 15.17) or the Sudbury lopolith in Ontario (Fig. 15.33), whereas others are elongate, such as the Muskox Intrusion in the Northwest Territories of Canada (Fig. 15.18) or the Great Dyke of Zimbabwe. Some that are circular to elliptical, such as the Sudbury body, were formed by meteorite impact, their shape being determined largely by the explosion crater, which was later filled with magma (Section 15.9). Perhaps most large circular to elliptical lopoliths have this origin. The elongate ones are associated with crustal extension, dike swarms, and flood basalts. For example, the Muskox Intrusion is overlain by the Coppermine River basalts and the Duluth Gabbro at the west end of Lake Superior by the Keweenawan basalts.

Almost all large layered complexes are of Precambrian age. For example, the Stillwater Complex in Montana is 3.2 Ga, the Great Dyke of Zimbabwe is 2.5 Ga, the Bushveld Complex is 2.0 Ga, the Sudbury lopolith is 1.7 Ga, the Kiglapait Intrusion, Labrador, is 1.3 Ga, and the Muskox and Duluth bodies are both 1.1 to 1.2 Ga. Layered complexes did form in Phanerozoic time, such as the Tertiary Skaergaard Intrusion on the southeastern coast of Greenland (Fig. 15.15), but they are much smaller. The large volumes of the Precambrian bodies may reflect higher heat production early in Earth's history. In addition, at least for those formed through meteorite impact, the flux of large meteorites would have been much greater during the Precambrian.

Whether beneath a rifting continent or a meteorite explosion crater, large-scale melting of the upper mantle produces tholeiitic magmas that rise to form lopoliths. Debates on the precise composition of these magmas center around the amount of normative olivine they contained and the degree to which they were contaminated by crustal rocks. Chilled margins containing pristine quenched magma do not occur on large intrusions. The finer-grained rocks that are found at contacts may have assimilated country rocks or been altered by circulating hydrothermal solutions. Commonly, nearby small dikes and sills that are thought to be related to the main intrusions are used to estimate initial magma compositions. These are all quite similar, containing about 50% SiO_2, 7% to 9% MgO, 11% CaO, and less than 0.3% K_2O. Their most variable constituent, Al_2O_3, ranges from 15% to 19%. More olivine-rich picritic magmas have been proposed for intrusions containing large volumes of olivine cumulates, but these can be explained with less olivine-rich magmas if the chambers remain open and are repeatedly replenished and flushed with primary magma.

Crystallization in these lopoliths takes place from the floor upward, to form gently dipping layers of cumulates. Toward sidewalls, dips can become steeper and even vertical. Evidence of slumping and cross-bedding in the layered rocks attests to their partly molten state for some time

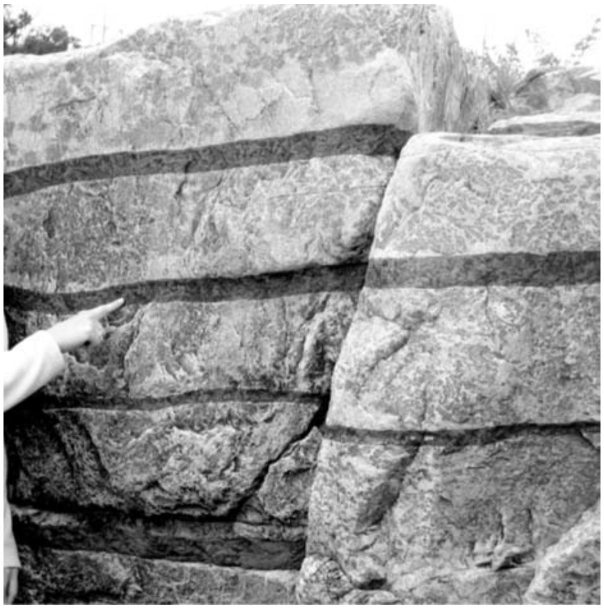


Fig. 15.14 Chromite layers in anorthosite in the Dwars River section in the eastern part of the Bushveld Complex, South Africa. Note the sharp contacts between chromite and anorthosite. (Photograph by Alan Boudreau.)

following accumulation (Figs. 4.67, 4.70). Other discordances in the layering can be explained as a result of magma convection. Some form of convection probably takes place in all of these bodies and may play an important role in promoting differentiation (Section 14.4).

Because olivine and pyroxene are the first minerals to crystallize from these magmas, dunites, peridotites, and pyroxenites form the lowest layers. Once plagioclase starts crystallizing, noritic and gabbroic rocks form, and then, as the plagioclase becomes more sodic, these rocks grade into iron-rich ferrodiorites. Toward the top of intrusions, granophyre is common. It may be derived by differentiation from the mafic magma or from assimilation and partial melting of country rocks.

The actual order of crystallization of minerals depends on the precise composition of the magma. Three of the most common sequences are (I) olivine–orthopyroxene–plagioclase–clinopyroxene, (II) olivine–clinopyroxene–plagioclase–orthopyroxene, and (III) olivine–orthopyroxene–clinopyroxene–plagioclase. The quaternary phase diagram for the system diopside–forsterite–anorthite–silica (Fig. 10.37(A)) shows how slight changes in the composition of a primary magma plotting near the peritectic can bring about these different sequences (Problem 15.5). Because of the peritectic, olivine does not continue crystallizing from the differentiating magma but is replaced by calcium-poor pyroxene. If a magma has a composition that causes minerals to appear as in sequence I, the order of formation of cumulate rocks would be dunite, harzburgite, bronzitite, norite, and gabbro (refer to Fig. 6.3 for rock classification). With sequence II, the rocks would be dunite, wehrlite, olivine gabbro, and two-pyroxene gabbro. With sequence III, they would be dunite, harzburgite, lherzolite, websterite, and two-pyroxene gabbro. The only difference in these three sequences is in the stage at which orthopyroxene crystallizes, which, as pointed out in Section 14.10, may reflect different degrees of crustal contamination, orthopyroxene appearing earlier in the more contaminated magmas.

Crystallizing with olivine as a primary phase is chromite. This mineral is normally an accessory phase because magmas contain only trace amounts of chromium. But in most large lopoliths chromite-rich layers have repeatedly formed by a process of accumulation that probably involves magma mixing (Section 14.11). Chromitite layers vary in thickness from centimeters to meters, the thicker ones being of economic value (Fig. 15.14). They occur in olivine and orthopyroxene cumulates and in anorthosite, but once clinopyroxene starts crystallizing, the small amount of chromium in the magma is taken into the pyroxene structure, and chromite no longer forms. Layers of almost pure magnetite rock also form in these intrusions but normally not until intermediate stages of fractionation.

Accompanying the succession of rock types are variations in the composition of minerals, the ferromagnesian ones becoming more iron rich, and the plagioclase, once it starts crystallizing as a cumulus mineral, becoming more sodic. Detailed inspection of these trends commonly reveals reversals, which are most likely the result of intrusion of fresh batches of magma. These new surges are also typically associated with the formation of an olivine or chromite cumulate. Most large intrusions have evidence of numerous periods of replenishment and mixing during their early stages of crystallization. This causes them to continue accumulating ultramafic rocks rather than fractionating to more evolved differentiates.

Once replenishment of the magma ceases, fractionation of the remaining liquid progresses to form gabbro, ferrodiorite, and eventually granophyre. At the ferrodiorite stage, magnetite and apatite are typical cumulus phases, and they can produce layers of extremely iron-rich rocks. The iron-rich magmas at this stage of differentiation may enter a two-liquid field (Fig. 14.26), and an immiscible silicic liquid separates to form granophyre. If the two-liquid field is not encountered, the fractionating magma will eventually attain a granophyric composition in any case. This late-stage silicic liquid remains molten to lower temperatures than the associated mafic fraction and commonly forms irregular transgressive sheets.

The upper part of these intrusions can be simple or complex, depending largely on the size of the intrusion. In smaller bodies, rocks accumulating on the floor of the intrusion finally meet a narrow zone of rocks that solidifies down from the roof, the residual liquid being trapped between the two advancing crystallization fronts. Granophyre typically occurs in this zone. In large intrusions, so much heat is liberated through the roof that overlying rocks are partly melted. The silicic melts derived in this way then mix with the granophyres formed by fractionation of the mafic magma and complex relations develop.

We will summarize three different layered intrusions. The Tertiary Skaergaard Intrusion, while small compared with most Precambrian bodies, is the best-studied intrusion

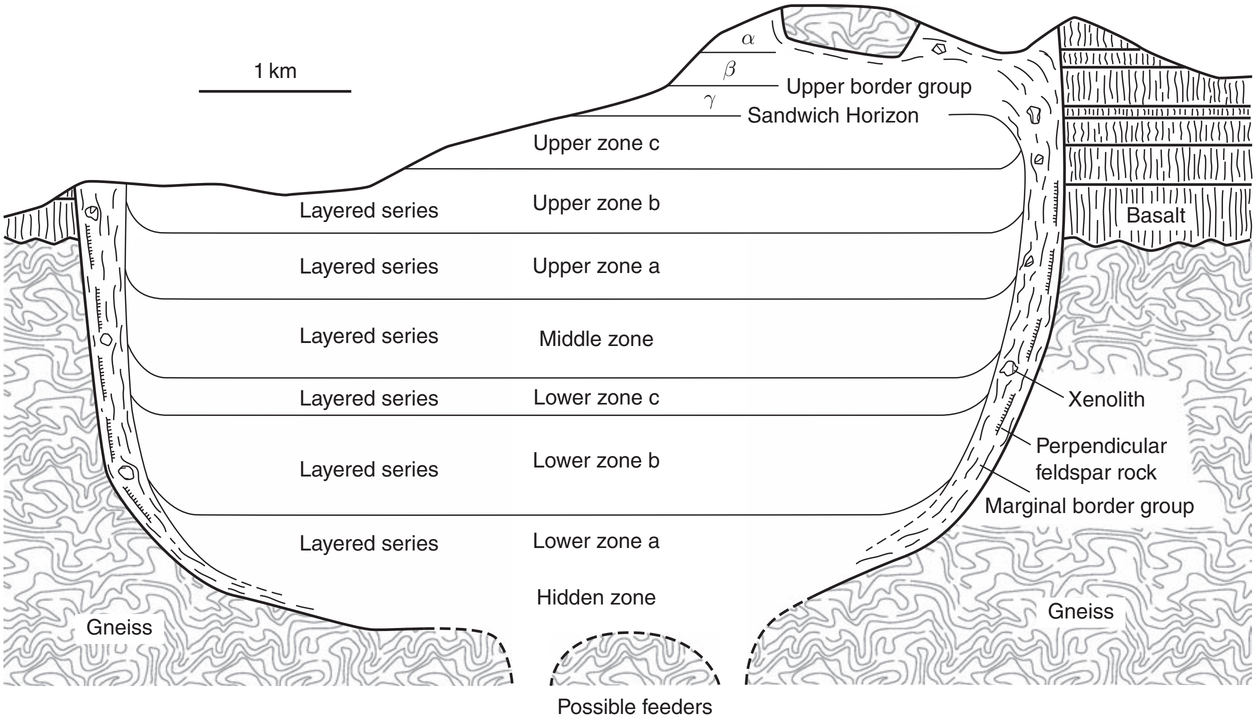


Fig. 15.15 Simplified cross section of the Skaergaard Intrusion, East Greenland. No vertical exaggeration. (Drawn from data in Wager and Brown, 1967, and Nielsen, 2004.)

in the world, and its history is relatively simple – intrusion of a single batch of magma that differentiated in a closed chamber. At the other end of the spectrum is the enormous Bushveld Complex, the world's largest igneous body – it formed through the mixing of several pulses of magma. Finally, the Muskox Intrusion was formed by repeated surges of magma that each time flushed out most of the old fractionated magma.

15.5.1 Skaergaard Intrusion, East Greenland

The Skaergaard Intrusion is one of a number of bodies on the east coast of Greenland that formed during the initial opening of the North Atlantic (Wager and Brown, 1967). Many diabase dikes and flood basalts in the vicinity of the intrusion formed during this same period of crustal extension. The body itself developed near the unconformity separating the Tertiary basalts from the underlying Precambrian gneisses (Fig. 15.15). The magma made room for itself by lifting a body of gneiss and basalt, as evidenced by a roof pendant of gneiss at the stratigraphic level of the basalts. The resulting intrusion is an irregular box-like body measuring 11 km from north to south, 8 km from east to west, and from 3.4 to 4 km deep (Nielsen, 2004).

The rocks of the intrusion form three distinct groups (Wager and Deer, 1939). A Marginal Border Group, a Layered Series, and an Upper Border Group. The *Marginal Border Group* is a 200- to 300-m-thick zone of generally finer-grained olivine gabbro that crystallized from the wall inward. Layering, which parallels the contacts, is formed by growth of different minerals on the wall of the intrusion or by flowage of convecting magma. In some layers, plagioclase crystals up to 2.5 cm long grew out from the wall into stationary magma to form a *crescumulate* texture, which Wager and Deer refer to as the perpendicular feldspar rock. Xenoliths of country rock occur in this zone. The compositions of the minerals change systematically through this zone in a manner that mimics the change found in the layered series of rocks. For this reason it is thought that the rocks of the Marginal Border Group accumulated on the walls of the intrusion at the same time that the Layered Series accumulated on the floor.

The main part of the intrusion is formed by the *Layered Series*, which consists of over 2700 m of layered gabbroic rocks that accumulated on the floor of the intrusion. This series is divided into three zones on the basis of whether or not the rocks contain olivine. In the *Lower Zone* magnesian olivine (Fo_{67-53}) is a cumulus phase (Fig. 15.15). Toward the top of this zone, olivine develops a reaction relation to form pigeonite ($Ca_9Mg_{56}Fe_{35}$), which on cooling changed to inverted pigeonite. The top of the Lower Zone is marked by the disappearance of olivine, and the overlying *Middle Zone* consists only of layered two-pyroxene gabbro. Magnetite also is a cumulus phase in this zone. The magma became progressively more iron rich with differentiation and when the pigeonite reached a composition of $Ca_9Mg_{45}Fe_{46}$ olivine once again became a stable phase, but this time with a composition of Fo_{40} (Problem 15.6). The reappearance of olivine marks the beginning of the *Upper Zone*. Olivine continues to change its composition, becoming pure fayalite at the top of the Upper Zone where the residual liquid encountered the downward solidifying roof. Plagioclase in the Upper Zone is all less calcic than An_{45}, so these rocks are ferrodiorites. Apatite is also a cumulus phase in this zone. A fourth zone, referred to as the *Hidden Layered Series*, underlies the Lower Zone rocks. It would presumably contain earlier crystallizing minerals, examples of which are found in the outer parts of the Marginal Border Group (olivine with a composition of Fo_{81}, for example). Recent drilling and geophysical surveys indicate that the Hidden Zone constitutes only a small fraction of the intrusion (Nielsen, 2004).

The *Upper Border Group*, which solidified downward from the roof, can also be divided into three zones, which mineralogically match the three zones in the Layered Series. In the Upper Border Group, however, the division must be based on plagioclase composition, because contamination by xenoliths of granite gneiss prevented olivine from crystallizing in many of the rocks. The zone nearest the roof, α, contains plagioclase more calcic than An_{53}; the next zone, β, contains plagioclase between An_{53} and An_{44}, and the lowest zone, γ, contains plagioclase less calcic than An_{44}.

Toward the end of crystallization a lens of residual liquid became trapped between the upward accumulating Layered Series and the downward solidifying Upper Border Group. This lens, which is referred to as the *Sandwich Horizon*, contains the most evolved liquids in the Skaergaard. Olivine is pure fayalite, and the clinopyroxene became so iron-rich that it reached the high-temperature stability field of iron wollastonite, which on cooling inverted to a polygonal aggregate of hedenbergite ($Ca_{43}Mg_0Fe_{57}$). This zone contains considerable quantities of granophyre, but some of the silicic liquid rose to form transgressive sheets in the overlying rocks.

The Skaergaard Intrusion was formed by a single pulse of magma that underwent extreme fractionation in a closed chamber. The differentiation resulted from fractional crystallization of a convecting magma that deposited minerals mainly on the floor of the intrusion. Wager and Deer (1939) believed that minerals separated by sinking from the convecting magma currents as they crossed the floor of the intrusion. More recently, McBirney and Noyes (1979) have interpreted the layering to result largely from the nucleation and crystallization of different minerals directly onto the solidification front.

15.5.2 Bushveld Complex, South Africa

The Bushveld Complex covers an area of approximately 66 000 km^2 and contains an 8-km-thick sequence of layered igneous rocks (Wager and Brown, 1967). In detail, it consists of four lobes, which contain similar but not identical series of rocks. The lowest parts of the intrusion, marked by the Main Chromitite layer and the Merensky Reef (Fig. 15.17), are absent from the northern and southern lobes but are present in the eastern and western lobes. At times the lobes were separate, but at others they were connected, as evidenced by prominent stratigraphic marker layers that can be traced throughout. Magma levels must have fluctuated considerably, possibly as a result of the fresh influxes of magma that gradually built the intrusion to its final size. Repeated surges of magma produce more complex sequences of rocks, and in this respect the Bushveld is different from the Skaergaard.

The Bushveld is underlain by sedimentary rocks of the Transvaal System (Fig. 15.17). Like the lopolith, they are bowed down into a large basin-like structure, which could have resulted from the loading of the dense igneous rocks or the withdrawal of magma from beneath the crust to form the lopolith. The upper contact of the body is complex and in most areas is obscured by the intrusion of the roughly contemporaneous Bushveld red granite, which overlies most of the lopolith. In places, however, the mafic rocks contact overlying Rooiberg felsite, which is also approximately the same age as the Bushveld. As shown in Section 15.9, this felsite has been interpreted as shock-melted rock formed by a meteorite impact that simultaneously triggered the mafic magmatism (Rhodes, 1975). Considerable assimilation of the felsite produced hybrid rheomorphic granophyres, which are difficult to distinguish from granophyres derived from the mafic magma.

The lowest part of the intrusion consists of approximately 100 m of noritic diabase (Marginal Group in Fig. 15.16), which contains approximately 5% normative olivine. This presumably formed by relatively rapid cooling of the original undifferentiated magma and may therefore be representative of the parental magma. Its high content of orthopyroxene, however, may indicate that it was contaminated with aluminous sedimentary rocks.

Immediately above the Marginal Group is a 1200-m-thick *Basal Series* composed of interlayered dunite, harzburgite, and bronzitite. The minerals in these rocks change little in composition throughout the series, olivine ranging from Fo_{88} at the base to Fo_{86} at the top, and orthopyroxene from En_{87} to En_{83} over this same interval. However, the mineral compositions do fluctuate, indicating additions of fresh batches of magma. These additions prevented the magma from fractionating to more evolved compositions.

Toward the top of the Basal Series numerous thin layers of chromitite appear. These are precursors to the *Main Chromitite* layer (Steelpoort), which is one of the prominent stratigraphic markers of the intrusion. Between it and another prominent stratigraphic marker, the Merensky Reef, is a remarkable sequence of 1000 m of finely layered cumulates forming the *Critical Series*. Approximately 60 m above the base of this series, plagioclase appears as a cumulus phase, and at 300 m, augite also appears. Many monomineralic adcumulate layers and reversals in the fractionation trends of minerals point to an unstable magma that was being repeatedly replenished. The layers are predominantly bronzitite near the base of the series but become mainly norites toward the top. Many chromitite layers occur throughout the series.

Olivine remains a cumulus phase throughout the Critical Series in the western lobe of the Bushveld, but it ceased crystallizing at the top of the Basal Series in the eastern lobe. This difference indicates that the magmas in these two parts of the intrusion had slightly different compositions, assuming the pressures were the same. But the Main

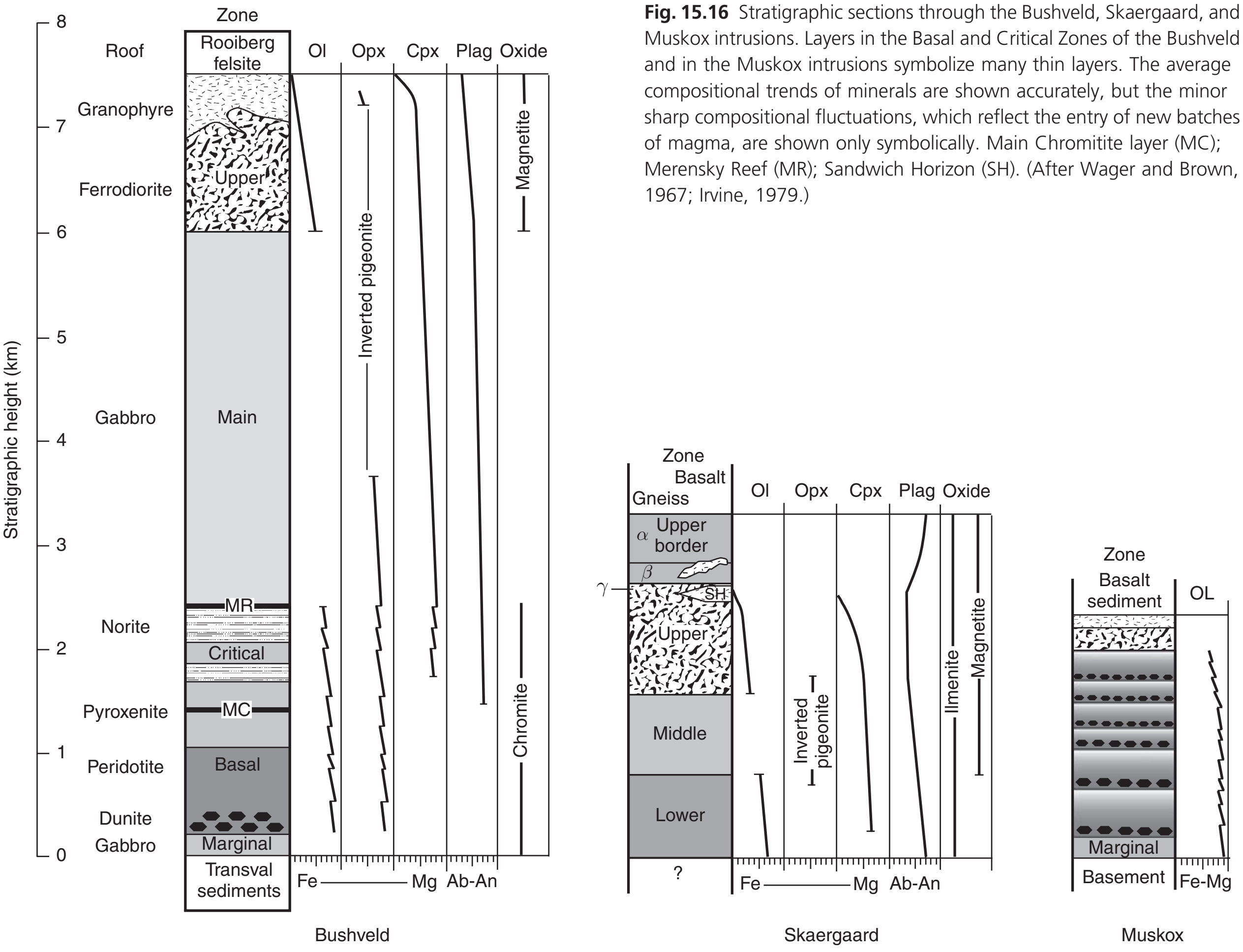


Fig. 15.16 Stratigraphic sections through the Bushveld, Skaergaard, and Muskox intrusions. Layers in the Basal and Critical Zones of the Bushveld and in the Muskox intrusions symbolize many thin layers. The average compositional trends of minerals are shown accurately, but the minor sharp compositional fluctuations, which reflect the entry of new batches of magma, are shown only symbolically. Main Chromitite layer (MC); Merensky Reef (MR); Sandwich Horizon (SH). (After Wager and Brown, 1967; Irvine, 1979.)

Chromitite layer and the Merensky Reef occur in both the western and eastern lobes. Their formation therefore cannot be extremely sensitive to magma composition. Instead, magma mixing is most likely responsible for the precipitation of these layers, especially since they are associated with reversals in mineral composition trends.

The top of the Critical Series is marked by the *Merensky Reef*, the world's most important source of platinum. This layer, which varies in thickness from 1 to 5 m, is composed of cumulus bronzite poikilitically enclosed by plagioclase with a thin chromite layer at its base. The platinum occurs with disseminated sulfides. Irvine *et al*. (1983) believe that the ore was formed as a result of magma mixing. Fractionation of a sulfur-poor magma had brought about enrichment in platinum when a new sulfur-rich magma was intruded. During mixing, an immiscible sulfide liquid formed into which the platinum strongly partitioned. The dense sulfide liquid then sank to the floor of the intrusion.

Above the Merensky Reef is the *Main Zone*, 3600 m of poorly layered gabbroic rocks. In addition to lacking the striking fine-scale layering of the Critical Series, the minerals show a steady upward change to more fractionated compositions. Apparently, during this stage, no new pulses of magma entered the chamber, so fractionation was able to take its course.

The appearance of magnetite as a cumulus phase marks the beginning of the *Upper Zone*, which consists of 1500 m of ferrodiorites. Numerous magnetite cumulate layers occur near the base of this zone, one of which, the Main Magnetite layer, forms a prominent stratigraphic marker that can be traced throughout the eastern and western lobes. No downward-solidifying upper border group exists as it does in the Skaergaard. Instead, the ferrodiorite grades upward into a granophyric zone that contains assimilated roof rocks.

15.5.3 Muskox Intrusion, Northwest Territories, Canada

The Muskox Intrusion is an elongate body (Fig. 15.18) emplaced at the unconformity between a basement of granitic and metamorphic rocks and overlying sedimentary rocks and basalts of the Coppermine River, which were erupted during the same general magmatic episode as the Muskox (Irvine, 1979). A dike extends for 60 km to the south from the intrusion but does

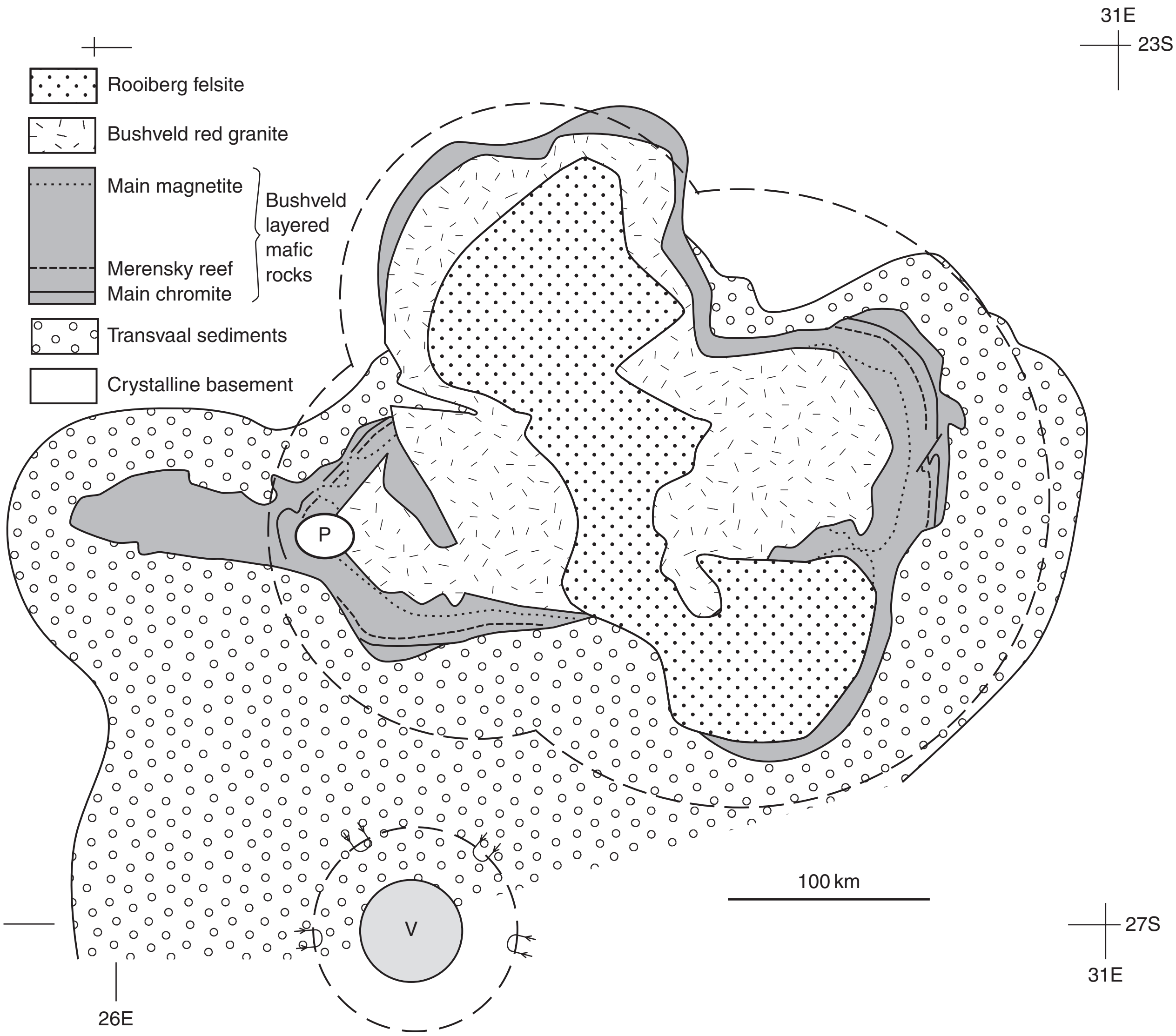


Fig. 15.17 Simplified map of the Bushveld Complex, South Africa. The Vredefort Ring (V) and three of the lobes of the Bushveld (dashed lines) may have been caused by meteorite impact. The Pilanesberg (P) plug is a younger alkaline intrusion (see Fig. 4.56). (Based on maps in Wager and Brown, 1967, and Rhodes, 1975.)

not cut the layered ultramafic and mafic rocks of the intrusion. It is therefore interpreted to be the feeder to the intrusion.

The 150-m-wide dike and 200-m-wide marginal zone of the intrusion are composed of bronzite gabbro, which grades to peridotite away from the contact. Inside the marginal zone is 1800 m of gently dipping layered series, which consist of 25 cyclic units, each of which formed by the fractional crystallization of a new batch of magma. Each cyclic unit consists of a number of cumulate layers, the mineralogy of which was determined by the order of crystallization. In the lower part of the intrusion, the sequence olivine–clinopyroxene–plagioclase–orthopyroxene was common, producing dunite, olivine clinopyroxenite, olivine gabbro, and two-pyroxene gabbro. Most cyclic units, however, do not contain the complete suite of rocks, because influxes of fresh magma interrupted the fractionation process. With time, the sequence of crystallization changed to olivine clinopyroxene orthopyroxene plagioclase and then to olivine–orthopyroxene–clinopyroxene–plagioclase toward the top of the layered series. Consequently, cumulus orthopyroxene appears at progressively earlier stages toward the top of the intrusion, indicating that later magmas may have been contaminated more with crustal rocks. Chromitite layers occur in the dunite, but they are absent where clinopyroxene is a cumulus phase.

Through a complete cyclic unit, the magnesium number of the cumulates decreases by about 10%, but injections of new magma and displacement of old kept restoring the magnesium number to essentially its original value throughout most of the layered series. Thus olivine, for example, changes from Fo_{85} to only Fo_{80} in the first 1200 m. However, during the accumulation of the final 600 m, when the magma was sufficiently fractionated to form gabbroic rocks, the composition of the olivine progressed to Fo_{60}. The layered series become progressively more silicic toward the top, with a prominent

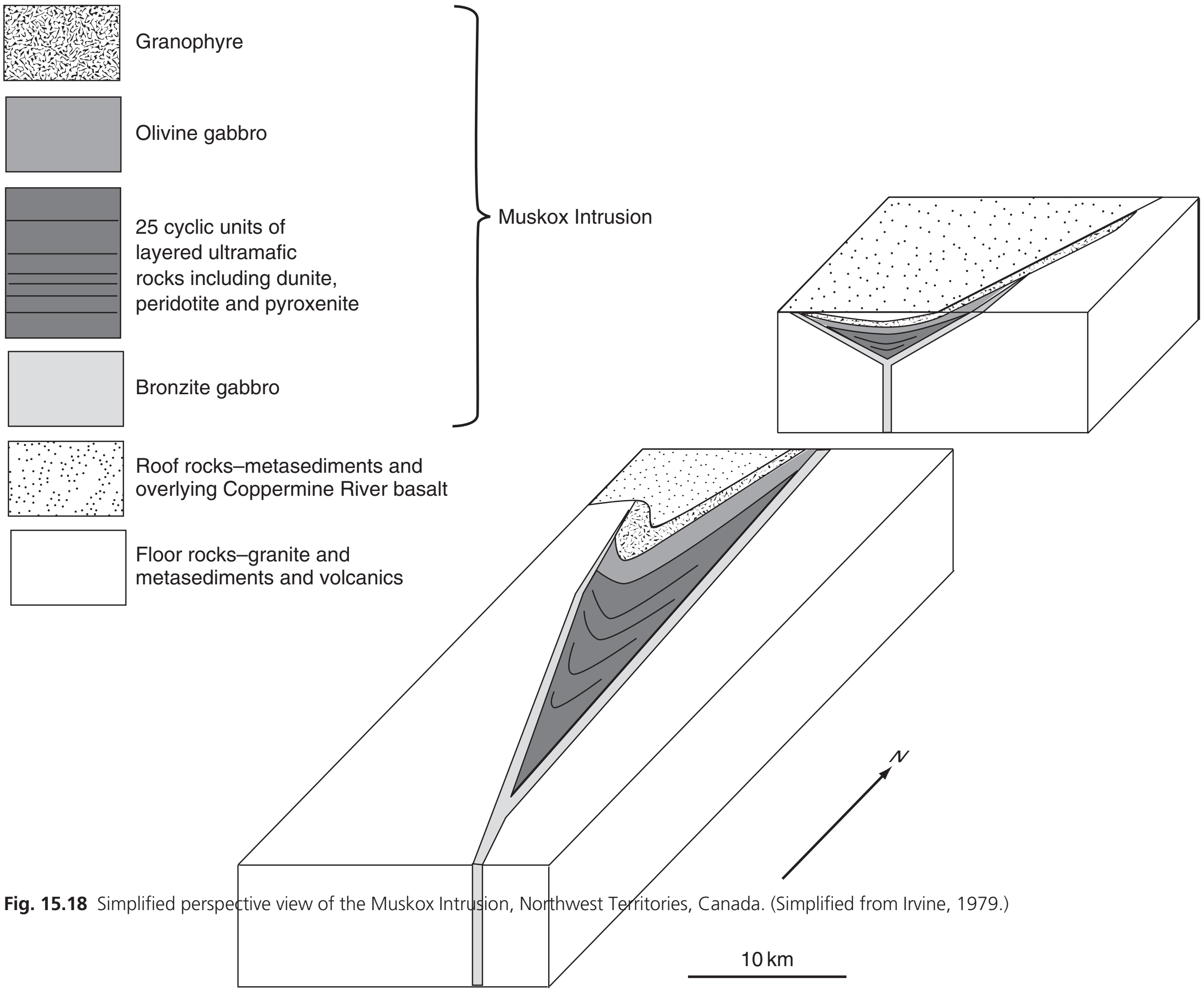


Fig. 15.18 Simplified perspective view of the Muskox Intrusion, Northwest Territories, Canada. (Simplified from Irvine, 1979.)

sheet of granophyre containing an abundance of xenoliths of the overlying metasedimentary rocks capping the intrusion.

Irvine (1979) estimates that the Muskox contains more than 50% olivine. Because magmas do not contain this high a percentage, accumulations from repeated surges of less olivine-rich magma must be invoked. From each surge, ultramafic and mafic cumulates were deposited on the floor of the intrusion, and the fractionated residual liquid was displaced by the next surge of magma, erupting possibly as the Coppermine River basalts. Despite the great thickness of the layered series, the actual volume of magma in the intrusion at any one time may have been quite small. The repeated pulses of magma would, however, have brought in the heat needed for the large-scale melting and assimilation of roof rocks.

15.6 CONTINENTAL ALKALINE ROCKS

Apart from Benioff-zone-related magmatism and continental flood basalts, igneous activity on continents is extremely limited and restricted to rift valleys and local hot spots (Sørensen, 1974a; Bailey, 1974). The volume of magma emplaced in these environments is small compared with those formed in the other two environments. Continental flood basalts are also related to rifting, but they are more common during early stages of rifting and along successful branches of rift systems, whereas alkaline magmatism predominates along the failed arms normally preserved within continents. A wide range of rock types, some of which have unusual mineralogy, commonly occur together in alkaline provinces. It is not surprising, therefore, that petrologists have devoted considerable effort to working out the petrogenesis of these rocks. As a result, alkaline rocks have received a degree of attention and a share of the rock nomenclature that far outweighs their volumetric significance. Despite these efforts, we are still a long way from fully understanding these rocks.

Alkaline igneous rocks of Archean age have not yet been found. There are, however, numerous Proterozoic examples, such as those of the Gardar province of southwest Greenland (1.2 Ga) and the Kapuskasing–James Bay belt of Ontario (1.1 Ga). The prevalence of alkaline rocks with an age of

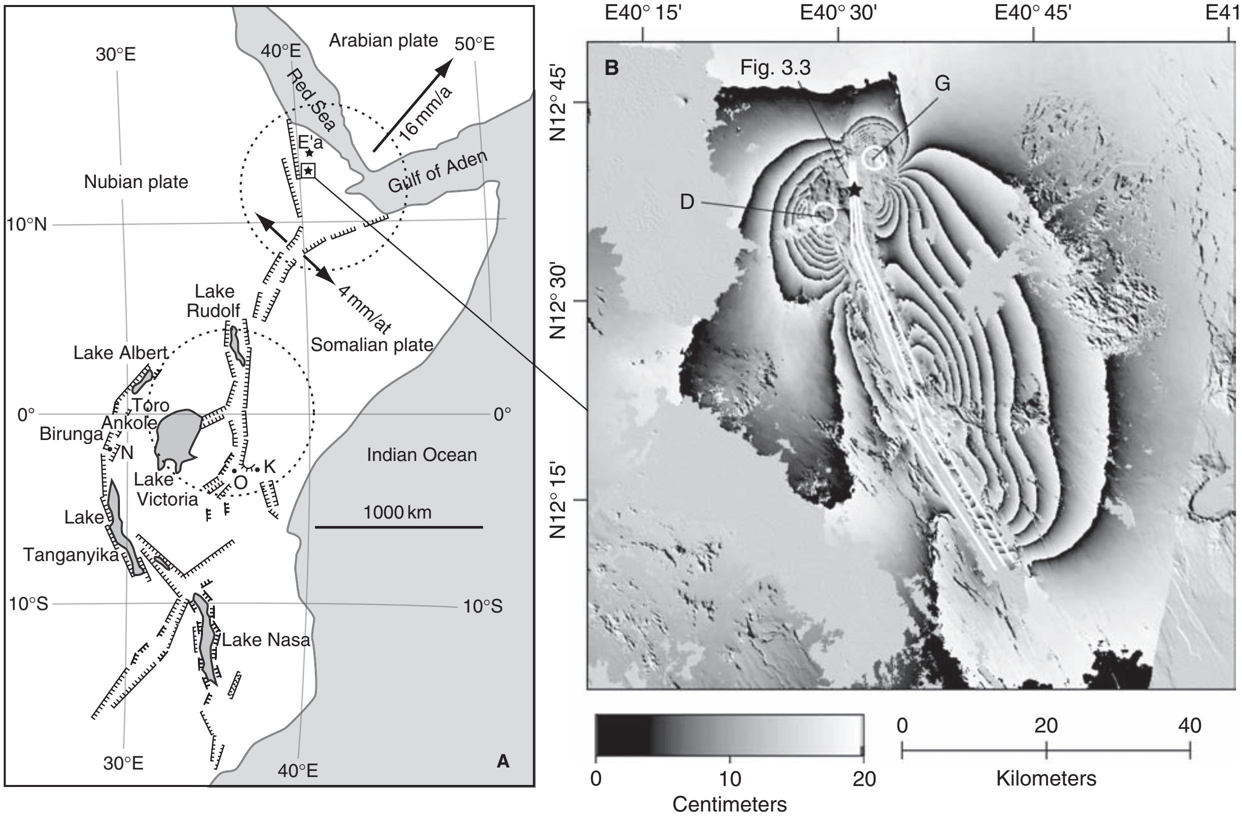


Fig. 15.19 (**A**) Rift system of East Africa and Ethiopia. Erte'ale (E'a); Kilimanjaro (K); Nyiragongo (N); Oldoinyo Lengai (O). Relative rates and direction of spreading of the Arabian, Nubian, and Somalian plates are indicated, and the locations of possible plumes are represented with dashed circles. (**B**) InSAR image of the region extending south from the Dabbahu (D) and Gabho (G) volcanoes in Ethiopia based on the difference between satellite images collected in May and October of 2005 (Wright *et al.*, 2006). The deformation records crustal extension of up to 6 m. On September 26, an eruption occurred between the two volcanoes (Fig. 3.3), but deformation extending for 60 km south was consistent with a ~3.5-m-wide dike being intruded laterally over this distance in a period of two weeks. Such dikes are what build the sheeted-dike complexes beneath spreading axes. See text for discussion. (InSAR image created by T. J. Wright with data acquired by the European Space Agency's Envisat satellite.)

0.57 Ga around the perimeters of many shield areas points to a worldwide period of rifting at this time. The fact that the beginning of the Paleozoic era, with its new life forms, is normally taken to be 0.57 Ga may not be entirely a coincidence. During the Paleozoic, alkaline rocks were formed along numerous rift systems, such as the Carboniferous Midland Valley of Scotland and the Permian Oslo graben of Norway. During the Mesozoic, alkaline rocks were formed along failed arms of the rift system that caused the breakup of Pangea, such as those of New England, the coastal regions of Brazil, and the Benue trough in Cameroon. In the Cenozoic, alkaline rocks were formed along the Rhine graben, the Rio Grande rift in New Mexico, and the East African rift system.

Although eroded paleorift systems provide excellent exposures of both volcanic and plutonic rocks, only in modern rift systems such as that of East Africa can the relation between tectonism and magmatism be precisely determined. On September 14, 2005, a series of earthquakes and faulting culminated on September 26 in the eruption of a new volcano near the Dabbahu volcano in the Afar region of Ethiopia (Fig. 15.19(B) and Fig. 3.3). What makes this event so interesting is that the deformation associated with rifting and intrusion of magma was recorded essentially in live time through use of satellite interferometry (Wright *et al.*, 2006). Twelve days prior to the eruption, the first of 163 earthquakes (magnitude >3.9) occurred. At first, these were located between the Dabbahu volcano and the Gabho volcano just to the east (D and G in Fig. 15.19(B)), but they steadily progressed southward for ~60 km. The eruption, which was of pumice, lasted for about three days after which the seismic activity subsided. Interferograms constructed from satellite images of the area taken on May 6 and October 28, 2005, give a clear record of the deformation associated with this magmatic episode (Fig. 15.19(B)). Although the satellite images are almost six months apart, most of the deformation is likely to have occurred in the two weeks following September 14. The images reveal that there was up to 6 m of opening and that over a 25-km-wide zone along the rift there was, on average, ~1.5 m of uplift, except in a narrow 2–3-km wide central zone that subsided ~2 m. Initial doming around the Dabbahu and Gabho volcanoes was followed by 2 to 3 m of subsidence. These deformations were found to be consistent with the lateral emplacement of a 60-km-long, ~3.5-m-wide dike intruded at depths between 2 to 9 km. The deflation around the volcanoes accounts for some of the estimated 2.5 km^3 of magma that must have been intruded in the dike, but some magma must have come from a deeper source. The magma in the dike must have been a low-viscosity basalt to be able to intrude for 60 km in two weeks. The pumice that was erupted at the vent (Fig. 3.3) must have come from an older magma chamber beneath the Dabbahu or Gabho volcanoes that was reactivated by the intrusion of basalt. The energy liberated by the associated earthquakes was an order of magnitude less than that which would be expected from the measured displacements. This indicates that much of the deformation occurred aseismically as a result of dike emplacement. On August 13, 2007, a fissure eruption occurred near the kink in this dike located at the center of the interferogram in Figure 15.19(B) (Yirgu, 2007). This eruption likely tapped the magma that flowed southward during the 2005 diking episode. This record of dike emplacement is probably representative of the activity that has led to the formation of sheeted dike complexes beneath spreading axes.

In East Africa, well-developed erosion surfaces of known age provide convenient datum plains with which to monitor long-term tectonic movements during rifting. These erosion surfaces indicate that the East African rift system now occupies the crest of an uplifted ridge, along which are a number of culminations or domes (Fig. 15.19(A)). These domes have diameters on the order of 1000 km and central uplifts of a couple of kilometers (Gass, 1970; Baker and Wohlenberg, 1971; Bailey, 1974). Major branches or triple junctions in the rift system are located at the centers of these domes. The largest of these is the Afro-Arabian dome centered on the Afar triangle at the northern end of the rift system. This is postulated to be the location of a mantle plume (White and McKenzie, 1989), which when first formed caused the eruption of the Afar flood basalts. The very active Erte'ale shield volcano (Fig. 4.14) is located at the center of this region. Subsequent to the flood basalts, igneous activity has continued with the emplacement of alkaline magmas over a large area of Ethiopia and Arabia (Shaw *et al.*, 2003). Farther to the south is the Kenya dome, which is centered on the triple junction to the east of Lake Victoria (Fig. 15.19(A)). Here again there is abundant alkaline magmatism.

The geophysical evidence that these domes may be surface expressions of upwelling mantle plumes, is the lower density and lower seismic velocity of the underlying mantle. Melting of the mantle appears to be localized beneath these domes, because, although eruptions can occur anywhere along the rift system, igneous activity is centered on the domes, with each dome having its own individual history. In the case of the Kenya dome, for example, igneous activity began in the early Miocene at about the same time as doming began. Large central volcanoes, composed essentially of nephelinite with minor basanite, phonolite, and trachyte, developed near the triple junction at the center of the dome. Next, fissure eruptions of phonolite issued from fractures formed during the initial rifting. The grabens at this stage were not prominent topographic features, and flows were able to extend well beyond the rift valley. Although composed of phonolite, they behaved very much like flood basalts, with individual flows extending laterally as much as 250 km. In the mid to late Pliocene and early Pleistocene, numerous central volcanoes developed on the flanks of the dome, such as Mount Kilimanjaro, which was built largely of alkali basalt, trachyte, and phonolite. Present-day activity is localized mainly within the rift and is highly variable in character. Pantellerite, for example, erupted recently near the center of the dome, and natrocarbonatite has erupted at Oldoinyo Lengai at the southern end of the dome (Fig. 15.19(A)). The association of igneous activity with uplift and rifting is clear in this case, and the domes play an important role in determining the location and composition of magmas. In general, rocks become more alkaline away from the center of domes, probably reflecting smaller degrees of melting in the mantle beneath these parts. Also, magmatism generally becomes less alkaline with time.

In addition to compositional differences within a dome, there may be larger-scale regional differences. For example, rocks of the Kenya dome are sodic, whereas those of the next dome to the west, which is associated with the western branch of the rift system, are highly potassic. Even here there are regional differences. For example, many of the rocks in the Birunga district are leucite basanites and leucite nephelinites, whereas in the Toro and Ankole districts to the north (Fig. 15.19(A)), potassium is so abundant that, in addition to leucite, kalsilite is present, and feldspar is absent; this produces rare rock types, such as ugandite and mafurite (Table 6.5). Farther to the north, rocks associated with the Afro-Arabian dome are far less alkaline, and indeed some are even tholeiitic, in particular those associated with the actual generation of ocean floor in the Gulf of Aden and the Red Sea (Gass, 1970). Voluminous alkali basalts, covering an area of about 800 000 km^2, erupted first from the center of this dome prior to the separation of the Arabian and African plates. A number of central volcanoes then developed along the rift valleys, erupting transitional basaltic rocks and peralkaline silicic differentiates. Finally, with continental breakup, tholeiitic lavas began erupting in the Afar triangle and on the new ocean floor.

These regional differences may depend on the vigor of individual mantle plumes. The more active, hot ones may undergo greater degrees of partial melting, which produces more tholeiitic magmas, whereas the less active, cooler ones may undergo smaller degrees of partial melting, which produces more alkaline magmas. Some differences may reflect lateral variations in the composition of the mantle, but melting at different depths in a vertically zoned mantle could also produce these differences.

In other areas, regional variations appear to be induced by differences in crustal rocks. The Monteregian province of southern Quebec is located at the intersection of three grabens, which formed early in the Paleozoic and are now occupied by the valleys of the Ottawa and St. Lawrence rivers and Lake Champlain (Figs. 4.74 and 15.22). The St. Lawrence and Champlain grabens succeeded in rifting apart to become the ocean that later closed to form the Appalachian Mountains. The Monteregian intrusions, which were emplaced in the Cretaceous, are localized by faults extending eastward from the Ottawa graben across the floor of the old triple junction into the Appalachian fold belt. Seismic studies reveal a steady increase in the thickness of the crust in this direction. The composition of the Monteregian rocks also changes systematically in that direction, with the degree of silica saturation increasing toward the east. At the western end of the province, feldspar is completely lacking from the rocks, which are carbonatite, kimberlite, and alnöite. Feldspar first appears near Montreal, but nepheline remains a major constituent. Not until the Appalachians are reached does quartz become a significant component of any of the rocks. Along with increasing silica activity, the size of the intrusions steadily increases (Fig. 4.74). In this case, it is tempting to invoke crustal contamination as the cause for regional compositional differences.

In many alkaline provinces, intrusions that are close to one another and clearly related by tectonic setting have very different rock types. Let us take the Monteregian province again as an example and examine the rock types encountered in going eastward through the province (Philpotts, 1974). At the western end, the first main intrusive complex is a carbonatite (no. 1 in Fig. 4.74). Next is a series of small plugs of alnöite and kimberlite (no. 2 and 3). Mount Royal (no. 4), in the center of Montreal, is composed of essexite and nepheline monzonite. St. Helen's Island (no. 5) in the St. Lawrence River is a diatreme breccia composed of Paleozoic and Precambrian rock fragments with no igneous matrix. Mount Bruno (no. 7) is composed entirely of feldspathic peridotite, as is Rougemont (no. 9). Olivine in Mount Bruno ranges in composition from Fo_{84} to Fo_{76} and the plagioclase from An_{60} to An_{10}. The olivine in Rougemont is more iron rich, ranging from Fo_{78} to Fo_{74}, but the plagioclase is extremely calcic and ranges from An_{98} to An_{80}. Between these two peridotite intrusions is Mount St. Hilaire (no. 8), an alkali gabbro and zirconium-rich sodalite nepheline syenite complex, which has one of the world's largest assortment of rare minerals. Farther to the east, Mount Yamaska (no. 11) is composed almost entirely of gabbroic rocks. Finally, Brome (no. 12) and Shefford (no. 13) are composed of gabbroic rocks with nepheline and quartz syenites.

Differences in these intrusions cannot be attributed to variable depths of erosion through a common type of magma body. Fundamental compositional differences distinguish each intrusion. Also, because these bodies are vertical plug-like stocks with steeply dipping contacts between rock types, rapid vertical changes in composition cannot be expected. Each intrusion seems to have its own separate source, for there is no geophysical evidence for a large magma reservoir underlying the entire province. Of course, the intrusions were not all emplaced at precisely the same time, which might account for some of the compositional variability. Different depths of origin of the magma in juxtaposed intrusions may also be responsible for differences.

In other alkaline provinces, there may be much less variability in the rock types. In the Permian Oslo graben, for example, 80% of the lavas consist of "rhomb-porphyry," an intermediate alkaline rock that is just saturated (Oftedahl, 1960). The plutonic rocks consist predominantly of larvikite (plutonic equivalent of the rhomb-porphyry) and syenite, with smaller amounts of peralkaline granite. Only 15% of the lavas are of alkali basalt, and an extremely small percentage of the intrusive rocks is composed of its intrusive equivalent, essexite. Similarly, in the Kenya dome in the East African rift, phonolite forms a very large percentage of the lavas. Geophysical evidence in both of these regions points to the existence of large elongate mafic magma chambers beneath the rifts. These chambers may have provided sites in which differentiation could produce the large quantities of intermediate and felsic magmas (Ramberg, 1976; Baker *et al*., 1978).

In most alkaline provinces there is a noticeable scarcity of intermediate rock types in both volcanic and plutonic series; that is, they exhibit a Daly gap. Volcanic rocks, for example, may consist predominantly of alkali basalt and trachyte, and plutonic ones of essexite and syenite. Figure 14.28(A), which illustrates the compositional range of Monteregian plutonic rocks, shows such a gap between rocks of gabbroic and nepheline monzonite composition. A similar gap exists between nephelinite and phonolite in the rocks of the Kenya dome. Baker *et al*. (1978) believe that this gap is a result of density filtering by the crust. Some dense mafic magmas may initially have risen directly to the surface, but those that did not must have had to reside in lower crustal chambers until fractionation lowered their density to the point that they could buoyantly rise through the crust. In the Monteregian province, however, where rocks as different as peridotite and syenite are juxtaposed, density does not seem to have been a critical factor. There, liquid immiscibility may have played a role in generating the gap (Section 14.7).

Some continental alkaline rocks do not appear to be associated with rift valleys, but instead form elongate belts that may track the paths of continents over mantle plumes. For example, extending through New England, and cutting obliquely across the trend of the Appalachian mountain belt, are the alkaline intrusions of the White Mountains (Fig. 15.22). This 300-km-long chain, when extrapolated to the sea, appears to connect with the New England seamount chain, which can be traced all the way to the Mid-Atlantic Ridge. On land, however, the ages of the intrusions, which range from 185 to 100 Ma, do not follow the simple linear progression that would be expected if a plate moved over a plume (Foland and Faul, 1977). A similar belt of Jurassic age extends for 1300 km in a northerly direction through Nigeria and Niger (Bowden and Turner, 1974).

These rocks typically form ring dike complexes, many of which have foundered blocks of rhyolitic lavas in their center (Fig. 4.52). Elongate chains of intersecting ring-dike complexes testify to the migration of magmatic activity with time (Fig. 4.55). The rocks are predominantly oversaturated, with granites being the most abundant. Quartz syenites are also common and may predate granites during cycles of activity. Many of the quartz-rich rocks are peralkaline. The granites of northern Nigeria contain economic deposits of tin, which occurs as the mineral cassiterite. Unlike most tin deposits, these contain no tourmaline. Mafic rocks, both younger and older than the granites, form a minor but essential part of this type of alkaline rock association. Many granites contain xenoliths of basaltic rocks, indicating that mafic magmas existed at depth, despite their scarcity near the surface. While extremely uncommon, critically undersaturated syenites and mafic rocks can occur. Some of the intrusions in Niger contain early intrusive sheets of anorthosite, one of the rare Phanerozoic occurrences of this rock type (Section 15.8).

The geochemistry of continental alkaline rocks is not very different from that of the alkaline rocks of oceanic islands. They appear to have been formed by small amounts of partial melting of relatively undepleted mantle that contains garnet. In general, alkaline magmas are rich in volatiles, in particular water. Amphibole is therefore a common mineral in the plutonic rocks, even in ones of mafic composition. Alkaline magmas have high concentrations of incompatible elements, which can become further concentrated by fractionation. Many of the felsic members of the alkaline series on continents show signs of some crustal contamination. This is particularly noticeable for magmas that resided in large crustal chambers for any length of time.

15.7 ULTRA-ALKALINE AND SILICA-POOR ALKALINE ROCKS

Associated with many but not all alkaline provinces are rocks with exceptionally high alkali contents and low silica contents. They include several varieties of lamprophyre, kimberlite, and carbonatite, and their volcanic equivalents. They constitute a very small percentage of the rocks in alkaline provinces. Nonetheless, in the case of kimberlite and carbonatite, they are of great economic importance as sources of diamond and niobium, respectively. Kimberlites and some lamprophyres are also of great petrologic interest because they contain an abundance of xenoliths ripped from the conduit walls during the rapid ascent of these magmas. Some of these xenoliths provide samples of the mantle down to depths of at least 200 km.

15.7.1 Alkaline lamprophyres

As a group, lamprophyres are porphyritic melanocratic hypabyssal rocks that contain euhedral phenocrysts of either amphibole or biotite but never feldspar (Table 6.6). Included under such a classification are lamprophyres belonging to two different associations. First, there are the calcalkaline lamprophyres, including minette, kersantite, vogesite, and spessartite. These are commonly associated with granitic rocks. Second, there are the alkaline lamprophyres, including sannaite, camptonite, monchiquite, polzenite, and alnöite. It is this second group we discuss here.

The most common alkaline lamprophyre is *camptonite*. It contains phenocrysts of kaersutitic (high TiO_2) or barkevikitic (high Fe) amphibole, titanaugite, and possibly olivine, and biotite in a fine-grained groundmass of amphibole, titanaugite, plagioclase, and iron titanium oxides, with minor amounts of feldspathoid and apatite. Calcite is also a common constituent, occurring both as a primary and secondary mineral; it may fill vesicles or form ocelli. Chemically, camptonites are equivalent to alkali basalts, which explains why they are so common in alkaline provinces. *Sannaites* are similar to camptonites, but plagioclase is subordinate to potassium feldspar in the groundmass.

Ocelli are a characteristic feature of the groundmass of camptonites (Figs. 12.29 and 14.28(D)). These millimeter- to centimeter-sized globules of felsic rock are interpreted to have formed as droplets of immiscible liquid (Section 14.7). Toward the top of sills, ocelli may coalesce to form sheets of felsic rock (Fig. 3.16). Thus on a small scale we see the development of contrasting rock compositions that are similar to those of the alkali basalts and trachytes that produce the Daly gap in many alkaline provinces on continents and ocean islands.

With decreasing silica activity, camptonites pass into *monchiquites*. The phenocrysts are similar to those in camptonites, but calcic plagioclase is lacking from the groundmass of monchiquites. The calcium and aluminum that would form plagioclase in more silicic rocks enter titanaugite as the Ca-Tschermak's molecule ($CaAl_2Si_2O_2 - SiO_2 = CaAl_2SiO_6$). These lamprophyres also contain ocelli of carbonate and nepheline monzonite (Fig. 14.27). *Fourchite* is similar to monchiquite but lacks olivine. The volcanic equivalent of monchiquite is nephelinite.

At still lower silica activities, melilite becomes an important mineral, taking the place of pyroxene ($2CaMgSi_2O_6 - {}^3/_2SiO_2 = Ca_2MgSi_2O_7 + {}^1/_2Mg_2SiO_4$). This gives rise to *alnöite*, which is composed essentially of phlogopite, melilite, olivine, titanaugite, monticellite, iron–titanium oxides, and calcite, with accessory nepheline, apatite, and perovskite. The presence of perovskite instead of sphene is a clear indication of the low silica activity of these rocks (Problem 15.8). Alnöites are remarkable in that they are both rich in magnesium and potassium, a characteristic shared with kimberlites. In other groups of rocks, the higher the magnesium content, the lower is the potassium content. Alnöites must be derived from a source containing phlogopitic mica. With this in mind, it is interesting to consider what would happen to an alnöitic magma once it reached high enough levels in the crust for the mica to become unstable. Phlogopite, for example, breaks down as follows:

$$\begin{aligned} &\underset{\text{(phlogopite)}}{2KMg_3AlSi_3O_{10}(OH)_2} \\ &= \underset{\text{(kalsilite)}}{KAlSiO_4} + \underset{\text{(leucite)}}{KAlSi_2O_6} + \underset{\text{(olivine)}}{3Mg_2SiO_4} + \underset{\text{(vapor)}}{H_2O} \end{aligned} \tag{15.2}$$

The reaction products are essential minerals in the rare volcanic rock types *ugandite* (olivine–melilite–leucite) and *mafurite* (olivine–pyroxene–kalsilite) found in the Toro and Ankole districts on the East African rift. Alnöites could therefore be the hypabyssal equivalents of these lavas (Problem 15.9 and 15.10).

Alnöites appear to have been emplaced rapidly and in some cases even explosively, for they commonly contain fist-sized nodules of mantle and lower crustal rocks; they may even be associated with diatreme breccias. Their content of phlogopite and calcite certainly indicates the presence of H_2O and CO_2. In these respects, alnöites are similar to kimberlites, with which they are commonly associated.

15.7.2 Kimberlite

Kimberlite is a rare and unusual alkaline rock that has the distinction of being the primary source of diamonds. It occurs in narrow dikes and diatremes where it forms the matrix surrounding fragments of the rocks passed through on the way to the surface (Section 4.9). Kimberlites are restricted to cratons, with the most widespread occurrences being in the shields of southern Africa and Siberia, but they are also found in North America, Brazil, Australia, and India. In many of these areas, kimberlites have been emplaced repeatedly over a timespan ranging from the Proterozoic to Cenozoic; this indicates the existence of a very stable source for these magmas. Although their age may correspond to alkaline igneous activity in nearby rift systems, kimberlites themselves are not confined to rifts. Their locations do, however, appear to be determined by prominent fractures. Kimberlites and associated rocks are treated in detail in the book by Mitchell (1986). Shorter review articles are given by Meyer (1979) and Cox (1978).

The igneous part of a kimberlite is normally severely altered by hydration and carbonation, making it difficult to ascertain the rock's original composition. Nonetheless, they are clearly volatile-rich, potassic ultramafic rocks composed of phenocrysts of olivine (commonly serpentinized) and possibly phlogopite, clinopyroxene, magnesian ilmenite, magnesian garnet, and orthopyroxene, in a groundmass of serpentine, phlogopite, monticellite, magnetite, perovskite, chlorite, calcite, and apatite. The groundmass has essentially the composition of alnöite; kimberlite then could simply be an alnöitic liquid with phenocrysts of olivine and phlogopite. Many of the phenocrysts in kimberlites are xenocrysts formed by the disaggregation of mantle nodules. Three distinct generations of olivine are commonly present: rounded megacryst (about Fo_{85}), phenocrysts ($>Fo_{90}$), and groundmass (Fo_{90-85}).

Kimberlites contain up to 50% xenoliths and xenocrysts derived from the mantle and crust. The mantle-derived ones fall into two classes: the eclogites, which are normally less abundant, and the more abundant ultramafic nodules, which have variable amounts of olivine but generally can be characterized as garnet lherzolite. Eclogite nodules are composed essentially of omphacitic clinopyroxene (NaAl) and pyrope garnet. Chemically they have the composition of basalt, and experiments by Yoder and Tilley (1962) indicated that they have a small melting range. These fragments, therefore, are not pieces of residual mantle, which would be expected to have a wide melting range. Instead, they appear to be the high-pressure crystallization product of basaltic magma or metamorphosed basaltic rocks that have been subducted.

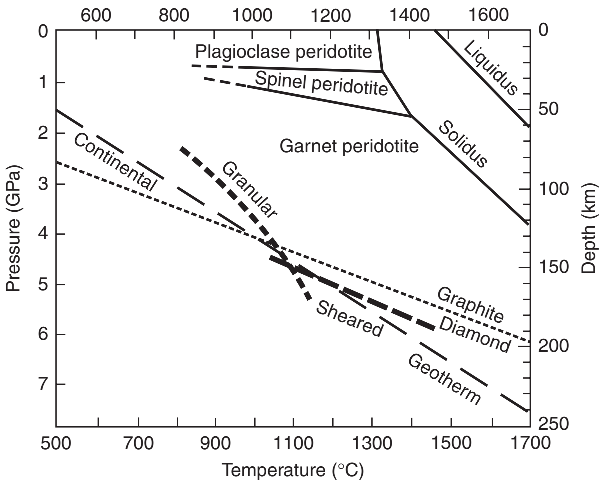


Fig. 15.20 Equilibrium temperatures and pressure of sheared and granular lherzolite nodules in kimberlites from South Africa (after Boyd and Nixon, 1975). Stability fields of plagioclase-, spinel-, and garnet-peridotites and their anhydrous liquidus and solidus are from Kushiro and Yoder (1966), and the diamond–graphite equilibrium boundary is from Bundy *et al*. (1961). The continental geotherm is from Figure 1.9 and Problem 1.4.

The ultramafic nodules include, in addition to garnet lherzolite, garnet harzburgite, garnet pyroxenite, spinel peridotite, and dunite. These come from a wide range of depths, as indicated by the presence of spinel in some and garnet in others. The assemblage plagioclase + olivine, which occurs in basalts at low pressure, is replaced at higher pressures by spinel + pyroxene according to the reaction

$$\underset{\text{(olivine)}}{2Mg_2SiO_4} + \underset{\text{(plagioclase)}}{CaAl_2Si_2O_8} = \underset{\text{(opx)}}{2MgSiO_3} + \underset{\text{(cpx)}}{CaMgSi_2O_6} + \underset{\text{(spinel)}}{MgAl_2O_4} \qquad (15.3)$$

At still higher pressure, spinel is replaced by garnet according to the reaction

$$\underset{\text{(opx)}}{2MgSiO_3} + \underset{\text{(cpx)}}{CaMgSi_2O_6} + \underset{\text{(spinel)}}{MgAl_2O_4} = \underset{\text{(garnet)}}{CaMg_2Al_2Si_3O_{12}} + \underset{\text{(olivine)}}{Mg_2SiO_4} \qquad (15.4)$$

These reactions give rise to the stability fields shown in Fig. 15.20.

Because kimberlites are emplaced rapidly, minerals in the mantle nodules do not have time to reequilibrate on rising to the surface – diamond, for example, is not converted to graphite. By determining the pressures and temperatures at which the minerals in these nodules equilibrated, the geothermal gradient in the mantle through which the kimberlite passed can be determined. Boyd and Nixon (1975) have done this for nodules in South African kimberlites. Garnet lherzolite nodules contain both Ca-rich and Ca-poor pyroxenes, making it possible to estimate the temperature of equilibration from the pyroxene solvus

(Section 10.20 and Fig. 10.32). Once the temperature is known, the pressure is estimated from the Al_2O_3 content of orthopyroxene coexisting with garnet. The results of their study (Fig. 15.20) indicate that the nodules formed at pressures and temperatures close to the geothermal gradient that would be expected beneath a stable craton (Section 1.6) and at pressures just above and below the diamond–graphite equilibrium boundary. This is consistent with the fact that only some of the kimberlites contain diamond.

Boyd and Nixon's study also revealed that the garnet lherzolite nodules fall into two distinct classes: those having a coarse-grained granular texture and those with finer-grained porphyroclastic and mosaic textures, which indicate shear. The granular nodules all come from depths between 100 and 150 km, whereas the sheared ones come from depths of 150 to 200 km. They interpret the boundary between these two types as being the top of the asthenosphere. The two types have slightly different compositions, the granular ones being more depleted in incompatible elements (except K) than the sheared ones. This suggests that the shallower mantle has had melt extracted from it.

Another significant difference is that the granular nodules define a distinctly shallower geotherm than the sheared ones. Because a sharp inflection in the geotherm could not be a steady-state phenomenon, Boyd and Nixon interpret the steeper gradient in the mantle from which the sheared nodules came to be the result of a thermal perturbation in the rocks below 150 km. They suggest that the heating could result from the shearing itself. It is also possible that the perturbation was related to the magmatic event responsible for the production of the kimberlites, which must have come from a depth of at least 200 km.

With contents of SiO_2 of about 30% and of MgO of about 23%, kimberlites are clearly ultramafic rocks. They differ from other ultramafic rocks, however, in having high concentrations of TiO_2, K_2O, P_2O_5, and other incompatible elements. Their high Cr and Ni contents indicate that they have not undergone any significant fractionation of olivine. They must therefore be derived by small degrees of partial melting of fertile mantle; that is, mantle from which little if any melt has previously been extracted. Zone melting may also have played a role in developing the high concentrations of incompatible elements.

An interesting, but rather rare group of volcanic rocks, the *lamproites*, may be related to kimberlites. In northwestern Australia, for example, they occur with kimberlites and also contain diamonds and mantle xenoliths. Lamproites are potassium- and magnesian-rich rocks that typically contain leucite and phlogopite. Olivine-rich varieties, which have compositions approaching kimberlite, are the ones that contain diamond.

15.7.3 Carbonatites

The most silica-undersaturated rocks found in alkaline provinces are the carbonatites. The term Karbonatite was introduced by Brögger in 1921 to denote carbonate rocks from the Fen district of southern Norway, which he believed were of igneous origin. The idea of magmatic carbonates met with immediate opposition from Bowen, who thought that the carbonates were of a replacement origin. Despite experimental evidence that such melts could exist at low temperatures and pressures (Wyllie and Tuttle, 1960), a magmatic origin for carbonatites was not universally accepted until carbonate lavas were witnessed erupting from the Oldoinyo Lengai volcano in Tanzania (Fig. 15.19) (Dawson, 1962). Reviews of the history of ideas relating to carbonatites and coverage of related field and experimental studies can be found in the books by Heinrich (1966), Tuttle and Gittins (1966), and Wall and Zaitsev (2004).

Carbonatites are found in most alkaline provinces throughout the world, but like kimberlites, they appear to be restricted to continental regions, with a few possible exceptions (Cape Verde Islands, for example). They are closely associated with rift valleys and can have ages from Proterozoic to modern, as for example: Paloboro in the Transvaal (Proterozoic); Mountain Pass, San Bernardino, California (1.1 Ga); Alno Island in the Baltic, Sweden (0.56 Ga); Fen, southern Norway (0.56 Ga); Minas Gerais, Brazil (Jurassic); Oka, Quebec (Cretaceous); Magnet Cove, Arkansas (Cretaceous); Bearpaw Mountains, Montana (Eocene); Kaiserstuhl, Baden, Germany (Miocene); and Oldoinyo Lengai (modern – 1960).

Most carbonatites occur in small ring complexes (Fig. 4.74) associated with silica-poor rocks composed essentially of nepheline and clinopyroxene but also containing variable amounts of melilite, melanite garnet (Ti-rich), perovskite, apatite, and opaque oxides. *Ijolite*, which is composed of nepheline and aegerine–diopside, grades to *melteigite* (pyroxene-rich) and *urtite* (nepheline-rich). Rocks composed largely of melilite and titanaugite are called *melilitites*, and they grade from *okaite* (melilite-rich) to *jacupirangite* (augite-rich). Nepheline syenite may also be present. The volcanic equivalent of ijolite is nephelinite, and at the world's only active carbonatite volcano, Oldoinyo Lengai, it is interlayered with carbonatitic tuffs and capped by the recent natrocarbonatite lavas. These occurrences leave no doubt as to the close association of carbonatitic and nephelinitic magmas.

In many carbonatite complexes, the concentrically arranged rock types become progressively poorer in silica toward the core, which is commonly occupied by carbonatite, but this pattern can be disturbed by multiple intrusions. A typical succession of rock types from rim to core would consist of nepheline syenite, ijolite, and carbonatite, with all rocks being cut by lamprophyric dikes of alnöite or fourchite. The carbonatite itself may also be concentrically zoned from an older, outer zone of calcite carbonatite (*sovite*), followed by a zone of dolomite carbonatite (*beforsite*) and a younger core of ankerite or siderite carbonatite. Based on experimental studies, such a sequence is consistent with the fractional crystallization of a carbonate melt.

The typical carbonatite is composed of approximately 75% carbonate (calcite, dolomite, ankerite, siderite), with lesser amounts of clinopyroxene, phlogopite, alkali amphibole, apatite, magnetite, olivine, monticellite, perovskite, and pyrochlore [$(Ca,Na)_2(Nb,Ta)_2O_6(OH,F)$]. Carbonatites are highly variable in composition, even down to the scale of a hand specimen, where modal variations almost always define a flow layering. Carbonate grains may be elongate parallel to the layering, but many have been recrystallied into smaller polygonal grains. Original grain sizes may still be evident, however, from the distribution of other minerals, such as apatite, which outlines the original grains. Phlogopite crystals may be zoned in shades of pale green with orange-red rims.

The rocks surrounding virtually all carbonatite complexes have undergone intense sodium metasomatism; that is, their compositions have been changed by the addition of sodium. Silicic country rocks may be converted to aegerine-bearing syenites that are termed *fenites*. Zones of *fenitization* are typically hundreds of meters wide. The prevalence of such metasomatized rocks around carbonatite complexes indicates that large volumes of alkali-bearing solutions are given off during cooling.

Ancient intrusive carbonatites are composed predominantly of calcite, whereas modern volcanic ones contain abundant sodium carbonate. Rainwater rapidly dissolves sodium carbonate, so it is not surprising that natrocarbonatite lavas are unlikely to be preserved in the geologic record. However, the lack of evidence for sodium carbonate in plutonic complexes has puzzled petrologists. One explanation is that carbonatite magmas do indeed have high concentrations of sodium carbonate, but during solidification and cooling, hydrothermal solutions remove the sodium from the carbonatite (leaving it composed essentially of calcite) and transporting the sodium into the country rocks to form fenites. This idea was strongly favored when it appeared from experimental studies that only sodium carbonate melts could separate as immiscible liquids from silicate magmas. In view of the work by Kjarsgaard and Hamilton (1988) showing that the immiscibility field extends to low alkali compositions (Fig. 14.29), it is not necessary for all carbonatites to be initially sodium-rich and then be leached of alkalis. They can form directly as calcite carbonatites.

Some carbonatites are of economic value as ores of niobium and rare earths. Initially carbonatites were mined only for iron and limestone, which is used for cement and as a flux in smelting iron. Their potential as the major source of Nb was not recognized until the 1950s. Although most carbonatites contain accessory amounts of pyrochlore, some contain zones with high concentrations (about 0.4% Nb_2O_5). These zones are independent of the shape of the carbonatite in which they occur, and they may even be discordant. The Nb enrichment probably takes place during a late magmatic or deuteric stage. REE and Th can also be present in economic concentrations, occurring mainly in perovskite, but also to a lesser extent in pyrochlore. The more highly differentiated ankeritic and manganiferous carbonatites typically contain the highest concentrations of REE.

Carbonatites, as a group, have exceptionally high concentrations of Ti, Nb, Zr, REE, P, F, Ba, Sr, and Th. These elements are abundant in alkaline magmas in any case, but it appears that during the process of generating carbonatites they are further concentrated, perhaps through strong liquid–liquid partitioning into an immiscible carbonate liquid (Section 14.7). Once fractional crystallization of this liquid takes place, Sr and Nb are depleted by entering early-crystallizing carbonate and pyrochlore, respectively, and the residual liquid becomes enriched in the other elements.

Strontium isotopic data clearly demonstrate that carbonatites are derived from the same mantle source as the associated alkaline rocks, both sharing low initial $^{87}Sr/^{86}Sr$ ratios (0.701 to 0.704). They are clearly not the product of limestone assimilation, for sedimentary limestones have initial values of about 0.709. Carbonatites are probably derived from a variety of alkaline magmas. Dawson and Hawthorne (1973) have shown that, at least on a small scale, carbonatite can form as an immiscible liquid from kimberlitic magma, as occurred in the Benfontein Sill, South Africa, where carbonatite forms small diapirs that rose into the kimberlite. Many alnöites contain abundant calcite, and fourchites, monchiquites, and camptonites contain carbonate ocelli, which may be quenched immiscible droplets (Problem 15.11). The carbonate globules trapped in apatite crystals in the Usaki complex of West Kenya (Rankin and Le Bas, 1974) show that immiscible carbonate liquids can separate from ijolitic magmas. It is likely, therefore, that carbonatite liquids can separate from silicate melts over a wide range of compositions. The experimental results of Kjarsgaard and Hamilton (1988) indicate that silicate melts encountering the immiscibility field at an early stage of fractional crystallization (low Na) would separate calcium-rich carbonate melts (Fig. 14.29), whereas more evolved silicate melts would separate more sodium-rich carbonate melts.

15.8 SPECIAL PRECAMBRIAN ASSOCIATIONS

One of the tenets of geology, first formulated by Hutton and widely publicized by Playfair and Lyell, was the doctrine of *uniformitarianism – the present is the key to the past*. This principle has helped geologists unravel much of the geologic record. Conditions during the earliest part of the Earth's history, however, were distinctly different from those during later time. As a result, we find that certain rocks were formed only during the Precambrian and at no other time. Without modern analogues, these unique Precambrian rocks present a special petrologic challenge.

Radioactive elements were more abundant in the Archean than in later time, and they would have generated more heat, thus creating a steeper geothermal gradient. This, in turn, would have resulted in a thinner lithosphere (Tarling, 1980). The geotherm would have intersected the beginning of melting curve for mantle materials over a wider range of pressures, allowing melting to occur at both shallower and greater depths than in more recent times. Mantle convection would have been more vigorous, and lithosphere was probably created and destroyed more rapidly. On being subducted, thin oceanic lithosphere may not have retained its water long enough to flux the mantle above the deep Benioff zone, thus decreasing the quantity of calcalkali rocks formed, especially the more potassic varieties. Bombardment by meteorites would have been more intense, with the larger ones creating huge impact structures. Unlike the Moon, the Earth, with its active tectonism, has tended to destroy the evidence of these ancient impacts, but some are still preserved, especially in stable Precambrian shields.

15.8.1 Archean crust

Shield areas on each of the main continents have a nucleus of Archean rocks (>2.3 Ga), some of which, despite their age, are surprisingly well preserved. Indeed, some sedimentary and volcanic rocks have hardly been metamorphosed, and they provide a reasonably good picture of the surface conditions at the time. At the same time, the igneous rocks provide some insight into the thermal status of the upper mantle and lithosphere. These rocks indicate that the Archean world differed significantly from that of later times.

First, the Earth's atmosphere had a very different composition. Gases liberated by the earliest volcanoes developed an atmosphere composed largely of H_2O, CO_2, CO, SO_2, H_2S, and N_2 (Section 11.1). This atmosphere lacked oxygen until the end of the Archean. Ultraviolet light from the sun caused some dissociation of water vapor to hydrogen and oxygen, but this small amount of oxygen would have combined with elements such as iron ($2Fe + O_2 = 2FeO$) to form oxides and silicates. Not until the appearance of green plants and photosynthesis was oxygen produced in sufficient quantity to oxygenate the atmosphere. Stromatolitic limestones formed from algae that were capable of carrying out this process are found in rocks as old as 3.0 Ga, and they are believed to have produced sufficient oxygen in the atmosphere by about 2.7 Ga for primitive oxygen-respiring bacteria to develop. These algae become particularly abundant in the geologic record by 2.3 Ga, which is normally taken as the beginning of the Proterozoic era.

The lack of oxygen in the atmosphere throughout most of the Archean allowed iron to be transported in surface waters in the soluble ferrous state. Today ferrous iron is rapidly oxidized by the atmosphere to the insoluble ferric state, which gives the brown and red colors to weathered rock surfaces and soils. The greater solubility of iron in Archean times was undoubtedly a factor in making sedimentary iron formations a common rock type (Klein, 2005). *Banded iron formations* (BIFs), which consist of millimeter- to centimeter-thick layers of hematite, jasper, or chert, are common in the Archean and reach a maximum abundance at ~2.5 Ga. They disappear from the record at 1.8 Ga but reappear briefly between 0.8 and 0.6 Ga, and are virtually nonexistent in Phanerozoic time. Other sedimentary rocks found in Archean terranes are stromatolitic limestones and clastic sedimentary rocks, which are important because they indicate that some granitic crustal material was present and available for weathering. The metaconglomerate of the Jack Hills of Western Australia contains 4.4-Ga-old zircons whose oxygen isotopes indicate that a continental crust that was undergoing weathering may have existed at this early date (Wilde *et al*., 2001). Unlike Phanerozoic sedimentary sequences, however, Archean ones are never thick, which indicates the lack of long-lived subsiding sedimentary basins.

Archean sedimentary rocks occur in small basins with a variety of volcanic and intrusive rocks, which have typically been metamorphosed to the green schist facies (Section 16.4). They are consequently referred to as *greenstone belts*. These belts form short arcuate outcrop patterns that have a general parallelism within a given province, but they lack the long continuous patterns of Proterozoic and Phanerozoic mountain belts. The thinner sequences of sedimentary and volcanic rocks, the discontinuous nature of the fold belts, and shorter wavelengths of folds have been interpreted as indicating that the lithosphere was thinner at the time and thus buckled more easily. Greenstone belts have also been interpreted as deformed basins that were initially formed by meteorite impact (Green, 1972). On the Moon, large impact structures have been flooded with basalt to form maria. On Earth, such structures would have been metamorphosed and deformed.

The igneous rocks of greenstone belts, which typically constitute more than 50% of a belt, are generally basaltic, but other compositions ranging from ultramafic through basaltic to andesitic and rhyolitic are also present. Some lavas of ultramafic compositions, known as *komatiites*, are found only in Archean terranes (see next section). In general, the basaltic and ultramafic rocks form the lower part of the succession, with intermediate and silicic ones forming the upper part.

Between the greenstone belts are vast stretches of granitic rocks, which form plutons and gneissic diapirs. In contrast to the greenstone belts, these granitic rocks are intensely metamorphosed at upper amphibolite or granulite facies grades. They are characterized by a low K/Na ratio; that is, they are trondjhemites or tonalites. They are typically gray and constitute the so-called *gray gneisses*, which are so characteristic of Archean terranes, such as the gneiss at Isua, west Greenland, which for a time held the record for the Earth's oldest rock (3.8 Ga).

The K/Na ratio in igneous rocks has increased through time (Engel *et al*., 1974). The preponderance of trondjhemites in the Archean is responsible for the low K/Na ratio during that time, but toward the end of the Archean the K/Na ratio increases rapidly as a result of the widespread emplacement of more normal granites. The Archean gray gneisses seem to be part of a calcalkali suite. Why more potassic varieties did not develop is a puzzle. Perhaps insufficient continental crust, which typically concentrates the strongly incompatible potassium, had not yet formed. On the other hand, in view of the distribution of Na and K rocks above modern Benioff zones, perhaps melting occurred at too shallow a depth.

The major and trace element chemistry of the gray gneisses of the Archean basement of Finland indicate that they must have been derived by the partial melting of tholeiitic basalt (Martin, 1987). They have low initial $^{87}Sr/^{86}Sr$ ratios of 0.702 to 0.701, which indicates derivation from the mantle and not from a reworked crust. The REE data, however, do not allow generation of the gray gneisses directly from the mantle. Instead, a period of REE enrichment during the formation of a tholeiitic basalt is necessary. This basalt must next have been metamorphosed to a garnet-bearing amphibolite and then partly melted. A residue of hornblende, garnet, plagioclase, clinopyroxene, and ilmenite must have existed in order to generate the major and trace elements of the gray gneisses.

The presence of hornblende and garnet in the source from which the gray gneisses came is of great importance and indicates a fundamental difference between these rocks and later calcalkali ones. The geothermal gradient must have been much steeper in the Archean for subducted basalt to be metamorphosed to garnet amphibolite and then melt before dehydration could occur. Today, hydrated ocean floor basalts lose their water to the overlying mantle wedge long before the basalts can actually melt.

15.8.2 Komatiites

The ultramafic lavas known as komatiites are restricted almost entirely to the Archean greenstone belts. They were first recognized in Zimbabwe and then described in detail from the Komati River, in the Barberton Mountainland of South Africa (Viljoen and Viljoen, 1969). They are now also recognized from the Archean terranes of Canada, Western Australia, and Finland. Some komatiite-like rocks, having ages of 88–90 Ma, have been reported from Gorgona Island off the coast of Colombia, from Costa Rica, and Curaçao in the Caribbean (Alvarado *et al*., 1997). Apart from high MgO contents, their most striking feature is a texture found in the upper part of many flows, consisting of long criss-crossing sheafs of olivine crystals that radiate or fan out downward into the flow. This is referred to as the *spinifex* texture because of its resemblance to inverted tufts of an Australian grass by that name. Komatiites have proved to be of economic importance, because some contain nickel sulfide deposits.

Komatiites are defined as ultramafic volcanic rocks that contain at least 18% MgO (Arndt and Nisbet, 1982); many contain as much as 33% MgO. They occur with tholeiitic basalts but never constitute more than 10% of the volcanic rocks. One of the best preserved successions of komatiite lavas is in Munro Township, Ontario, where they form part of the Abitibi greenstone belt (Pyke *et al*., 1973). Here 60 ultramafic flow units are exposed over a stratigraphic thickness of 125 m, with the thickest being 15 m, and the average, 3 m. Most flows have a characteristic stratigraphy, which is illustrated in Figure 15.21. The top of each flow is marked by prominent fracturing. Immediately beneath this is the spinifex zone, with olivine blades increasing in length downward and reaching a maximum length of 1 m in the thickest flow. The olivine blades, which have a marked skeletal form (Fig. 12.7(A)), constitute approximately 50% of the rock, with the interstices being occupied by skeletal clinopyroxene and altered glass. The spinifex zone terminates abruptly against a fine- to medium-grained peridotite that forms most of the lower part of the flow. Immediately beneath the spinifex zone in some flows is a zone containing elongate olivine dendrites with a strong foliation parallel to the flow. Even in the main lower peridotite part of the flow a weak foliation is commonly present. Toward the base of the peridotite zone the flow develops a knobby weathering.


Fig. 15.21 Steeply dipping 1-m-thick komatiite flow, Munro Township, Ontario. The top of the flow, which is at the extreme left, is marked by a prominent zone of fracture. Beneath this is dark-weathering peridotite with spinifex texture (long blades of olivine), which terminates abruptly against a fine- to medium-grained pale-weathering peridotite. Toward the base of the flow is a dark, knobby-weathering peridotite. The hammer is lying on the lower contact of the flow, which is most clearly marked by the beginning of the fractured flow top of the underlying flow. (Photograph by J. M. Duke.)

The spinifex texture, with its skeletal olivine crystals that nucleate and grow down from the upper quenched surface of the flow, is an ideal example of crystal growth controlled by the rate at which heat is dissipated from the growth surface (Section 12.3). These olivine quench crystals constitute important textural evidence that komatiites are indeed ultramafic lavas and are not basaltic ones in which olivine crystals were concentrated. Olivine in the spinifex zone definitely grew as quench crystals in the melt. The fact that the blades of olivine are not broken or bent and can extend down through as much as two-thirds of the thickness of a flow indicates that little, if any, differential movement occurred in this part of the flow during the growth of the olivine blades. This has been taken as evidence that the olivine grew *in situ*. However, the velocity profile calculated in Problem 2.9 indicates that little differential movement occurs in the upper part of a flow in which there is a temperature gradient; most of the movement is concentrated in the central hot zone. The spinifex zone may have been rafted along on the hotter interior, which may have formed the strongly foliated peridotite immediately beneath the spinifex zone.

Experiments indicate that komatiitic magma has a high liquidus temperature, in excess of 1650 °C at atmospheric pressure (Green, 1975). It is not surprising then that such hot magma, when erupted on the Earth's surface, would experience rapid cooling both by radiation and conduction, and that quench crystals would likely develop, at least in the upper part of flows. Passage of such hot magma through a continental crust is likely to have caused partial melting and assimilation. Numerous geochemical studies have shown komatiites to have highly variable concentrations of incompatible elements. This could indicate derivation from inhomogeneous mantle, but it is more likely that komatiites all come from depleted mantle but have suffered variable degrees of contamination while passing through the crust (e.g. Barley, 1986). Huppert and Sparks (1985) calculate that the rate of meltback of conduit walls by komatiitic magma could have been as high as tens of meters per day. These hot, low-viscosity magmas, on erupting, would have flowed turbulently across the Earth's surface (Problem 15.12). Turbulent flow maintains a lava at a relatively constant temperature throughout, which causes considerable heating of the underlying rock. Huppert *et al*. (1984) have shown that this heat would cause sediment beneath a komatiite flow to melt, and the lava would eventually flow in a thermal erosion channel of its own making. At Kambalda in Western Australia such thermal erosion has led to the development of nickel sulfide ores at the base of the komatiite flows. The assimilated sediment contained sulfur, which on entering the komatiite lava, combined with iron and nickel to form a dense immiscible sulfide liquid that sank to the bottom of the channel.

The existence of such high-temperature lavas is strong evidence that the Archean geothermal gradient was steeper than that in later times. Today, partial melting of peridotitic mantle generates basaltic to picritic magmas, not komatiitic ones. To increase the MgO content to the levels in komatiites, the degree of partial melting would have to be increased to more than 50%. This, however, presents a problem, because melt instantly segregates from residual solids once the degree of partial melting exceeds 30%; the 50% melt required to generate komatiitic liquid in this way could therefore not be achieved. Takahashi and Scarfe (1985), however, have found from experiments that the normative olivine content of the anhydrous melt that first appears above the solidus of mantle peridotite systematically increases with increasing pressure. At low pressure this melt is basaltic, but at 4 GPa it contains 20% MgO, and at 7 GPa it contains >30% MgO. They conclude therefore that basaltic magmas are generated at depths of less than 100 km, whereas komatiitic ones are formed between 150 and 200 km.

Despite the presence of komatiites in Archean greenstone belts, basalts are still by far the most abundant volcanic rock. The conditions in the upper mantle must therefore have been capable of producing both basaltic and komatiitic magmas. With a steeper geothermal gradient, the depth range over which the geotherm would intersect the peridotite solidus would have been greater than in later times, thus allowing basalts to form at shallow depth and komatiites at greater depth. In post-Archean times, the shallower geotherm would have intersected the solidus over a shorter depth interval and apparently at depths too shallow to produce high-magnesium melts. The young komatiitic rocks of central America have been interpreted to indicate that the geothermal gradient in that region was elevated by a mantle plume (Alvarado *et al*., 1997).

15.8.3 Anorthosites

Anorthosites are rocks composed essentially of plagioclase feldspar. They occur as cumulate layers in large, layered gabbroic intrusions, where they typically have compositions in the bytownite range. They also occur in large Proterozoic massifs, where there are no mafic or ultramafic rocks; these typically have an intermediate plagioclase in the labradorite or andesine range. These bodies, which rival the large layered lopoliths and granite batholiths for size, represent a remarkable and unique period of magmatism that occurred between 1.45 and 1.1 Ga. There are only two Phanerozoic massif-type anorthosites. One Cambrian body at Sept Iles, Quebec, looks identical to nearby Proterozoic ones (Fig. 15.22). It occurs on the margin of the Canadian Shield and is associated with alkaline magmatism and rifting. The other body is of Silurian age and occurs in the alkaline ring complexes of Niger. Anorthosites also occur in Archean terranes, where they typically form lenses in high-grade gneisses. Some of these may be part of metamorphosed layered gabbroic intrusions, whereas others may have formed by the metamorphism of sedimentary rocks. They differ from the Proterozoic anorthosites in having extremely calcic plagioclase. Anorthosites are believed to make up much of the lunar highlands. They were possibly formed early in the Moon's history by plagioclase crystals floating and accumulating on the top of basaltic magma. It is the Proterozoic massif-type anorthosites that are dealt with in this section.

Proterozoic anorthosites lie in a belt that extends across Pangea from Scandinavia, through southern Greenland, Labrador, Quebec, and New York State (Fig. 15.22), and beneath the sedimentary cover across the craton to Wyoming and California. A second, less well defined belt may extend from Virginia, through Africa, Madagascar, and India, to Australia. A collection of works describing these bodies has been edited by Isachsen (1968), and partisan reviews have been written by Morse (1982), Duchesne (1984), and Emslie (1985).

Most massif-type anorthosites in the Grenville province form batholithic bodies that appear to have been emplaced diapirically (Philpotts, 1981). They consequently do not reveal floors, and thus we do not know what underlies them. Gravity surveys do not indicate the presence of ultramafic rocks beneath anorthosites, but this is not to say that such rocks might not occur at much greater depth and were simply too dense to rise with the anorthosite. Most anorthosites are emplaced into upper amphibolite or granulite facies metamorphic rocks. The marginal zones of these bodies may be intensely deformed and recrystallized as a result of the diapiric emplacement, but toward their cores, primary textures are preserved that show massif anorthosites to have formed initially as igneous rocks. In Labrador, north of the Grenville front, anorthosites appear to have formed *in situ* rather than being emplaced diapirically.

Within most anorthosite massifs, the rock types have a distinct primary stratigraphy (Fig. 15.23). If the body has been emplaced diapirically, these rocks may be disposed

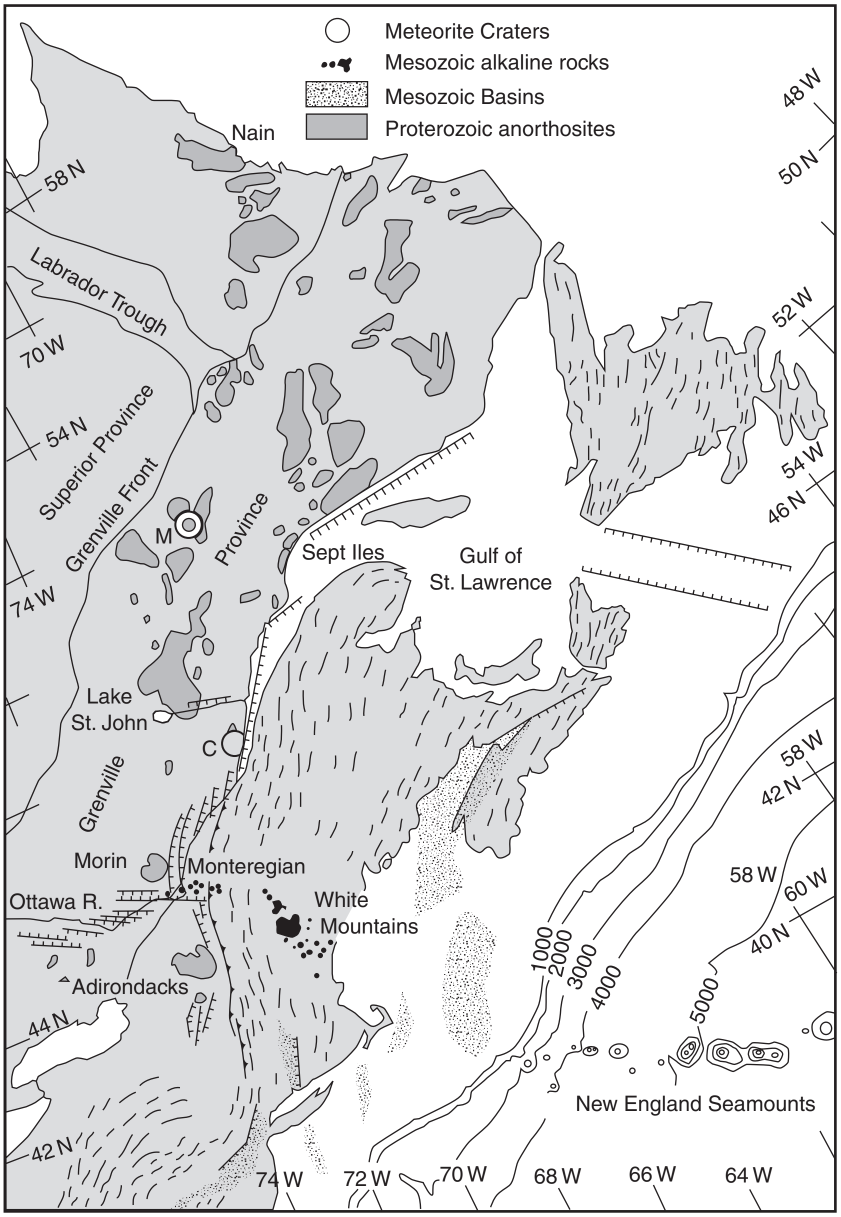


Fig. 15.22 Massif-type anorthosites of the southeastern part of the Canadian Shield. Those near Nain have ages of 1.45 Ga, whereas many of those south of the Grenville Front have ages as young as 1.1 Ga, with the exception of the Sept Iles anorthosite, which is only 0.5 Ga. The New England Seamounts may mark the trace of a hot spot that earlier formed the Mesozoic igneous rocks of the White Mountains. The Mesozoic basins (stippled areas) formed during crustal extension associated with the opening of the Atlantic. Also shown are two large meteorite impact structures, Manicouagan (M) and Charlevoix (C).

in annular zones with the oldest in the core. The lowest exposed rock is invariably anorthosite, containing only small amounts of orthopyroxene, augite, or olivine. Variations in the proportions of these minerals can produce layering on the scale of tens of centimeters to meters in thickness. This layering originally appears to have been horizontal. More commonly, the anorthosite appears massive. One of the most characteristic features of the Proterozoic anorthosites is their coarse primary grain size (Fig. 15.24). Plagioclase crystals several centimeters long are common, and they may even reach a meter. This coarse grain size is not due to pegmatitic fluids. Indeed, all indications are that anorthosites formed under anhydrous conditions. Plagioclase crystals are commonly of two different sizes. The largest crystals, which typically have lengths of several centimeters, occur separately or in clumps that are surrounded by centimeter-length laths or granules that are typically intergrown with ferromagnesian minerals. The granular plagioclase, which is prevalent in the Grenville anorthosites, is interpreted as being *protoclastic*, that is, broken and recrystallized prior to complete solidification of the magma. This texture indicates that these anorthosites were emplaced as a crystal mush. The primary plagioclase is invariably darkened by small oriented inclusions of iron oxide. Recrystallization associated with later tectonic activity, which presumably took place at lower temperature, expels these inclusions and turns the plagioclase white. Anorthosites also commonly contain megacrysts of aluminous bronzite (Fig. 15.24(B)), which contain between 7% and 9% Al_2O_3 and indicates crystallization at high pressures. Xenoliths of country rocks are rare, but some anorthosites enclose meter-sized blocks of other anorthosite that differs slightly in mineralogy or texture.

The content of ferromagnesian minerals (opx normally most abundant) in anorthosites increases upward as the rock grades into *leuconorite*, which in many bodies is the main rock type. Undeformed samples of this rock have a prominent alignment of plagioclase laths parallel to any compositional layering that is present (Higgins, 1998). Toward the top of the leuconorite the abundance of pyroxene, apatite, and iron–titanium oxides increases rapidly, and the grain size decreases to form a ferrodioritic rock, *jotunite*. Dikes of ferrodiorite may cut the underlying leuconorite. Occurring in this zone and in the underlying leuconorite may be sills and dikes of magnetite–ilmenite (rutile)–apatite rock (*nelsonite*), which in some areas has been mined for ilmenite. The jotunite zone, which is commonly only tens of meters thick, marks the top of the anorthositic part of an intrusion; below this plagioclase is the dominant mineral; above it alkali feldspar dominates. A variety of rock types may lie above the jotunite, but in general they have the composition of a hypersthene bearing quartz monzonite (*quartz mangerite*); quartz syenite and hypersthene granite (*charnockite*) may also be present. The contact between the jotunite and more silicic rocks is abrupt, but sheets of jotunite continue to occur for some distance above the main contact.

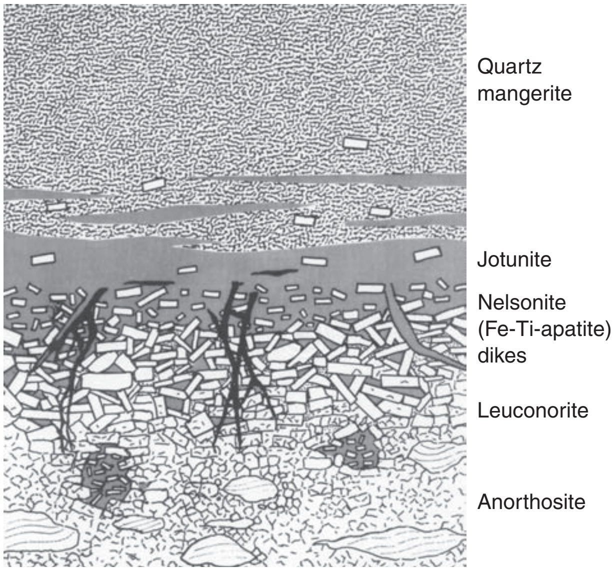


Fig. 15.23 Idealized stratigraphic section through a typical massif-type anorthosite. See Chapter 6 for definitions of rock names. The black dikes are of an iron–titanium oxide–apatite rock known as nelsonite, which intrudes downward into underlying anorthosite. (After Philpotts, 1981.)

Plagioclase in the anorthositic rocks typically has compositions between An_{60} and An_{30}, but within a given body it rarely varies by more than a few percent. For example, in the world's largest anorthosite body, which covers an area of 25 000 km^2 north of Lake St. John, Quebec (Fig. 15.22), all plagioclase falls in the labradorite range. To the southwest, plagioclase in the Morin anorthosite, located directly north of Montreal, is all in the andesine range. Plagioclase does become slightly more sodic in the jotunites, reaching An_{45} in those associated with labradorite anorthosites and An_{30} in those associated with andesine anorthosites. The composition of the plagioclase remains essentially the same in the quartz mangerites. Dark megacrysts of andesine, measuring up to several centimeters in diameter, commonly occur in quartz mangerites near anorthosite bodies.

By contrast, pyroxenes, which are minor components in most anorthositic rocks, vary widely in composition. Orthopyroxenes show a complete range of composition from En_{75} in the most primitive anorthosites to En_{30} (inverted pigeonite) in the quartz mangerite. This enrichment in iron takes place progressively from leuconorite, through jotunite, to quartz mangerite. The iron enrichment is sufficient in many quartz mangerites for the iron-rich pigeonite (now inverted pigeonite) to be replaced by the assemblage fayalite plus quartz. Coexisting augite in this rock series changes steadily from about $Ca_{47}Mg_{43}Fe_{10}$ in anorthosites to ferroaugite in quartz mangerites. The fact that the most magnesian

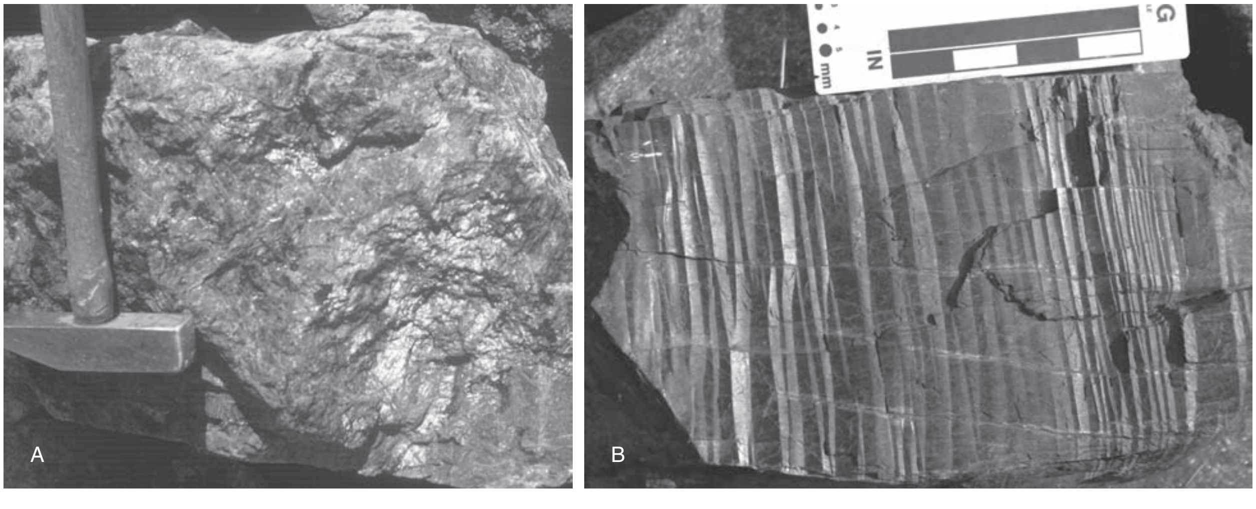


Fig. 15.24 (**A**) Single meter-sized crystal of plagioclase from the Nain anorthosite, Labrador (type locality for the mineral *labradorite*). Reflection of light is from (010) cleavage, which shows iridescent blues and greens (see book's web site, www.cambridge.org/philpotts, for color image). (**B**) Single megacryst of aluminous bronzite from Lake St. John anorthosite, Quebec. The crystal has kink bands as a result of deformation.

orthopyroxene in anorthosites is only En_{75} indicates that these rocks form from a magma with a magnesium number that is lower than that of basaltic magmas in layered gabbroic intrusions, which typically crystallize orthopyroxene as magnesian as En_{85}.

One of the major problems to unraveling the origin of massif-type anorthosites is the relation of the quartz mangerites to the other rocks. Are they fractionation products of the magma that produced the anorthosites, or are they crustal rocks melted by the heat given off by this magma? The answer to this question is critical to determining the composition of the parental magma of anorthosites. Because most anorthosite bodies lack chilled margins, the only way of estimating the composition of the parental magma is from a weighted average of all comagmatic rocks. This, of course, is difficult to do in the case of a body emplaced diapirically, because some of the original rocks may not be represented at the present level of erosion. Nonetheless, if this possibility is ignored, the problem remains of whether the quartz mangerites belong to the series. If they are omitted from the weighted average, the parental magma would have the composition of leuconorite; if they are included, the magma would be granodioritic.

The continuity to the variation in the composition of the major minerals in passing from anorthosite, through leuconorite and jotunite, to quartz mangerite suggests these rocks are comagmatic. But it is possible that a mantle-derived magma could rise into the crust by assimilating material at its roof and crystallizing plagioclase and pyroxene at its base. The rise of such a zone of melt into the crust could produce a progressive change in the composition of the minerals accumulating on the floor, even though most of the silicic fraction of the rock series would be derived from crustal rocks.

There is a simple test of whether quartz mangerites are fractionation products of an anorthositic magma or are partial melts of crustal rocks. Plagioclase is the primary cumulus phase in anorthosites and leuconorites, so fractionation products should show the effects of the early separation of this mineral. Because europium preferentially enters plagioclase, separation of this feldspar should produce negative europium anomalies in fractionated liquids (Section 14.12 and Fig. 14.42). Although some quartz mangerites have distinct negative europium anomalies, the majority do not. We can conclude, therefore, that although quartz mangerite can be a differentiation product of the magma from which anorthosites formed, the large volumes of quartz mangerite associated with most anorthosites are the products of partial melting of crustal rocks. Initial $^{87}Sr/^{86}Sr$ ratios support this conclusion.

Two of the most puzzling features of anorthosites are the unzoned nature of its plagioclase crystals and the relatively constant composition of plagioclase within a given body. Plagioclase crystals are always enriched in anorthite relative to the liquid with which they are in equilibrium. As crystallization proceeds, the liquid and the crystals become more albitic (Fig. 10.12), and if equilibrium is maintained the crystals should become homogeneous throughout. Diffusion rates in plagioclase, however, are so slow that once a plagioclase crystallizes, its composition is essentially fixed; that is, crystals have little chance of homogenizing, and as a result, plagioclase is commonly zoned (Fig. 15.8). The fact that even the largest plagioclase crystals in anorthosite are remarkably free of zoning calls for explanation.

Homogeneous plagioclase crystals can grow if the proportion of liquid is much greater than the proportion of crystals forming from it. This, however, cannot be the case for anorthosite massifs, because not only are individual crystals unzoned, but the plagioclase within a given complex shows little compositional variation. The process by which anorthosites form must therefore involve some mechanism of buffering the plagioclase at a relatively constant composition, but this composition can be different for different anorthosite complexes. Such a mechanism might, for example, involve special phase relations, or perhaps a physical process of magma replenishment or mixing.

As an example of one mechanism that could produce plagioclase of essentially fixed composition, we will examine the possible role of liquid immiscibility (Philpotts, 1981). Rocks of the anorthosite–quartz mangerite series, when plotted in variation diagrams such as those based on silica content, differentiation index, or magnesium number (see Section 14.2), show a continuous spectrum of compositions. When these same data are plotted in terms of the three groups of components SiO_2–($FeO + MnO + TiO_2 + CaO + P_2O_5$)–($MgO + Al_2O_3 + Na_2O + K_2O$), the series is seen to have a distinct fork to it (Fig. 15.25(A)). Rocks from anorthosite through leuconorite to jotunite are marked by strong enrichment in iron and other high-field-strength cations. The rocks from leuconorite through mangerite and quartz mangerite to charnockite, on the other hand, show a decrease in iron and an increase in silica. Thus there seem to be two contrasting trends of fractionation: one toward iron-rich residues (Fenner trend) and the other toward silica-rich residues (Bowen trend). When compared with the liquid immiscibility field in the system fayalite–leucite–silica (Roedder, 1951) and its expanded form resulting from the addition of 3 wt% TiO_2 and 1% P_2O_5 (Freestone, 1978), these two trends skirt the two-liquid field, as if they were composed of conjugate liquids. Experiments (Philpotts, 1981) reveal that quartz mangerite and jotunite produce immiscible liquids on melting, just as do the granophyre and ferrodiorite of the Skaergaard (McBirney, 1975; Jakobsen *et al*., 2005).

Figure 15.25(B) shows the results of an experiment where a homogeneous, high-temperature fused mixture of jotunite and quartz mangerite was held at 995 °C at atmospheric pressure. Plagioclase (An_{50}), augite, and pigeonite have crystallized, and the remaining melt has unmixed into silica-rich and iron-rich liquids. In this one experiment, we have the necessary phases to form the entire anorthosite series; that is, anorthosite from the plagioclase, jotunite from the iron-rich liquid, and quartz mangerite from the silicic liquid. As plagioclase and pyroxene crystallize in this phase assemblage, the liquids do not change their compositions significantly; instead, the two liquids simply change their proportions. As a

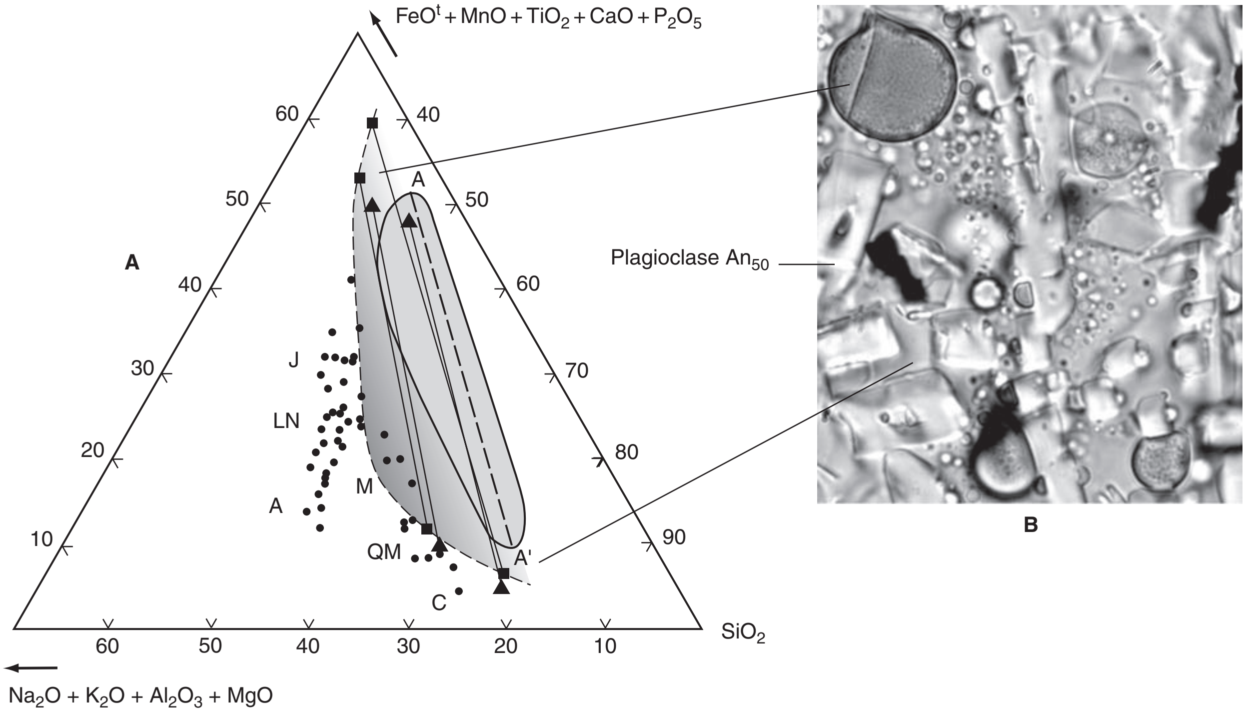


Fig. 15.25 (**A**) Plot of the composition of the rocks associated with massif-type anorthosites; A, anorthosite; C, charnockite; LN, leuconorite; J, jotunite; M, mangerite; QM, quartz mangerite. The oval shaded area is the two-liquid field from the system fayalite–leucite–silica (Roedder, 1951), and to its left (short-dashed line) is the expanded two-liquid field resulting from the addition of 3% TiO$_2$ and 1% P$_2$O$_5$ (Freestone, 1978). Tie-lines connect conjugate immiscible liquids in experiments (squares) and andesites (triangles). (**B**) Experimentally produced immiscible iron-rich and silica-rich liquids coexisting with plagioclase of An$_{50}$ composition formed by cooling a homogeneous fused mixture of jotunite and quartz mangerite to 995 °C at low pressure under an oxygen fugacity buffered by nickel–nickel–oxide. Diameter of large iron-rich globule is 15 μm. (After Philpotts, 1981.)

consequence, the plagioclase crystallizes with a constant composition, buffered by the two liquids. Note that in this model, the quartz mangerites could be derived by fractionation from the parental liquid of the anorthosites or by melting of crustal rocks. The buffering of the plagioclase requires only that two liquids be present, regardless of their ultimate source. Mechanisms other than liquid immiscibility could also produce relatively constant composition plagioclase. Any successful model for the generation of massif-type anorthosites must be capable of doing this.

Anorthosites remain one of petrology's greatest puzzles. With Phanerozoic igneous rocks, a particular geochemistry can be associated with a general tectonic setting – alkaline rocks with rifting, depleted tholeiites with ocean floor spreading, and calcalkali rocks with subduction. Anorthosites do not fit into any of these categories. Estimates of the composition of the parental magma vary widely, from basalt to anorthosite and granodiorite. The amount of crustal contamination is also hotly debated. But overshadowing all else is the problem that they formed almost exclusively in the short interval between 1.45 and 1.1 Ga, and then only in restricted belts. What unique conditions could have led to their formation?

Isotopic studies reveal anorthosites to have highly variable initial ^{87}Sr/^{86}Sr ratios, but there appears to be a distinct input of magma from the mantle. However, anorthosites do not resemble any other mantle-derived rocks, in that they are so feldspathic and have such low magnesium numbers. A mantle-derived magma would therefore have to fractionate before it could produce anorthosites. Experiments indicate that the eutectic between diopside and anorthite (Fig. 10.4) shifts to feldspar-rich compositions at high pressures (Yoder, 1965; Presnall *et al*., 1978). Fractionation of a mantle-derived basaltic magma in the lower crust could lead to the development of this feldspar-rich residual liquid. It is here that the aluminous megacrysts of bronzite might form. The feldspar-rich eutectic liquid would have a low density and would rise into the crust. As the pressure decreased, the liquidus field of plagioclase would expand, and plagioclase would become a prominent cumulus phase. As the magma rose, more and more plagioclase would form, converting the magma into a plagioclase mush. Such a medium would be capable of transporting the megacrysts of bronzite, which otherwise could not have been raised by a liquid magma (Problem 15.13).

All anorthosites have a strong positive europium anomaly, which is to be expected in view of the partition coefficients of the REE between plagioclase and magma (Section 14.12). These positive anomalies, however, must have their negative counterparts somewhere in the geologic record if the REE are to preserve their overall normal abundance distribution. Furthermore, the fact that anorthosites have positive europium anomalies indicates that anorthosites must be cumulates and cannot be liquid residues from a fractionation process. For example, if anorthosites formed from an anorthositic liquid that was generated from a basaltic magma by the separation of mafic minerals at depth there would be no reason for it to have a europium anomaly. However, an anorthositic liquid developed from the melting of a previous plagioclase cumulate would have a positive europium anomaly. But this simply pushes the problem of the origin of the anorthosite back one step. The question then is: where is the residue from the anorthosite cumulate that has the negative europium anomaly? So far, only a small number of quartz mangerites and jotunites have such anomalies. Perhaps other quartz mangerites had negative anomalies but large-scale assimilation of crustal rocks may have masked them. It is also possible that the residual liquid with the negative anomaly was more dense than anorthosite and sank. A similar relation would exist if anorthosites formed by accumulation of floating plagioclase at the top of a magma chamber. The stratigraphy to anorthositic rocks (Fig. 15.23), however, is not consistent with this model. Another complicating factor is that apatite crystallizes with a negative europium anomaly, which causes residual liquids to become enriched in europium relative to the other REE. This would tend to cancel the effect of plagioclase crystallization. Apatite becomes extremely abundant in jotunites, and thus rocks more evolved than this will have their REE patterns modified by this additional factor.

Anorthosites in the southern part of the Canadian Shield form a prominent belt extending from the coast of Labrador to the Adirondacks in New York State (Fig. 15.22), but within this belt anorthosites are of several ages. The distribution, sizes, shapes, and petrology of the anorthosites throughout this region are so similar that the entire belt appears to be one igneous province. Yet the anorthosites in Labrador have ages of 1.45 Ga, whereas most in the Grenville province are only 1.1 Ga or even 0.5 Ga if we include the body at Sept Iles, Quebec. Perhaps anorthosites were formed at depth throughout this region at 1.45 Ga and then reactivated and intruded to higher levels at later dates in the Grenville province. In fact, even in the Grenville, anorthosites may be of several ages. Anderson and Morin (1968), for example, argue on the basis of the presence of blocks of labradorite anorthosite in andesine anorthosite that andesine anorthosites are formed by the partial melting of labradorite anorthosites. Duchesne (1984) has pointed out that structural relations in the Grenville anorthosites are not consistent with them simply being older anorthosites that were tectonically reactivated during the Grenville orogeny; they must involve melting at this time.

The volumes of magma involved in the formation of anorthosite complexes are enormous; for example, anorthosites constitute 20% of the rocks in the belt extending from Labrador to the Adirondacks. They must therefore reflect a major thermal event (or episode). Are there any other significant occurrences in the geologic record that might give some indication of what this event was? In the vicinity of Lake Superior, Keweenawan basalts and the Mid-Continent gravity high were developing between 1.2 and 1.1 Ga, as were the Coppermine River basalts and the Muskox Intrusion in the Northwest Territories. Many Keweenawan basalts and diabases actually contain large xenoliths of anorthosite. To the northeast of Lake Superior, alkaline igneous rocks were formed between Kapuskasing and James Bay at 1.1 Ga, and just to the east of Labrador, alkaline rocks of the Gardar province in southwest Greenland were formed at 1.2 Ga. Crustal extension therefore seems to have become active shortly after the emplacement of the first Proterozoic anorthosites and remained active during the emplacement of the Grenville ones. In fact, Emslie (1985) interprets the anorthosite suite as a rift-related group of rocks. Widespread rifting at this time might reflect some major overturning of the mantle, which has never been repeated on the same scale. But this is pure speculation, and the solution to the anorthosite problem awaits new insight. This solution, when it comes, is likely to be of considerable significance to our understanding of the history of the Earth.

15.9 METEORITE-IMPACT-GENERATED ROCKS

Inspection of the surface of the Moon and many other planets reveals that cratering by meteorite impact has been a common occurrence in the solar system. The flux of meteorites reaching the surface of the Moon, and presumably its nearest planetary neighbor, the Earth, has decreased dramatically with time, as evidenced by the difference in the frequency of craters on the older lunar highlands and the younger maria. Most of the ancient surface of the Earth has been destroyed by tectonic processes, and none older than 4.4 Ga has yet been found. Nonetheless, it is in the cratonic areas, which have remained tectonically stable since the Proterozoic, that we have the greatest chance of finding evidence of meteorite impacts.

Meteorite craters are of two types. Small meteorites, on passing through the Earth's atmosphere, are slowed by friction, which causes them to glow brightly, and most completely evaporate before reaching the surface. Those that do penetrate the atmosphere cause little damage on landing, and the size of the crater, which may be up to several meters in diameter, depends on the material on which they fall. By contrast, large meteorites are not perceptibly slowed by the atmosphere, and they reach the Earth's surface with essentially the velocities they had in space, ~20 km s^{-1}. The energy dissipated on impact is so great that it causes an explosion, which excavates the crater (Fig. 15.26). These hypervelocity impact structures are the ones likely to be recognized in the geologic record, for they produce distinctive fracture patterns, rock types, and even minerals (French, 1998).

When a meteorite hits the Earth it has a kinetic energy that depends on its mass and velocity, that is, $^1/_2 mv^2$. Following impact, this energy is dissipated both as shock waves that fracture and displace the impacted crust to form the crater, and as heat, which vaporizes the meteorite and melts the underlying rock on the crater floor. Approximately 20% of the energy goes into heating and perhaps even more with very large impacts. Considerable insight into the formation of large explosion craters has been gained from underground

Fig. 15.26 Meteor Crater, Arizona. (Photograph courtesy of U.S. Geological Survey.)

nuclear testing (e.g. Short, 1966). If we define the distance from the explosion out to the point where the rock is completely crushed as R, the rock is vaporized to a radius of $0.05R$, melted to a radius of $0.07R$, and fractured to a radius of $2R$. The diameter of a crater is related to the kinetic energy of the meteorite by the relation

$$\text{diameter} = 9.47 \times 10^{-3}(1/2mv^2)^{0.294} \tag{15.5}$$

where all values are in SI units (m, kg, s). The depth R to which complete crushing of the rock occurs is approximately one-third the crater diameter.

Craters, whether formed by nuclear explosion or hypervelocity impact, have a common structure (Fig. 15.27). Because they are formed by explosion, they are almost perfectly circular and are not elongated parallel to the direction of flight of the meteorite. If the impacted rocks have marked anisotropic physical properties (prominent joints or schistosity, for example), fracturing during the explosion may produce a noncircular form. Meteor Crater, Arizona, for example, tends to be rectangular (Fig. 15.26). Explosion craters are quite shallow, the depth to the floor being only one-tenth their diameter. Material ejected during the explosion forms a blanket surrounding the crater. The size of the ejecta decreases rapidly away from the rim. If the crater penetrates stratified rocks, their fragments are deposited in the ejecta blanket in reverse stratigraphic order. Flat-lying sedimentary rocks may be tipped up on end or even completely overturned by the force of the explosion as in the case of the rocks surrounding Meteor Crater, Arizona. The floor of a crater is covered with breccia and melted rock (Fig. 15.28). In large impact structures, this zone of melt may be hundreds of meters thick. Beneath the breccia is a zone of intense fracturing. Certain rocks (e.g. quartzite, limestone) may develop a type of fracture known as *shatter cones*, which are believed to be diagnostic of impact structures; they have not been found associated with volcanic explosions. Shatter cones are conical fractures, with their apex pointing toward the explosion. They commonly form interpenetrating sets as seen in Figure 15.29.

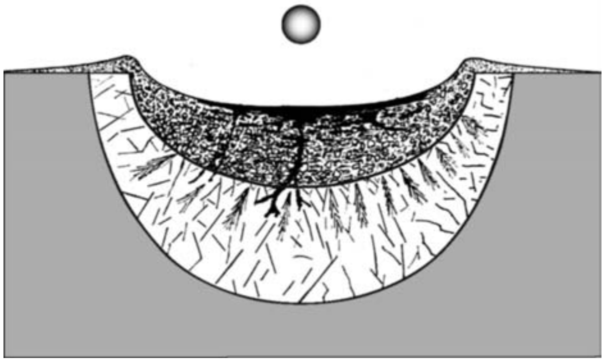


Fig. 15.27 Vertical section through meteorite explosion crater. Melt (black) formed by fusion of impacted rock mixes in all proportions with the underlying breccia. Crater depth is typically one-tenth of the crater diameter, and the depth to which rock is totally crushed is one-third the diameter. Intense fracturing extends twice as deep as the crushed zone. Shatter cones fan out away from the explosion. The size of the meteorite shown assumes it to be a stony variety with a density of 3500 kg m^{-3} traveling with a velocity of 20 km s^{-1}.

The steep walls of very large craters are gravitationally unstable and collapse along normal faults into the crater. Where the toes of these slump structures converge toward the center of the crater, they develop a central uplift. In extremely large craters, central uplift may continue long after the initial slumping, as a result of the lithosphere

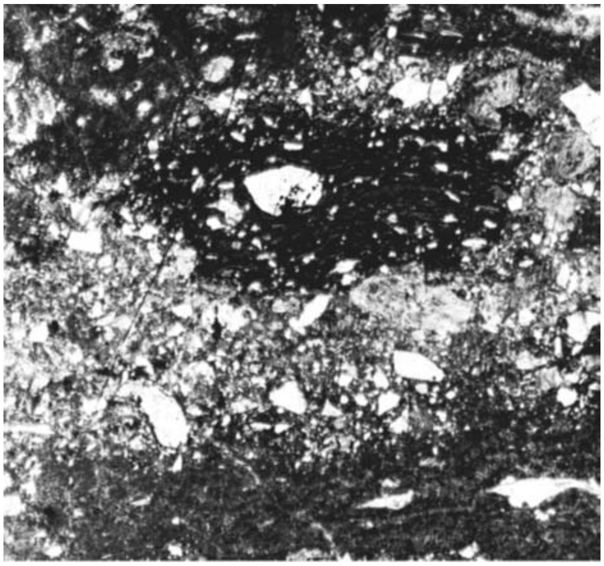


Fig. 15.28 Impact melt from the Manicouagan impact structure, Quebec (see point M in Fig. 15.22).

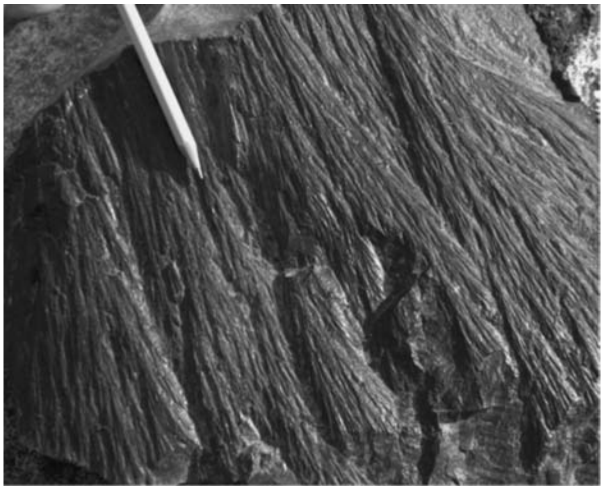


Fig. 15.29 Shatter cones in the Ordovician Trenton limestone from the Charlevoix impact structure, Quebec (see point C in Fig. 15.22). Shatter cones fan out in the direction away from the blast.

isostatically trying to compensate for the lack of mass in the crater and zone of broken rock. The amount of structural uplift in craters is related to the diameter of the crater according to the relation (Grieve and Therriault, 2000)

$$\text{Uplift} = 0.086\ D^{1.03} \tag{15.6}$$

and the diameter of the uplift (D_{U}) is related to the diameter of the crater by

$$D_{\text{U}} = 0.13\ D \tag{15.7}$$

The melt formed by impact rapidly quenches to a glass, which in some ancient impact structures has been mistakenly identified as a volcanic rock. Impact glasses, however, differ in significant ways from volcanic ones (Dence, 1971). First, they are extremely inhomogeneous and contain abundant fragments of the surrounding rocks, many of which exhibit shock features (see below). With increasing number of fragments, these glasses grade into breccia (Fig. 15.28). Because impact glasses are superheated they never contain phenocrysts, whereas most volcanic glasses do. Impact glasses also contain streaks of glass that correspond in composition to almost pure minerals, such as *lechatelierite*, formed from fused quartz or high-pressure silica polymorphs. These glasses testify to the high temperatures associated with hypervelocity impacts. Some glasses contain spherules of iron, which may be of meteoritic origin. Melt ejected from craters cools rapidly by radiation as it passes through the atmosphere, raining down as small glassy spheres. With very large explosions, melt may be ejected into space, and when it returns through the atmosphere, frictional heating and melting of the surface sculptures the glass particles into aerodynamic forms. These are known as *tektites*.

The high pressures generated by the explosions, which momentarily may exceed 10^3 GPa, produce new minerals and a variety of deformation structures. The high pressures do not last long enough to develop stable high-pressure mineral assemblages, and those transformations that do occur are limited to ones requiring no diffusion; that is, simple polymorphic transformations. Nor is there time for large crystals to be formed, and most identification of high-pressure forms is made by X-ray diffraction.

Both of the high-pressure polymorphs of quartz, *coesite* and *stishovite*, have been found at impact structures. Coesite is stable only above 2 GPa at low temperature, and though it may form in high-pressure metamorphic rocks, it would normally invert to quartz on returning slowly to the surface (Fig. 16.9). It has been found, however, in some alpine high-pressure metamorphic rocks, and so its presence cannot be taken as positive proof of meteorite impact. The polymorph stishovite, however, requires pressures in excess of 7 GPa (corresponding to a depth of 220 km), and it has been found only at meteorite craters. Stishovite may melt to lechatelierite after a shock wave passes. Some meteorite fragments, such as those of Canon Diablo, which formed Meteor Crater, Arizona, contain diamond, which requires about the same pressure as stishovite to form.

The high strain rates associated with the passage of a shock wave can produce disorder and dislocations in minerals. Plagioclase, for example, is converted at pressures above 5 GPa to a glassy pseudomorph known as *maskelynite* (Fig. 15.30). When examined in plane light under the microscope, maskelynite resembles normal plagioclase, retaining its crystal shape, cleavage, and twin lamellae. But under crossed polars it is isotropic. Also it does not diffract X-rays. Maskelynite therefore resembles a glass, but its refractive index, while being less than those of plagioclase of the same composition, is greater than that of glass formed by fusing the plagioclase. Also, heating of maskelynite at 900 °C for 2 h converts it back to a single crystal of plagioclase (Bunch *et al.*, 1967). Clearly, maskelynite lacks long-range order (hence no X-ray pattern), but unlike fused glass it retains the short-range order that allows it to revert easily

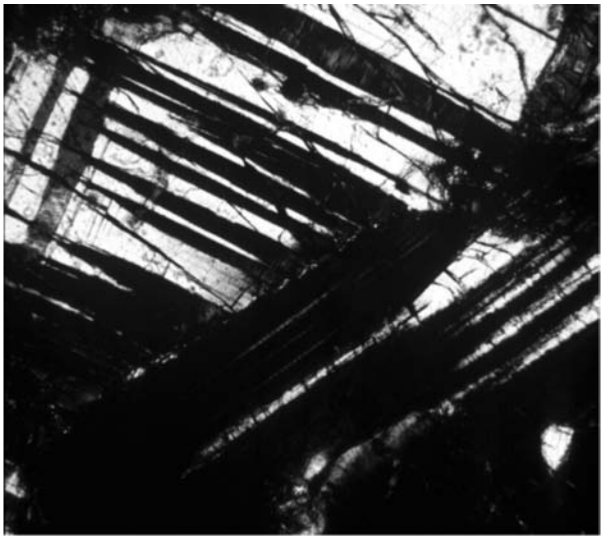


Fig. 15.30 Maskelynite (isotropic plagioclase glass) developed in anorthosite from the central uplift of the Manicouagan impact structure, Quebec. Note that only one of the two twinned directions in the lower grain has been converted to maskelynite, the remainder of the grains are still plagioclase.

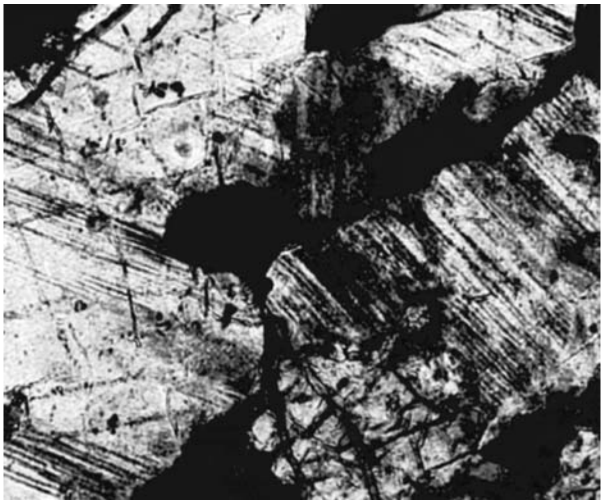


Fig. 15.31 Decorated planar features in quartz from the Charlevoix impact structure, Quebec.

back to plagioclase. Commonly, the transformation to maskelynite is selective, affecting only parts of grains or perhaps only one set of twin lamellae. This shows that crystallographic orientation with respect to the path of the shock wave is critical to the formation of maskelynite.

Table 15.1 *Planar features in quartz formed by meteorite impact*

Type[a]	Crystallographic orientation	Number of lamellae/mm
A	$\{0001\}$	100–250
B	$\{10\bar{1}3\}; \{01\bar{1}3\}; \pm\{0001\}$	200–500
C	$\{22\bar{4}1\}; \{10\bar{1}3\}; \{01\bar{1}3\}; \{0001\}$ rare	400–1000
D	$\{10\bar{1}2\}; \{01\bar{1}2\}; \{22\bar{4}1\}; \{10\bar{1}3\}; \{01\bar{1}3\}$	>1000
E	Same as in D, but with isotropic patches in which there are no planar features	>1000

[a] Type A is least deformed and type E is most deformed.
After Robertson *et al*. (1968).

One of the most common microscopic features produced by meteorite impact is sets of planar features in quartz (Fig. 15.31). They occur as fine, closely spaced, crystallographically controlled multiple sets of lamellae. They may be visible because of decorations (minute inclusions), differences in refractive index or crystallographic orientation, or actual cleavages. In their study of meteorite craters in the Canadian Shield, Robertson *et al*. (1968) were able to classify these planar features on the basis of crystallographic orientation into five different types, which could be correlated with the intensity of deformation (Table 15.1). The least deformed quartz has planar features parallel $\{0001\}$. With greater deformation planar features develop at 23° to the basal plane ($\{10\bar{1}3\}$) and then at 80° ($\{22\bar{4}1\}$). By this stage, the basal features are no longer present. Next, lamellae develop at 32° to the basal plane ($\{10\bar{1}2\}$). Finally, at the highest degree of deformation, parts of the quartz grains become isotropic. The planar features become more closely spaced with increasing intensity of deformation, and the refractive indices of quartz decrease. These five classes of planar feature can be used to map zones of deformational intensity around old craters.

The diameters of impact structures range from hundreds of meters to possibly as much as 300 km. It is difficult to determine the diameter of large ancient craters, because erosion has removed the actual crater. All that commonly remains is a circular structure that surrounds the central uplift. The actual crater diameter would have been approximately three times this size (Eq. (15.7). Figure 15.32 shows the present topographic expressions of a number of craters. They range from the 1.1-km Meteor Crater, Arizona, to the 100-km Chicxulub structure on the coast of the Yucatan Peninsula, Mexico. Structures with diameters of less than 30 km form simple craters, their precise form depending on the depth of erosion or amount of younger sedimentary fill. Structures with diameters greater than 30 km are complicated by a central uplift. In the case of the Clearwater Lakes, on the east side of Hudson Bay, which were probably formed by the impact of a pair of meteorites, we see that the western structure, with a diameter of 32 km, has a central uplift which forms a ring of islands composed of melt rock and breccia. The eastern structure, with a diameter of 27 km, lacks a central uplift. The large 250-Ma-old Manicouagan structure has a central uplift consisting of anorthosite and other basement rocks from beneath the crater. Melt rock that would have originally covered the crater floor now occurs only around the central uplift. The 2.0-Ga-old Vredefort Ring structure consists of a ~50-km-diameter core of Archean granitoid rocks surrounded by metamorphosed Archean and Proterozoic sedimentary and volcanic rocks. Sediments of the Transvaal System are even overturned against the core.

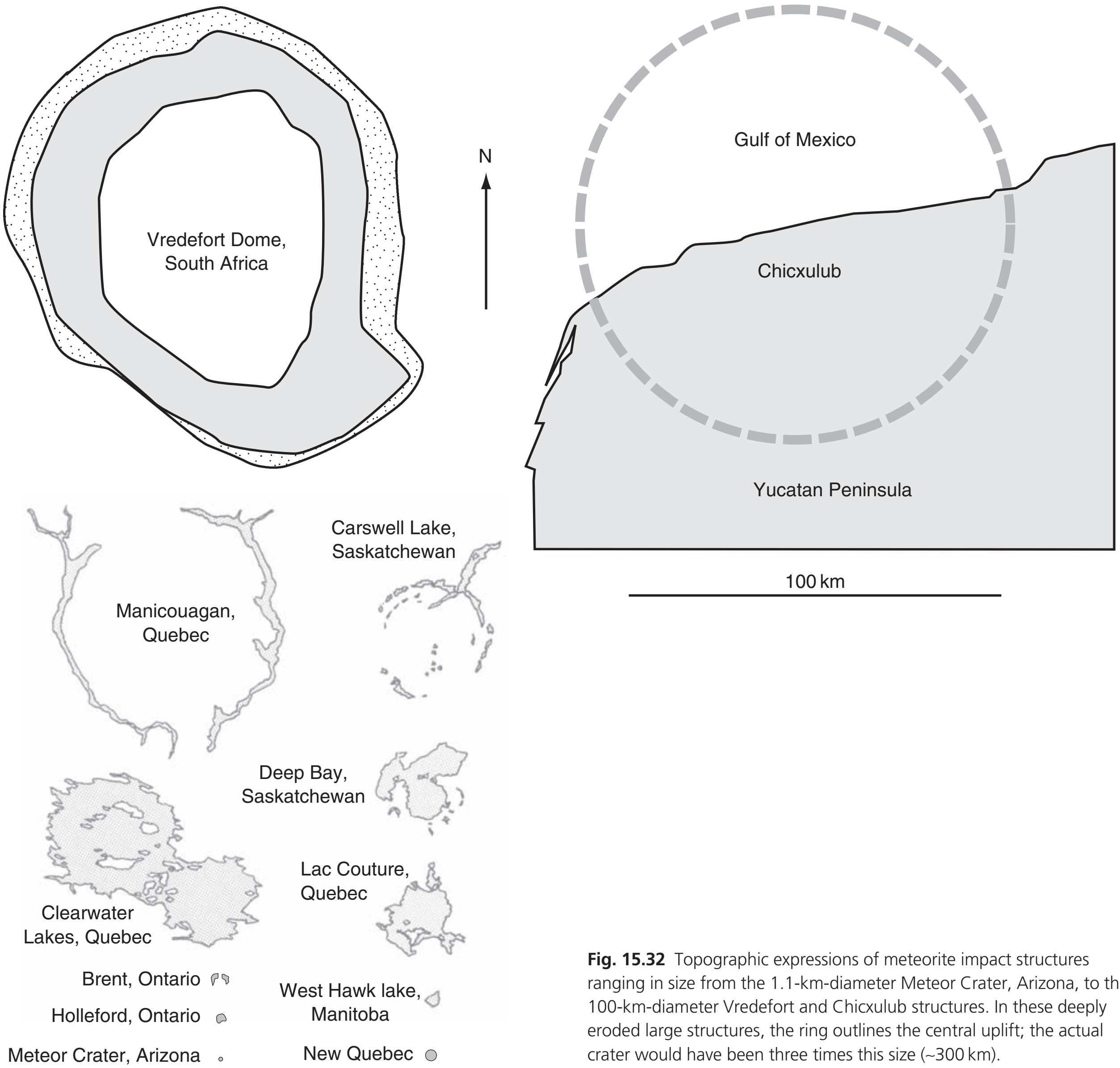


Fig. 15.32 Topographic expressions of meteorite impact structures ranging in size from the 1.1-km-diameter Meteor Crater, Arizona, to the 100-km-diameter Vredefort and Chicxulub structures. In these deeply eroded large structures, the ring outlines the central uplift; the actual crater would have been three times this size (~300 km).

These, in turn, are surrounded by a series of anticlinal and synclinal rings out to a diameter of 300 km (not shown in Fig. 15.32). Black aphanitic veins, which were originally described as *pseudotachylites* (tachylite = basaltic glass), are formed by *in situ* fusion of the host rocks (Philpotts, 1964). These veins can be found out to 80 km from the center of the structure. Coesite and stishovite have both been found in association with pseudotachylite (Martini, 1991). Shatter cones have also been found around the Vredefort structure, but these now have their apex pointing away from the structure because of the uplift. There is now no doubt that the Vredefort dome is the product of meteorite impact. Of a similar size is the Chicxulub structure, which is buried beneath sediments on the northern coast of the Yucatan Peninsula. This ring structure, which is revealed by gravity and seismic surveys and by drilling, is of great importance because it lies on the Cretaceous–Tertiary boundary and is referred to as the “smoking gun” responsible for the mass extinction at the K–T boundary.

The zone of rupture associated with impact structures having diameters of more than 50 km would extend into the lower crust, and the zone of fracture would penetrate to the upper mantle. The resulting decompression on the underlying mantle could cause large-scale melting if the geotherm were near solidus temperatures prior to impact. Large-scale melting would favor the formation of tholeiitic magma, which would then rise into the crater and pond beneath the less dense breccia and melt rock formed by the initial impact.

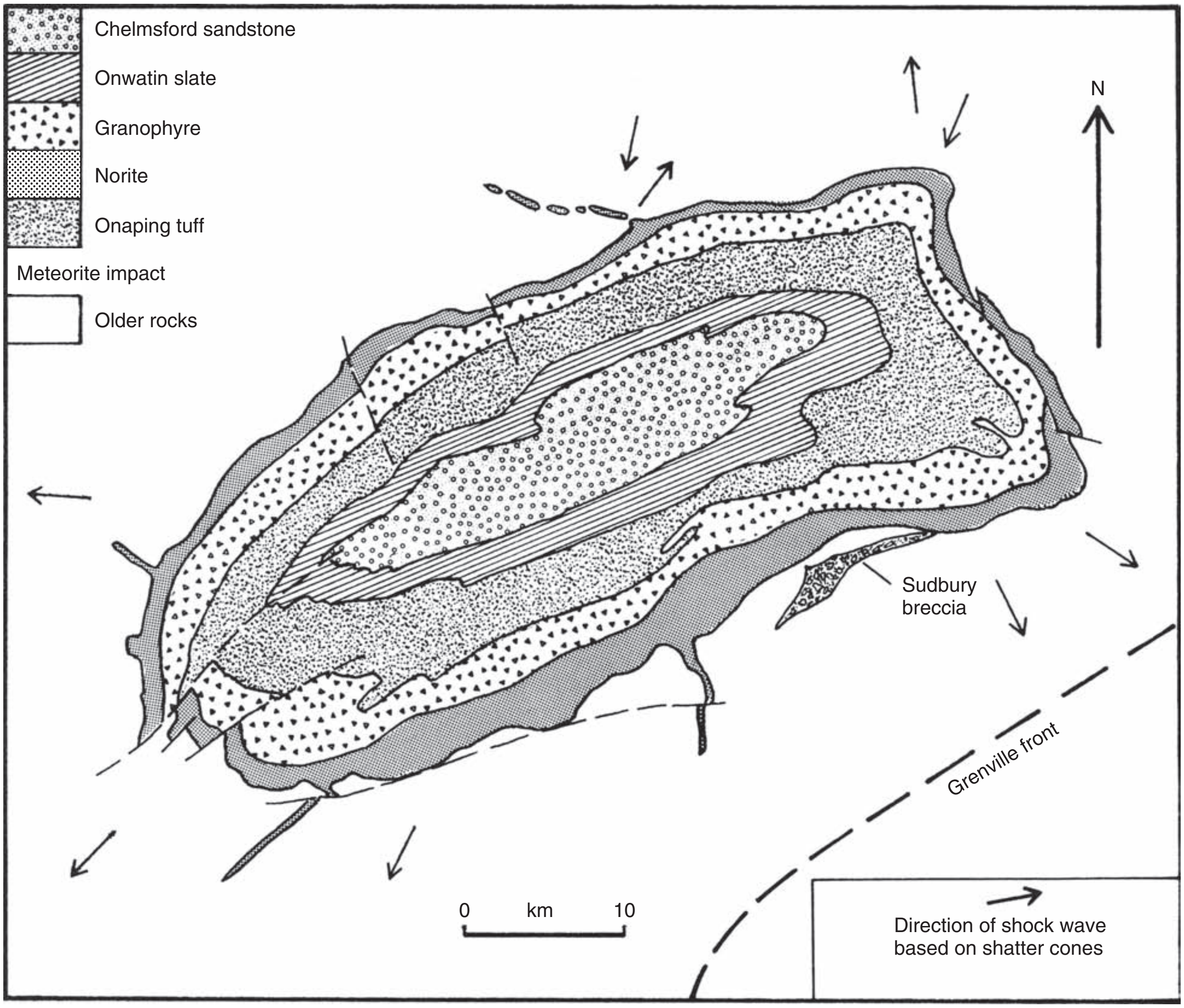


Fig. 15.33 Sudbury lopolith, Ontario, an ancient meteorite impact structure. (Map from various sources; shatter cone orientations from Guy-Bray *et al.*, 1966.)

Such a mechanism was proposed by Dietz (1964) for the origin of the 1.8-Ga-old Sudbury lopolith, Ontario (Fig. 15.33). Since its formation, this body has been deformed by the Grenville orogeny (1.1 to 1.0 Ga), and thus some interpretation is involved in reconstructing its original form (Naldrett, 2003). The lopolith is now elliptical in plan, but prior to Grenville deformation it may have been circular. The presence of shatter cones, most of which radiate and plunge outward from the body, is strong evidence of meteorite impact (Guy-Bray *et al.*, 1966). Rocks underlying the lopolith are cut by irregular veins and stockworks of a remarkable rock known as the *Sudbury breccia*. It is similar to the pseudotachylite of the Vredefort dome and contains fragments that range in size from large boulders to microscopic particles (Fig. 15.34). This rock was formed by shock melting of the impacted rocks. The magma that rose into the crater solidified to form norite. This rock contains a large percentage of interstitial granophyre, which increases in abundance toward the top, where it forms a continuous sheet (micropegmatite). The granophyre, in turn, intrudes the Onaping tuff (Fig. 15.35), which forms the roof of the intrusion. This rock was originally interpreted as a welded tuff of volcanic origin. However, it contains many fragments of the country rocks, and some quartz fragments contain planar deformation features (French, 1972). It has also been found to contain diamonds (Masaitis *et al.*, 1999). These diagnostic features of high strain rates and pressure leave no doubt that the Onaping tuff resulted from a meteorite impact explosion. The crater was later filled with sedimentary rocks, which now form the Onwatin slate and the Chelmsford Arkose. Most of the world's nickel comes from the ore deposits associated with the Sudbury lopolith. Although Dietz proposed that this nickel may have come from the original meteorite, the high Cu/Ni ratio of the Sudbury ores (about 1:1) makes this highly unlikely because this ratio in iron–nickel meteorites is about 1:500. The ores, which sank as immiscible sulfide liquids to the base of the body, are therefore probably of crustal origin. Re–Os analyses of the nickel ores show them to be highly enriched in radiogenic Os, which could originate only in a long-lived reservoir

Fig. 15.34 Sudbury breccia on the north side of the Sudbury lopolith. Fragments of gneiss are surrounded by a black aphanitic matrix (pseudotachylite) that was formed by shock melting at the time of meteorite impact. Doreen Philpotts for scale (1.67 m).

Fig. 15.35 Fragment of granite in flow-banded glassy fragment in the Onaping tuff, which forms the roof rock to the Sudbury lopolith (see Fig. 15.33). The presence of quartz with planar deformation features and even diamonds in this rock is evidence that the Onaping tuff was formed as an explosion breccia at the time of meteorite impact.

with a high Re/Os ratio. This points to the continental crust as the source of the ores (Shirey and Walker, 1998).

On a still larger scale, meteorite impact has been proposed as the triggering mechanism for the magmatism that gave rise to the Bushveld Complex, South Africa (Rhodes, 1975). This body and the nearby Vredefort Ring (Figs. 15.17 and 15.32) may have been formed by four simultaneous hypervelocity impacts. Although multiple impacts may seem highly improbable, the Clearwater Lakes (Fig. 15.32) show that they have occurred. The Bushveld Complex is very much like the Sudbury lopolith, with noritic rocks capped by a granophyre that is intrusive into overlying volcanic rock, the Rooiberg Felsite. Rhodes interprets the felsite to be an impact melt. Its age is indistinguishable from that of the Vredefort granophyre and the Bushveld Complex itself (1.95 to 2.0 Ga). No shatter cones have been found around the Bushveld, but Rhodes claims that the lobes of the Complex are formed by magma flowing into ring synclines around each of the three impact structures. A central uplift exposes highly deformed and brecciated rocks from beneath the complex. The evidence for an impact origin for the Bushveld, while tantalizing, is far from positive at this stage and will require further documentation. It seems quite likely, however, that early in Earth's history impact structures of this dimension would have formed.

15.10 PROBLEMS

15.1 Two basalts with different concentrations of the trace elements Ni and P might be related by fractional crystallization or different degrees of partial melting. The distribution coefficient (solid/liquid) for the compatible Ni is 10, and that for the incompatible P is zero. Calculate the factors by which the Ni and P in the less evolved composition (lower P) would have to be multiplied to give those in the more evolved composition if the two basalts are related by:

(a) one being a primary magma and the other its differentiation product following 5% crystallization, and

(b) one being a 1% partial melt and the other a 6% partial melt. Assume equilibrium fractionation and melting (review Eq. (14.36)).

(c) In light of your answers to parts (a) and (b), how might you distinguish chemical changes in rock series produced by fractional crystallization from those produced by partial melting?

(d) What effect would Rayleigh rather than equilibrium fractionation have?

15.2 The concentrations of Ni and Rb in typical primitive tholeiitic ocean island basalt are 120 and 5 ppm, respectively, and in primitive alkali basalt they are 100 and 30 ppm, respectively. Assuming that these lavas are formed by partial fusion of an identical source and that the distribution coefficient (solid/liquid) for Ni is 10 and for Rb is 0, what degrees of equilibrium partial fusion would produce the given compositions, and what are the concentrations of these elements in the source rock? (Hint: If the source rock were totally melted, the concentration of an element in the melt (c_i^{ol}) would be the same as in the source rock.)

15.3 Carefully study the zoning in the plagioclase phenocryst from Mount St. Helens shown in Figure 15.8, and list, in chronological order, all growth and resorption episodes visible. Give a possible explanation for the uneven distribution of inclusions on the one surface marked by inclusions (review Section 12.4 and Fig. 12.15). Finally, what geological events may be recorded in the zoning of this crystal?

15.4 In tabular form, contrast the compositions and mineralogy of typical MORB, ocean island alkali basalt, continental flood basalt, and calcalkali basalt.

15.5 Using Figure 10.37(A), show how three magmas of only slightly different composition could produce the following sequences of crystallization:

I olivine–orthopyroxene–plagioclase–clinopyroxene
II olivine–clinopyroxene–plagioclase–orthopyroxene
III olivine–orthopyroxene–clinopyroxene–plagioclase

15.6 In many tholeiitic layered intrusions, early-crystallizing olivine, which is forsteritic, is replaced in intermediate rocks by calcium-poor pyroxene (opx or pigeonite) and then reappears in late-stage rocks as a fayalitic variety. Assuming complete fractional crystallization, select an initial magma composition in the phase diagram for FeO–MgO–SiO_2 (Fig.15.36) that would produce such a sequence.

15.7 The most magnesian olivine in the Muskox Intrusion is $Fo_{89.5}$. Using a K_D for the partitioning of magnesium and iron between liquid and solid of 0.31 (Eq. (14.1)), calculate the magnesium number of the magma with which this olivine must have been in equilibrium.

15.8 Alnöite commonly contains perovskite rather than sphene, which occurs in the more silicic lamprophyres. Perovskite reacts with magma to form sphene if the silica activity is high enough according to the reaction

$$\mathrm{CaTiO_3(perovskite) + SiO_2(magma)} \\ \mathrm{= CaTiSiO_5(sphene)}$$

The presence of perovskite therefore sets an upper limit on the activity of silica in alnöites. Calculate the maximum silica activity that an alnöitic magma can have at 1100 °C. The free energy of formation in kJ mol^{-1} of silica liquid at this temperature is − 663.996, of perovskite is − 1277.885, and of sphene is −1958.551.

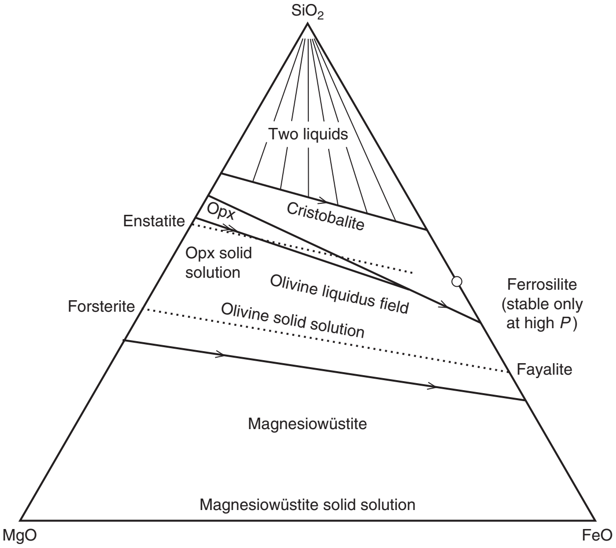


Fig. 15.36 for Problem 15.6.

15.9 Many of the volcanic rocks of the Toro and Ankole districts of the East African rift contain both kalsilite and leucite, which therefore defines the silica activity in these lavas through the reaction

$$\mathrm{KAlSiO_4(kalsilite) + SiO_2(magma)} \\ \mathrm{= KAlSi_2O_6(leucite)}$$

According to Robie *et al*. (1978), the free energies of formation of these materials from the elements at 1100 °C in J mol^{-1} are −1 560 611 for kalsilite, −663 906 for silica in magma, and −2 262 666 for leucite. Calculate the activity of silica in a lava containing these minerals at 1100 °C. Assume leucite and kalsilite to be pure phases.

15.10 In view of the possibility that ugandite and mafurite might be related to alnöitic magma through the breakdown of phlogopite at low pressure, is the silica activity calculated in Problem 15.9 compatible with the presence of perovskite in alnöitic magma (Problem 15.8)?

15.11 **(a)** Calculate the silica activity in a carbonatitic magma containing crystals of pure diopside and monticellite at 1100 °C. Use thermodynamic data from Table 7.1. (You may wish to use your solution to Problem 8.4 to help in setting up the calculation.)

(b) If a carbonatitic magma formed as an immiscible liquid from an alnöitic magma, how would the silica activities in these two liquids compare?

(c) In light of the silica activity calculated in part (a) and the possible silica activities in alnöites (Problem 15.8), could carbonatite form as an immiscible liquid from an alnöitic magma?

15.12 The viscosity of a komatiitic lava near its liquidus temperature is 0.5 Pa s. If the density of the magma is 2800 kg m^{-3}, how thick would an extensive sheet flowing down a 2° slope have to be to become turbulent? (*Hint*: Review Problem 3.15 and the critical Reynolds number in Section 3.6.)

15.13 Could megacrysts of bronzite, such as the one in Figure 15.24(B), be transported from depth into the crust by magma? Assume that the megacryst has a diameter of 20 cm and a density of 3400 kg m^{-3}. We do not know the density of the magma, but it cannot be far from 2500 kg m^{-3}. The viscosity of the magma is unknown and will depend strongly on composition. Let us assume two different values, one typical of a basalt, 50 Pa s **(a)**, and the other of an intermediate composition, 1000 Pa s **(b)**. Calculate the sinking velocity of the crystal using Stokes' law (Eq. (14.4)), and check that the conditions are appropriate for the equation (Eq. (14.5)). In light of your calculations, what conclusion can you draw from the presence of bronzite megacrysts in anorthosites?

15.14 The Manicouagan impact structure has a diameter of 67 km. Assuming that it was formed by a common stony meteorite with a density of 3500 kg m^{-3} and was traveling with a velocity of 20 km s^{-1}, **(a)** what was its kinetic energy, and **(b)** what was its diameter?

16 Metamorphism and metamorphic facies

16.1 INTRODUCTION

Metamorphism is the sum of all changes that take place in a rock as a result of changes in the rock's environment; that is, changes in temperature, pressure (directed as well as lithostatic), and composition of fluids. The changes in the rock may be textural, mineralogical, chemical, or isotopic. These changes proceed at varying rates, so time is an important factor in metamorphism. Any kind of rock can be metamorphosed; the starting rock is called the *protolith*. Common protoliths include the spectrum of igneous rocks from ultramafic to felsic, as well as sedimentary rocks such as sandstones, alumina-rich shale (pelite), and carbonate rocks (limestone and dolostone).

In its broadest sense, metamorphism includes the entire range of changes that take place between the zone of weathering and the zone in which melting gives rise to magmas. Traditionally, however, the low-temperature changes associated with weathering and the lithification and diagenesis of sediments have been omitted from the study of metamorphism. Typical metamorphic reactions take place at temperatures above 150 to 200 °C. At the highest temperatures, metamorphism gives way to magmatic processes where partial melting produces *migmatites*, mixed igneous–metamorphic rocks. Average continental crust starts melting under water-saturated conditions around 1000 °C at low pressures, but this temperature decreases with increasing pressure, dropping to 650 °C at 0.5 GPa (Fig. 11.7). The water-saturated beginning of melting of granite marks the upper temperature limit of metamorphism in many regions, because the latent heat of fusion provides an enormous heat sink. If the activity of water is less than 1, the melting curve is raised (Fig. 11.6), and thus higher metamorphic temperatures may be attained.

Metamorphic rocks are abundant because metamorphism can occur throughout the crust and affect a broad assortment of protoliths. For example, metamorphic rocks make up about 60 volume percent of the continents. Furthermore, the widespread alteration of oceanic crust due to hydrothermal circulation of seawater at mid-ocean spreading ridges is also generally included within the domain of metamorphism (Figs. 1.2, 21.1). Taken together, the metamorphic rocks of the continents and the hydrothermally altered rocks of the ocean basins constitute an enormous volume of crust. Consequently, knowledge of metamorphism is necessary to understand Earth's large-scale geological, geochemical and geophysical evolution.

Metamorphism that occurs during heating or burial is called *prograde* metamorphism. Heat can be introduced into the crust by bodies of magma, which, on cooling, liberate their heat content, and, on crystallizing, liberate the latent heat of crystallization. Metamorphism resulting from this heat that is localized around igneous intrusions is referred to as *contact metamorphism* (Figs. 16.1(A), 4.45). Heat can also be introduced over much larger areas. For example, when sedimentary rocks are buried, either through subsidence and filling of sedimentary basins or by overthrusting of other crustal rocks, their temperature rises. This is due in part to the geothermal gradient and in part to the blanketing effect of crustal rocks, which normally have relatively high concentrations of heat-producing radioactive elements (review Problem 1.5). The metamorphism resulting from this type of burial and heating is associated with convergent plate boundaries and develops on a regional scale; it is therefore referred to as *regional metamorphism* (Fig. 16.1(B)). Furthermore, extensive magmatism associated with rift zones, arcs, areas of lower crustal delamination, or mantle plumes can elevate geothermal gradients over large areas of the crust and produce regionally metamorphosed rocks. Regional metamorphism is normally accompanied by tectonism, with the result that the metamorphic rocks develop prominent foliated fabrics – slate, schist, gneiss. In contrast, most contact metamorphic rocks are not strongly deformed during metamorphism, and they therefore lack the foliated fabric of regionally metamorphosed rocks. Metamorphic changes that develop during cooling are referred to as *retrograde*; they may affect both contact and regionally metamorphosed rocks and frequently accompany exhumation.

When studying metamorphic rocks, petrologists attempt to determine the conditions and the timing of the metamorphism and then, from these, to deduce the major geologic events that have affected a region. The mineralogy, textures, and structures of metamorphic rocks record important information about the conditions prevalent during the formation of these rocks, and one of the main tasks of metamorphic petrology is to learn how to read this record. We have already seen, for example in Section 1.6, that many metamorphosed pelites (metapelites) contain one or more of the Al_2SiO_5 polymorphs, andalusite, kyanite, or sillimanite, yet the calculated steady-state continental geotherm (Fig. 1.9) lies entirely within the kyanite stability field. Metamorphic rocks containing andalusite or sillimanite must therefore have formed at temperatures well above the steady-state geotherm. In

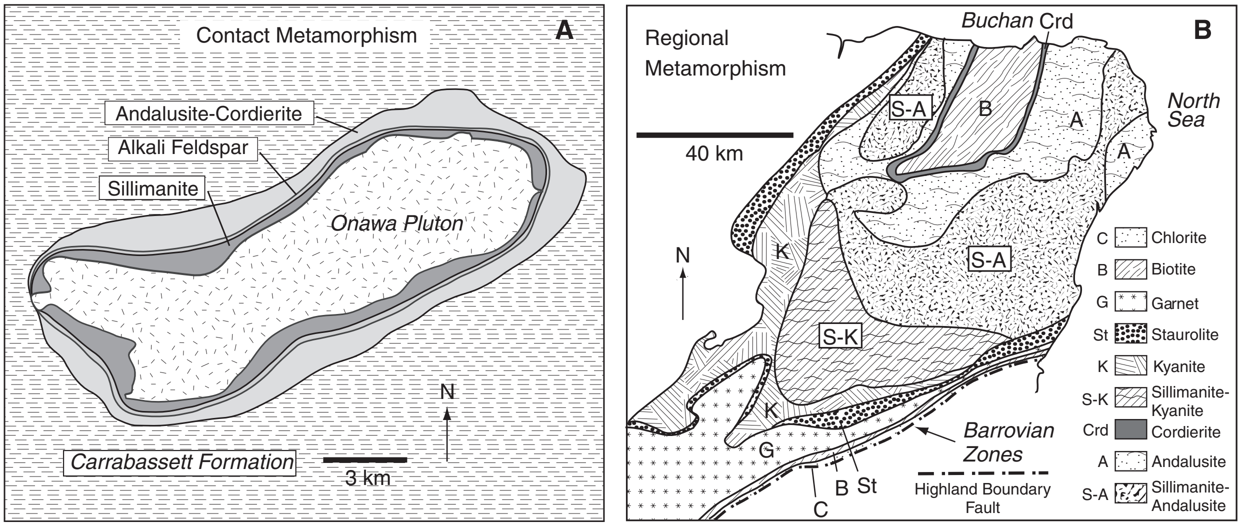


Fig. 16.1 Contact and regional metamorphism. (**A**) Contact aureole around the Onawa pluton, Maine, illustrating andalusite + cordierite, alkali feldspar, and sillimanite index mineral zones developed in metapelitic rocks of the Carrabassett Formation (after Philbrick, 1936; Symmes and Ferry, 1995). Partial melting of the metasediments occurred directly adjacent to the pluton. (**B**) Map of regional metapelitic index mineral zones and isograds in the northeastern Scottish Highlands including the area immediately north of the Highland Boundary Fault mapped by Barrow (1893, 1912) and the Buchan area (modified from Atherton, 1977). Pre-, syn-, and post-tectonic igneous rocks omitted for clarity.

addition, rocks containing all three polymorphs or just kyanite and sillimanite must have formed at pressures of about 0.4 GPa or higher; that is, at depths greater than 14 km. The occurrence of all three polymorphs in a single rock is rare, but kyanite–sillimanite assemblages are not uncommon and are exposed on the surface of the Earth in regions where the present crust is 35 km thick. At the time of metamorphism, then, the crust must have been at least 50 km thick. We can conclude from these simple observations that such metamorphic rocks must form in a thickened continental crust with a steepened geothermal gradient. These requirements are met, for example, near convergent plate boundaries.

In this and subsequent chapters we will see how petrologists go about reading the metamorphic record. The treatment will examine important types of field and laboratory data, how they are analyzed, and what metamorphic conditions they record. The principles introduced are developed further in subsequent chapters.

16.2 METAMORPHIC VOLATILES

During metamorphism, existing minerals react to form new minerals. Some reactions do not produce or consume fluids. Simple examples of these *solid–solid* reactions are the transformations among the aluminosilicate polymorphs, such as andalusite–kyanite or sillimanite–kyanite (Fig. 16.2(A)). Many other reactions, however, involve fluids (Fig. 16.2(B), (C), (D)). Fluids are important in metamorphism, because they exert strong controls on the temperatures and pressures of reaction, the extent of reaction progress, and the transport of heat, chemical elements, and isotopes through mountain belts.

Before examining metamorphic fluids, let us first consider the processes involved in the lithification or diagenesis of sediment. Freshly deposited marine sediment can contain as much as 50% pore fluids. With burial and compaction, this is soon reduced to about 30%, at which stage the sediment would consist, in the simplest case, of close-packed spheres. The pore fluid in the compacted sediment remains a continuous phase connected through the intergranular channelways. The fluid is therefore under hydrostatic pressure. The sedimentary grains, however, in addition to experiencing the pressure from the surrounding fluid must support the weight of the overlying sedimentary grains. This creates stresses on the grains at their points of contact. As indicated in Section 12.6, stress increases the solubility of minerals, and consequently, the grains dissolve at the points of contact (Fig. 16.3). The pore solution, however, becomes saturated in the dissolving mineral, and further dissolution leads to redeposition on the parts of grains protruding into pore spaces. In this way the porosity of a sedimentary rock is reduced almost to zero at the base of a thick sedimentary sequence. For this reason, fluids involved in metamorphic reactions usually come from minerals rather than from trapped pore fluids.

Fluid constituents such as water, CO_2, and CH_4 are referred to in a general way as *volatiles*, and liberation of these constituents from rock is known as *devolatilization* (Fig. 16.2(B)). The bulk of the volatiles released during metamorphism are initially present as structural water (OH^-) in hydrous minerals like sheet silicates and amphiboles, and as $(CO_3)^{2-}$ radicals in carbonate minerals like calcite. For example, basalts and ultramafic rocks that undergo hydrothermal metamorphism at spreading ridges can contain more than 10 wt% volatiles held in solid phases including serpentine, chlorite, and carbonate minerals. Pelitic sedimentary rocks commonly have as much as 5.0 wt% water hosted by clay minerals, and lesser amounts of CO_2 contained in carbonate minerals. Limestones and dolostones are rich in CO , but can

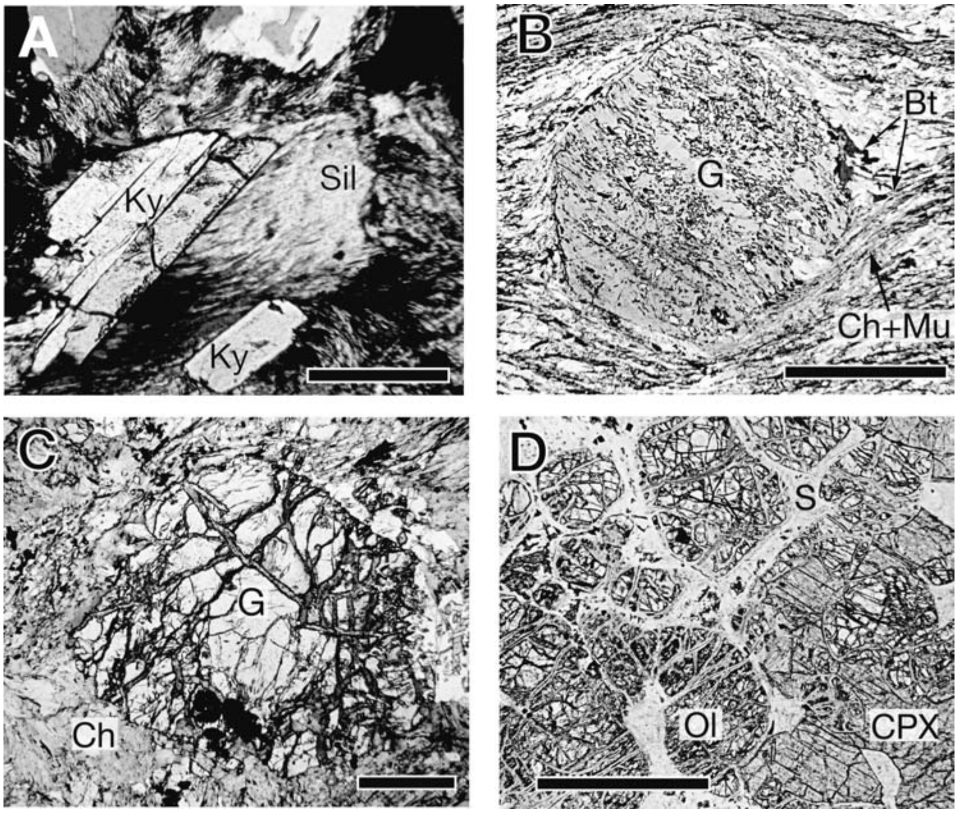


Fig. 16.2 Photomicrographs illustrating metamorphic reactions. (**A**) Solid–solid reaction. Kyanite (Ky) replacing sillimanite (Sil), Barrovian zones, Scotland. The sillimanite is the fine-grained variety known as *fibrolite* that grows in fibrous mats. Crossed-polarized light; scale bar = 0.25 mm. (**B**) Prograde reaction and devolatilization. Chlorite and muscovite (Ch + Mu) reacted to produce garnet (G) and biotite (Bt) via the generalized dehydration reaction: muscovite + chlorite + 3 quartz = biotite + 4 garnet + 12 H_2O. Wepawaug Schist, south-central Connecticut. Plane-polarized light; scale bar = 2 mm. (**C**) Retrograde hydration. Chlorite (Ch) replacing garnet (G) along cracks and grain boundaries in blueschist, Syros Island, Greece. Plane-polarized light; scale bar = 1 mm. (**D**) Hydrothermal alteration of ultramafic rock. Olivine (Ol) and clinopyroxene (CPX) alter to serpentine (S) along cracks and grain boundaries. Plane-polarized light; scale bar = 1 mm.

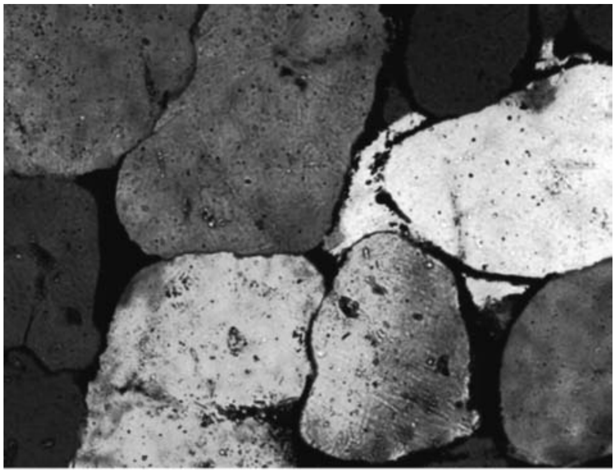


Fig. 16.3 Grains of sand in this rock dissolved at points of contact as a result of pressure solution due to compaction alone. Original outlines of grains are marked by inclusions of clay particles that adhered to the grains. The dissolved silica reprecipitated as cement, which grows in optical continuity on the original grains (note that birefringence of white grain is the same as that of surrounding cement).

also contain significant amounts of clay minerals. In addition to the volatile constituents held in solid phases, most sediments contain a free fluid phase that occupies the pore space between mineral grains. Although most of this fluid is expelled from sedimentary basins prior to metamorphism during compaction, diagenesis, and lithification, it can be important in certain contact metamorphic settings where igneous rocks intrude poorly lithified sediments.

Water is the dominant constituent in a wide array of metamorphic fluids. Devolatilization reactions that liberate water are called *dehydration* reactions. Dehydration reactions are endothermic; the heat supplied to rocks during prograde heating provides the energy needed for these reactions to proceed. A simple example is the destruction of muscovite and quartz to produce K-feldspar and aluminosilicate (discussed previously in Problem 8.8):

$$\begin{aligned} &\mathrm{KAl_2AlSi_3O_{10}(OH)_2} + \mathrm{SiO_2} \\ &\quad \text{(muscovite)} \qquad \text{(quartz)} \\ &\quad = \mathrm{KAlSi_3O_8} + \mathrm{Al_2SiO_5} + \mathrm{H_2O} \\ &\qquad \text{(K-feldspar)} \ \text{(sillimanite)} \ \text{(fluid)} \end{aligned} \tag{16.1}$$

The water is liberated from the rock as a supercritical fluid phase, which at the pressures extant during regional metamorphism has a density near that of normal water at standard temperature and pressure ($\sim$1000 kg m^{-3}). Consequently, during regional metamorphism a pelitic rock may lose a volume of water that is of the same order of magnitude as the volume of the solid rock itself (Problem 16.1). The liberated water enters a fluid phase that occupies fractures or the pore space between mineral grains. The fluid is less dense and less viscous than the rock. Thus, it migrates toward Earth's surface under pressure gradients generated by geologic factors including buoyancy, metamorphic reactions, and deformation.

The loss of volatiles from rocks is an important general consequence of prograde metamorphism. On the other hand, retrograde metamorphism often adds volatiles back into rocks that are exposed to fluids during cooling. For example, reaction (16.1) commonly proceeds in reverse so that K-feldspar and aluminosilicate combine with water to produce muscovite and quartz. Another common example is the partial or complete replacement of garnet by chlorite (Fig. 16.2(C)). Such retrograde hydration reactions are exothermic and actually liberate heat to the rock mass.

Not all metamorphic fluids are generated by devolatilization of metamorphic rocks. For example, magmas are a potential source of volatiles. Fluids can be released as magmas ascend and decompress, or as they crystallize (see Sections 11.2 and 11.6). These fluids can then infiltrate into the surrounding rocks and be involved in prograde or retrograde reactions. Another example is seawater, which circulates near spreading centers and is critical for hydrothermal metamorphism of oceanic crust (Fig. 16.2(D)).

16.3 METAMORPHIC GRADE

Heat, pressure, chemically reactive fluids, and deformation can all act to drive metamorphism, but they are not distributed evenly throughout the crust. Consequently, the intensity of metamorphism varies spatially. The intensity or degree of metamorphism is referred to in a general way as *metamorphic grade*. One of the earliest attempts to evaluate metamorphic intensity on a regional scale was by Barrow (1893) working in the southeastern Scottish Highlands. He mapped zones of progressively metamorphosed metapelitic rock that he named after readily identifiable *index* minerals; these are chlorite (originally called "zone of digested clastic mica" by Barrow), biotite, garnet, staurolite, kyanite, and sillimanite, listed in order of increasing intensity of metamorphism (Figs. 16.1(B), 16.4). Using modern petrologic methods (Chapter 19), we know that temperatures increased from about 300–350 ºC in the chlorite zone, to 650 ºC or more in the sillimanite zone. This classical *Barrovian* metamorphism is not limited to Scotland, and is in fact the most widely recognized type in regionally metamorphosed areas throughout the world. Although Barrow's sequence is common, it is not unique. Indeed, in the Buchan area, just to the northeast of the Barrovian type locality (Fig. 16.1(B)), the sequence is chlorite, biotite, cordierite, andalusite, sillimanite. This *Buchan* metamorphism developed at lower pressures than the Barrovian.

Subsequent to Barrow's work, Tilley (1924) coined the term *isograd* (equal grade) to describe the boundary between any two index mineral zones. The basic idea is that mineral assemblages on either side of an isograd can be related through a simple reaction. The term implies that all points along an isograd were exposed to the same grade or intensity of metamorphism, based on the assumption that an isograd reaction takes place over a narrow range of temperature, pressure, and fluid composition. In this case, an isograd is inferred to represent the trace of the intersection of the three-dimensional reaction surface (which extends into the crust), with the present erosion surface.

The isograd concept is straightforward, but the precise relation between an isograd and the intensity of the factors causing metamorphism is complex and can be determined only through detailed analysis. For example, some index minerals, like garnet, can form over considerable ranges of pressure, temperature, and fluid composition. The garnet isograd developed at pressures of about 0.4 GPa in the northeastern part of Figure 16.1(B), but to the southwest pressures were around 0.6 GPa or more. Thus, the depth of metamorphism actually varies along the isograd, and the garnet zone is polybaric. The rock's bulk composition is another critical consideration. Sillimanite may form in an aluminum-rich metapelite at considerably lower temperatures than in a less aluminous one. Consequently, local or regional variations in protolith characteristics, such as those caused by changes in sedimentary depositional environments, can shift the positions of mapped "isograds." In spite of these and other complications, mapping of index mineral zones and isograds is still quite useful as a convenient field method for studying metamorphic rocks and for giving a general idea of the metamorphic grade.

Mapping in metamorphic terranes can be challenging because of the wide variety of lithologic variations that may be present. Primary boundaries are those between different rock bodies including sedimentary bedding planes separating mappable formations, and igneous contacts. In addition to primary boundaries, metamorphic boundaries separate different metamorphic zones that have undergone different intensities of metamorphism, as evidenced by their mineralogy. Deformation, including folding, faulting, and boudinage, can affect both primary and metamorphic zone boundaries during orogeny.

Most primary boundaries are easily recognized, for they separate rocks of distinctly different composition, for example, pelite/sandstone or shale/granite (Fig. 16.5(A)). Primary features, such as sedimentary bedding and igneous textures, may also help distinguish rock types, but with increasing metamorphic grade such features become obscured by the growth of new minerals or the redistribution of elements, and they may be completely obliterated at the highest grades of metamorphism. Sedimentary layering is one of the most difficult features to identify with certainty in highly

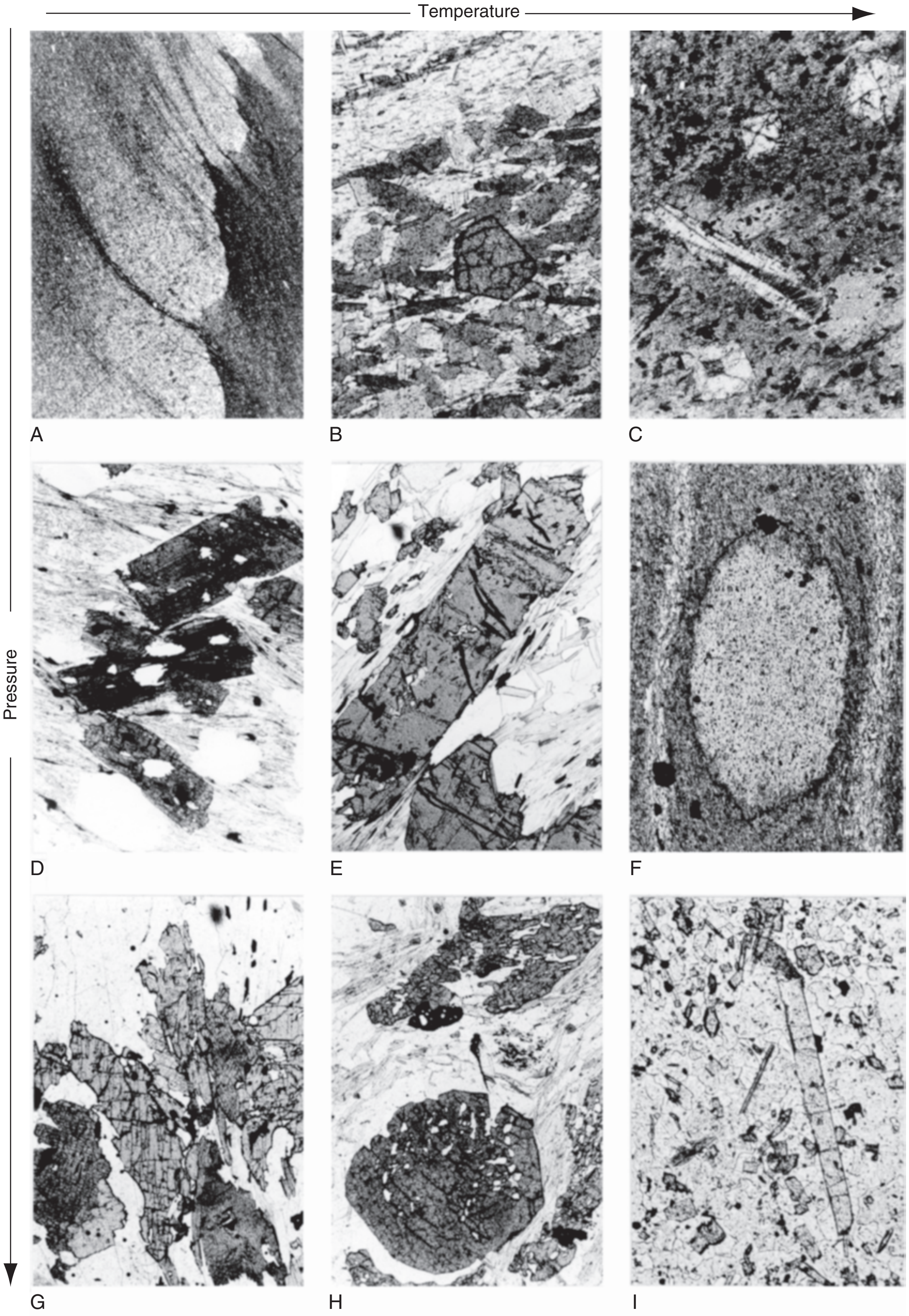


Fig. 16.4 Photomicrographs of typical minerals in metapelitic rocks. All samples contain quartz and muscovite except for (I), which contains quartz and K-feldspar. The width of each field is 4 mm. All are under plane polarized light. Minerals are arranged approximately with temperature of formation increasing from left to right and pressure increasing from top to bottom. (**A**) Graded silt bed (top of bed to left) in chlorite-bearing slate, Waterville Formation, south-central Maine. (**B**) Biotite and garnet in Waterville Formation. (**C**) Andalusite porphyroblasts in hornfels of the Skiddaw granite, English Lake District. Concentrations of inclusions along sector boundaries in andalusite produce the variety known as chiastolite. (**D**) Chloritoid porphyroblasts, Leeds, Quebec. (**E**) Staurolite in Gassetts Schist, Vermont. (**F**) Cordierite porphyroblast, Waterville Formation, south-central Maine. (**G**) Kyanite in Gassetts Schist, Vermont. (**H**) Garnet with core of quartz inclusions, Bronson Hill anticlinorium, south-central Massachusetts. (**I**) Sillimanite, Aberdeenshire, Scotland.

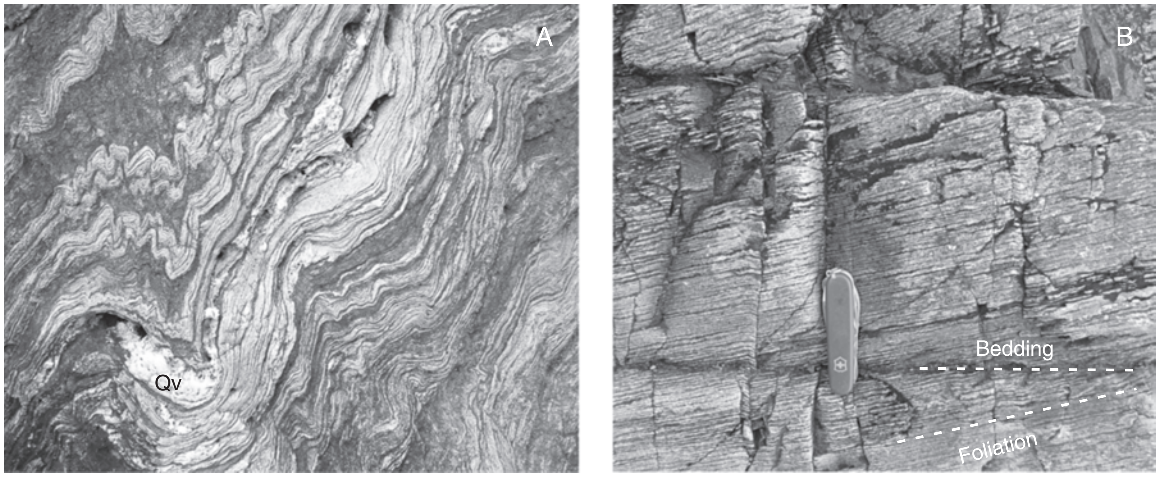


Fig. 16.5 Primary and metamorphic layering. (**A**) Folded primary sedimentary layering, Biotite zone, southwestern Scottish Highlands. Darker rocks are metapelitic, whereas lighter bands are metasandstones (also known as *metapsammites*). Note irregular spacing of layers characteristic of many sedimentary sequences. Qv denotes quartz vein. Field of view is approximately 1 meter wide. (**B**) Primary bedding in massive metasandstone is horizontal. Metamorphic layering (or *foliation*; see Chapter 17) is at an angle to bedding and consists of regularly spaced, alternating sheet silicate-rich and quartzofeldspathic layers. Biotite zone north of Stonehaven, northeastern Scotland.

metamorphosed rocks, because millimeter- to centimeter-scale layering, which resembles sedimentary bedding, can develop during metamorphism at pronounced angles to bedding. The thickness of these metamorphic layers tends to be rather constant, whereas that of sedimentary layers is commonly variable (Fig. 16.5(A)). In addition, most metamorphic layering is produced by abrupt changes in mineral abundance; for example, one set of layers may consist essentially of quartz and feldspar and the alternating set of sheet silicates like muscovite (Fig. 16.5(B)). Changes in composition within sedimentary beds, by contrast, tend to be more gradational. In some cases, however, it is necessary to look for compositional layering on the scale of an outcrop, or larger, before a decision can be made as to the primary or secondary origin of layering.

Boundaries between most metamorphic zones are not as evident in the field as the primary boundaries, and commonly they can be identified only after close inspection of the rocks with a hand lens or even in a thin section. Because many of the Scottish rocks are rather fine grained, Barrow himself often had to use a petrographic or binocular microscope for mineral identification and textural interpretation. As discussed above, metamorphic zones are identified on the basis of a characteristic mineral or mineral assemblage. In nature, zone boundaries are commonly gradational, and can extend over meters to even kilometer scales. In Scotland, for example, relict kyanite coexists with sillimanite well into the sillimanite zone due to slow reaction rates (Fig. 16.1(B)).

16.4 METAMORPHIC FACIES

Index mineral zones and isograds are not the only ways to assess metamorphic grade in the field. Another approach came largely as a result of two important studies, one by Goldschmidt (1911) and the other by Eskola (1920). Goldschmidt, in mapping the contact metamorphic rocks around alkaline plutons in the Oslo graben, in Norway, showed that their mineralogy was simple; the rocks rarely contain more than four or five minerals, and the mineral assemblages appear consistent with the Gibbs phase rule (Section 8.5, Eq. (8.11)). Eskola (1920) found that contact metamorphic rocks around granitic intrusions at Orijarvi, Finland, also have simple mineral assemblages that appear to conform to the Gibbs phase rule. However, these assemblages are different from those around the Oslo plutons. For example, metapelitic rocks at Oslo are characterized by K-feldspar and andalusite, whereas similar rocks at Orijarvi contain muscovite coexisting with quartz. Dehydration of muscovite proceeded at Oslo according to Eq. (16.1), but did not at Orijarvi. Eskola concluded that chemical equilibrium was probably attained in both areas, but that the conditions of metamorphism must have been different. This led him to propose the *metamorphic facies* concept, which he generalized to cover all conditions in contact and regional metamorphism. He defined a *metamorphic mineral facies* as comprising all the rocks that have originated under temperature and pressure conditions so similar that a definite bulk rock chemical composition results in the same set of minerals (Eskola, 1920). Stated another way, the metamorphic facies concept can be expressed in two simple sentences.

1. For rocks metamorphosed under the same physical conditions, different mineral assemblages represent different bulk rock compositions.
2. For a given bulk rock composition, different mineral assemblages indicate metamorphism under different physical conditions.

Eskola (1920) was the first to appreciate fully the broad relations between mineral assemblages, rock composition, and the pressures and temperatures of metamorphism. His

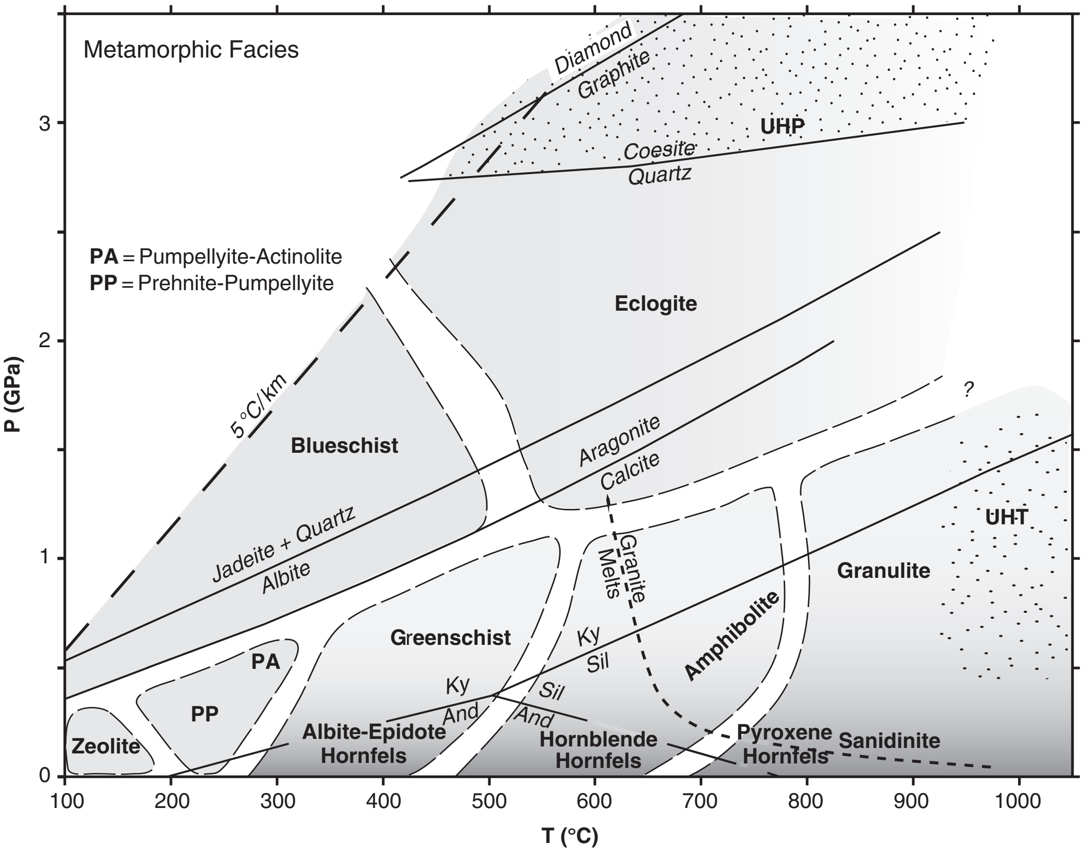


Fig. 16.6 Approximate temperatures and pressures for the major metamorphic facies. Rocks are assumed to be in equilibrium with water at the same pressure as the load pressure. Location of triple point between kyanite (Ky), sillimanite (Sil), and andalusite (And) from Holdaway (1971). Melting curve for peraluminous granite under water-saturated conditions from Clemens and Wall (1981). UHP and UHT designate approximate fields of ultrahigh pressure and ultrahigh temperature metamorphism, respectively. Limiting geothermal gradient of 5 °C per km shown; gradients less than this are extremely rare in nature.

metamorphic facies concept is still used in a general way today, but it contains a number of pitfalls for the unwary. Two of the problems will be mentioned here briefly as examples, but detailed discussion will be left to later chapters. First, the metamorphic facies concept is based on the critical assumption that a metamorphic rock is a mineral assemblage that closely approached equilibrium at the time of metamorphism. Unfortunately, this assumption cannot be proven. While adherence to the phase rule is a necessary requirement for equilibrium, it does not prove attainment of equilibrium. The converse, however, is true; that is, if a rock violates the phase rule it cannot have been at equilibrium. Either way, the phase rule is often difficult to apply in natural settings because rocks are chemically complex and the minimum number of chemical components needed to describe them is not always obvious. Nonetheless, the frequency with which many metamorphic facies occur is so great that they most likely do represent an approach to chemical equilibrium assemblages. The occurrence of zoned metamorphic minerals, however, is clear evidence that the approach to equilibrium is not always complete.

Second, it is often assumed that significant gradients in temperature, pressure, and fluid phase composition are not likely to have existed on the scale of an outcrop, and therefore all rocks in an outcrop would have experienced essentially the same temperature, pressure, and fluid phase composition. Although temperature and pressure are unlikely to vary much over short distances, the fluid phase composition might; in which case it cannot be treated as a homogeneous, environmentally controlled variable. Rocks with different mineral assemblages must, according to the metamorphic facies concept, have different bulk compositions. However, differences in fluid composition can drive chemical reactions that produce different mineral assemblages in rocks with the same bulk composition, even if temperature and pressure do not vary. In Chapter 21, we see that in some rocks the fluid phase composition can vary over short distances as a result of devolatilization reactions in adjacent layers of different mineralogy.

Eskola originally proposed five different metamorphic facies, but today about a dozen are recognized, covering the entire spectrum of possible metamorphic conditions (Fig. 16.6). He formulated his metamorphic facies concept largely from study of metamorphosed basalts, so most of the facies are named after metabasaltic rock types, or after common minerals found in metabasalts (Fig. 16.7). Experimental studies provide the data necessary to determine the approximate pressures and temperatures over which each of the facies is stable, but the boundaries between the facies are strongly dependent on the activity of water, which in Figure 16.6 has

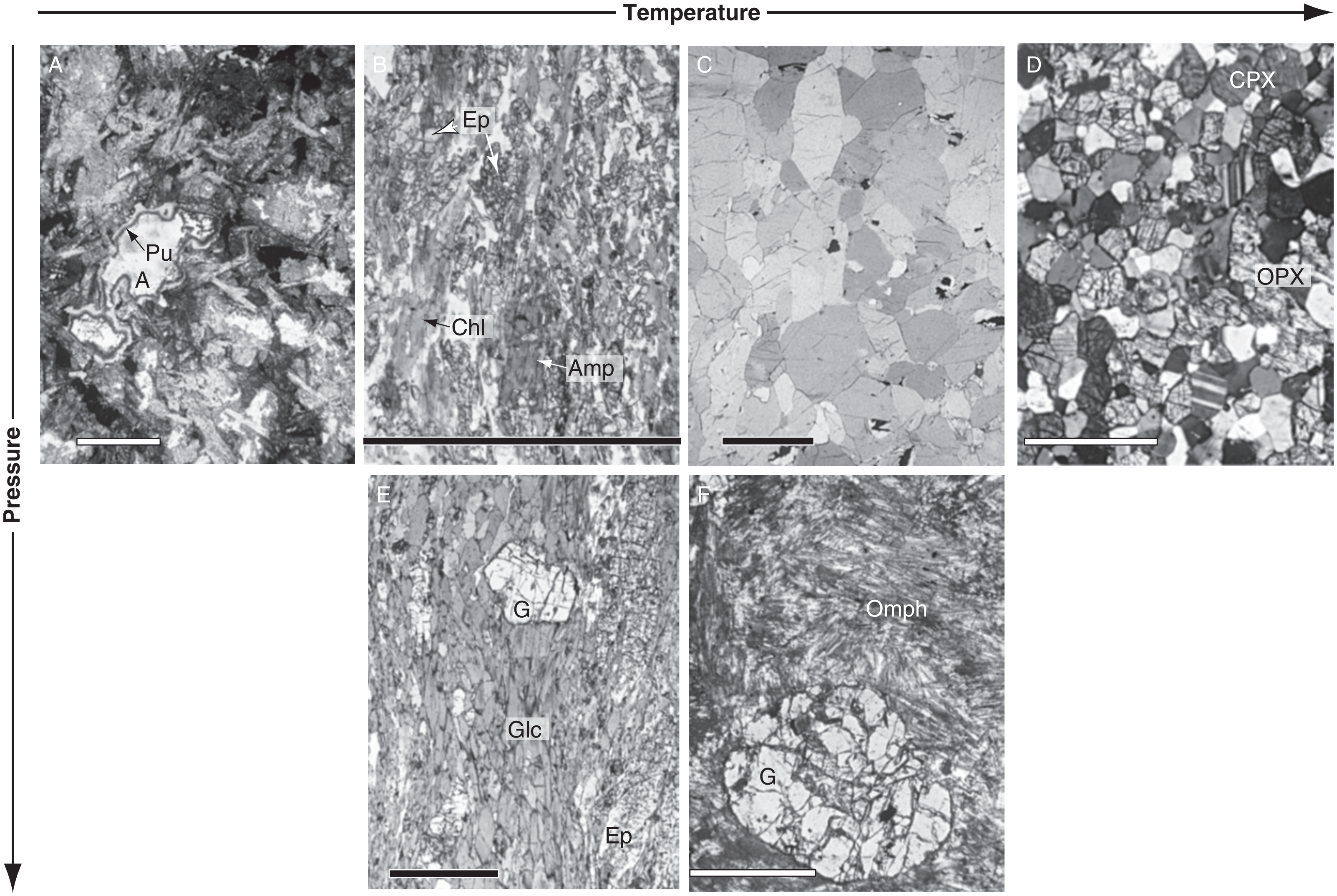


Fig. 16.7 Photomicrographs of metabasalts illustrating selected metamorphic facies. Scale bars are 1 mm. Crossed-polarized light for (D); plane-polarized light for others. (**A**) Zeolite facies, Vancouver Island, British Columbia. Original amygdaloidal basalt is undeformed and preserves relict igneous textures. Plagioclase laths appear cloudy and are replaced by albite, pumpellyite, and wairakite ($Ca(Al_2Si_4O_{12}) \cdot 2H_2O$). Amygdule (A) linings include chlorite fringed by high-relief pumpellyite (Pu). (**B**) Greenschist facies metabasalt from south-central Connecticut rich in the characteristic green minerals epidote (Ep), chlorite (Ch), and actinolitic amphibole (Amp). Low-relief (light) areas occupied by albite and quartz. (**C**) Amphibolite dominated by hornblende, Mitchell County, North Carolina. Plagioclase and quartz present between hornblende crystals. (**D**) Granulite, Adirondack Mountains, New York. Clinopyroxene (CPX) and orthopyroxene (OPX) together with twinned plagioclase and minor quartz. (**E**) Blueschist, Tinos island, Greece. Glaucophane (Glc) coexists with garnet (G) and epidote (Ep). (**F**) Eclogite, Tinos island, Greece, rich in garnet (G) and the Na- and Al-rich clinopyroxene omphacite (Omph).

been taken to be equal to 1. Furthermore, variations in other factors, such as bulk rock compositions and reaction kinetics, result in variations in the positions of facies boundaries of roughly ± 25 ºC and ± 0.1 GPa, or even more in some cases.

There is an old debate as to whether the term *facies* refers just to characteristic mineral assemblages in rock associations, or to the conditions under which such mineral assemblages develop. Today, the term is commonly used in both senses. For example, if a petrologist refers to an "eclogite facies metabasalt," they are talking about an eclogite – a rock that contains omphacite (sodic, aluminous clinopyroxene) and garnet. On the other hand, "eclogite facies conditions" refers in a very general way to the pressure–temperature regime of the eclogite facies.

The metamorphic facies fall into three general prograde sequences on the *P*–*T* diagram. One sequence extends from moderate to high temperatures at low pressures and includes the albite–epidote hornfels, hornblende hornfels, pyroxene hornfels, and sanidinite facies. These facies are normally restricted to contact metamorphic aureoles, or even xenoliths in mafic igneous rocks in the case of the sanidinite facies. (A hornfels is a hard, brittle contact metamorphic rock.) The albite–epidote hornfels facies is the lowest-temperature facies of contact metamorphism, but around many intrusions nothing lower than the hornblende hornfels facies is formed. Another sequence, which includes the zeolite, greenschist, amphibolite, and granulite facies, is marked by increasing temperature and pressure and is typical of many regionally metamorphosed orogenic belts. A third sequence, which includes the prehnite–pumpellyite, blueschist, and eclogite facies, shows a relatively small increase of temperature despite a marked rise in pressure. This sequence is the product of regional metamorphism under low geothermal gradients in subduction zones, and is commonly referred to as low-temperature/high-pressure (LT/HP) metamorphism. Some rocks are metamorphosed under *P*–*T* conditions that fall between these three series, but they are less common.

The characteristic mineral assemblages of each sequence are worth considering in detail because they are commonly encountered in nature. We focus on basalt protoliths here; other protoliths are summarized in Table 16.1. A basaltic rock in the zeolite facies would consist of chlorite (possibly with smectite in a mixed layer structure), pumpellyite, and laumontite, or chlorite, prehnite, and calcite, or even chlorite, calcite, and dolomite, depending on the exact composition of the basalt (Fig. 16.7(A)). In addition, they would of course contain quartz and probably analcite or wairakite. With increase of metamorphic grade to the greenschist facies, basaltic rocks are characterized by chlorite, epidote, and actinolite (plus albite and quartz; Fig. 16.7(B)). These rocks form dark green-colored schists, from which the facies derives its name. In the amphibolite facies plagioclase and the aluminous amphibole, hornblende, are stable and are the major phases in metabasaltic rocks along with minor amounts of either clinopyroxene or almandine garnet (Fig. 16.7(C)). In the granulite facies, hornblende is not stable; its place is taken by clinopyroxene and orthopyroxene (Fig. 16.7(D)). Note that water is released by dehydration reactions as the metamorphic grade increases, culminating in the essentially anhydrous mineral assemblages of the granulite facies.

If the progression of metamorphic facies had followed the low-temperature/high-pressure sequence, the zeolite facies would have been followed by the prehnite–pumpellyite facies, in which the basalt might have consisted of prehnite, pumpellyite, and chlorite (plus quartz and albite). Actinolite would form in the prehnite–actinolite facies. In the blueschist facies, the metabasalt would consist of the sodic amphibole glaucophane, commonly accompanied by lawsonite (or epidote), quartz, and garnet (Fig. 16.7(E)). Glaucophane imparts a prominent blue color to these rocks, which gives the facies its name. Aragonite is the stable calcium carbonate polymorph in carbonate-bearing blueschists. However, aragonite rapidly transforms to calcite during decompression, so depending on the pressure–temperature conditions, extremely fast uplift and exhumation may be required to preserve it. At still higher pressures in the eclogite facies, garnet coexists with green sodic clinopyroxene (omphacite) rich in the aluminous jadeite component (Fig. 16.7(F)). A metabasalt would therefore contain these two minerals plus accessory minerals such as quartz, kyanite, epidote, and rutile. Again, notice that the high-grade eclogite facies mineral assemblage is nearly or completely dehydrated.

The metabasaltic mineral assemblages found in the low-pressure facies characteristic of localized contact metamorphism are similar or identical to those of the corresponding regional metamorphic facies. For example, the assemblage plagioclase + orthopyroxene + clinopyroxene is found in metabasalts of the granulite, sanidinite, and pyroxene hornfels facies. A rock containing this assemblage could therefore not be assigned to any one of these facies unless associated rocks of different composition had unique mineral assemblages, or the geological setting (regional or contact metamorphic) could be determined independently. Mineral assemblages in the albite–epidote hornfels and greenschist facies are also very similar to each other. Metabasalts of the hornblende hornfels and amphibolite facies are both characterized by abundant hornblende, although garnet is often rare or absent from the hornfels.

In addition to these three broad sequences, hydrothermal metamorphism of oceanic crust on the seafloor takes place at pressure–temperature conditions ranging from the zeolite facies to the albite–epidote and even hornblende hornfels facies. Traditionally, however, these rocks are not called "hornfels" because they typically lack the hard, brittle characteristics of rocks metamorphosed in contact aureoles. Often, they are simply referred to as hydrothermally altered (or hydrothermally metamorphosed) oceanic crust. The degree of ocean floor metamorphism depends on the proximity of the rocks to the spreading center, and the temperature, composition, and availability of circulating hydrothermal fluids. Furthermore, as the oceanic lithosphere moves away from the spreading ridge, it cools and contracts, producing cracks. Calculations suggest that in old (~100 Ma) lithosphere, the cracks may reach tens of kilometers in depth and facilitate further fluid circulation and hydration (Korenaga, 2007).

Table 16.1 *Common silicate and carbonate minerals for regional metamorphic facies*

Facies	Mafic protolith	Pelitic protolith	Carbonate rock protolith	Ultramafic protolith	Notes
Zeolite	Chlorite, serpentine, clay minerals, zeolites and the related minerals analcime and wairakite, quartz, albite, prehnite, pumpellyite, calcite, dolomite	Chlorite, illite, clay minerals, quartz, albite, calcite, dolomite	Calcite, dolomite, quartz, chlorite, illite, clay minerals, albite	Serpentine group minerals, brucite, dolomite, magnesite	Zeolite facies rocks are generally undeformed and retain relict igneous or sedimentary textures.
Prehnite–pumpellyite	Chlorite, serpentine, prehnite, pumpellyite, quartz, albite, calcite, dolomite	Chlorite, muscovite, clay minerals, quartz, albite, calcite, dolomite	Calcite, dolomite, quartz, clay minerals, albite	Serpentine group minerals, brucite, dolomite, magnesite	
Blueschist	Glaucophane, lawsonite (or epidote), quartz, garnet; chlorite and albite (at lower pressures); sodic clinopyroxene (at higher pressures)	Glaucophane, phengite, lawsonite (or epidote), quartz, garnet; chlorite and albite (at lower pressures); sodic clinopyroxene (at higher pressures)	Aragonite, dolomite, phengite, glaucophane, epidote; albite and chlorite (at lower pressures); sodic clinopyroxene (at higher pressures)	Serpentine group minerals, brucite, dolomite, magnesite	Glaucophane or lawsonite are diagnostic of this facies. Phengite is a high-pressure K-mica with excess Si relative to muscovite.
Greenschist	Chlorite, actinolite, epidote, albite, quartz	*Chlorite zone*: chlorite, muscovite, quartz, albite *Biotite zone*: chlorite, muscovite, biotite, quartz, albite *Garnet zone*: muscovite, biotite, garnet, quartz, sodic plagioclase (chloritoid in Fe- and Al-rich rocks)	Calcite, dolomite, muscovite, quartz, albite (with biotite and sodic plagioclase at higher temperatures)	Serpentine group minerals, brucite, forsterite, tremolite	The Barrovian chlorite, biotite, and garnet zones are all in the greenschist facies.
Amphibolite	Hornblende, plagioclase, quartz, garnet	*Staurolite zone*: muscovite, biotite, quartz, garnet, staurolite, plagioclase *Kyanite zone*: muscovite, biotite, quartz, garnet, kyanite staurolite, plagioclase *Sillimanite zone*: muscovite, biotite, quartz, garnet, sillimanite, plagioclase	Calcite, dolomite, quartz, biotite, amphibole, diopside, K-feldspar, wollastonite	Talc, forsterite, anthophyllite, tremolite, orthopyroxene	The Barrovian staurolite, kyanite, and sillimanite zones are all in the amphibolite facies. Regional amphibolite facies metamorphism at lower pressures than the Barrovian sequence (~0.3 GPa) commonly produces andalusite, sillimanite, and cordierite.
Granulite	Clinopyroxene, orthopyroxene, plagioclase, garnet, quartz	Quartz, K-feldspar, plagioclase, sillimanite, garnet, biotite, orthopyroxene, cordierite (sapphirine coexists with quartz in UHT rocks; Fig. 16.9(B))	Calcite, dolomite, quartz, diopside, wollastonite, K-feldspar, forsterite	Forsterite, orthopyroxene, clinopyroxene (calcic)	
Eclogite	Garnet, omphacite, quartz, kyanite, rutile (coesite in UHP rocks)	Phengite, quartz, omphacite, garnet, kyanite, rutile (coesite, diamond in UHP rocks)	Aragonite, dolomite, omphacite, epidote, quartz, phengite, garnet (coesite, diamond in UHP rocks)	Forsterite, orthopyroxene, clinopyroxene (calcic), garnet	UHP coesite is generally converted to quartz in the rock matrix as the rocks are brought to the surface, but can be preserved as inclusions in high-strength minerals like garnet and omphacite (Fig. 16.9(A)).

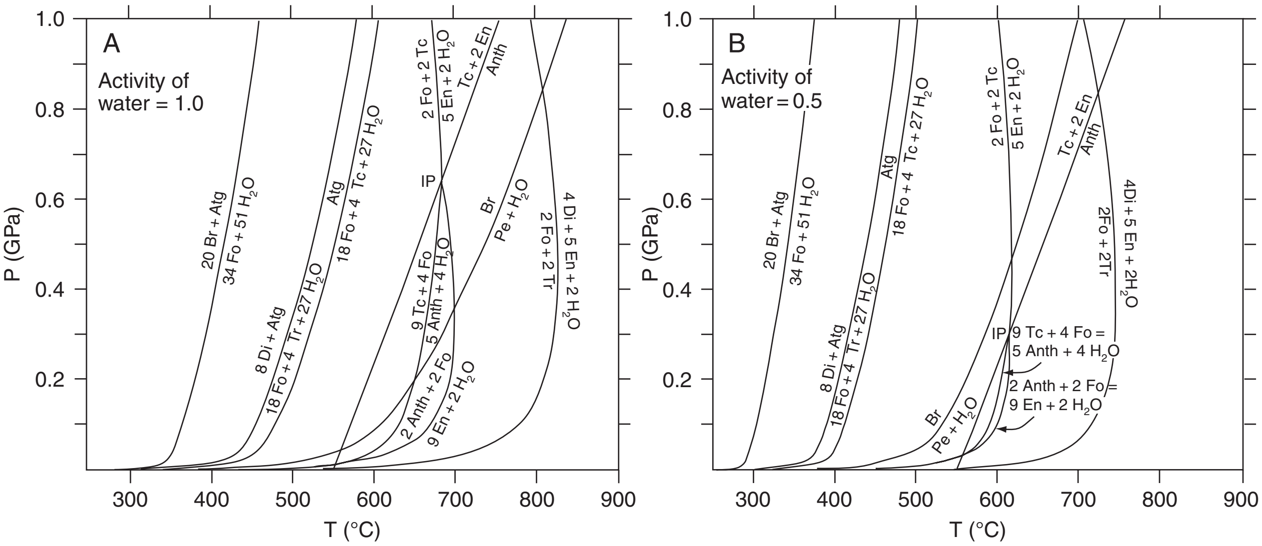


Fig. 16.8 Reactions among representative minerals in metamorphosed ultramafic rocks. Brucite (Br), periclase (Pe), antigorite (Atg), forsterite (Fo), talc (Tc), enstatite (En), anthophyllite (Anth), diopside (Di), tremolite (Tr). Invariant point denoted by IP. See text for further discussion. (**A**) Equilibrium curves for activity of water = 1.0. (**B**) Activity of water = 0.5. Note that all of the dehydration reactions are shifted to lower temperatures relative to part (A).

The presence or absence of water has important consequences for the metamorphic facies diagram. For example, consider a greenschist facies metabasalt undergoing metamorphism at 500 ºC and 0.7 GPa. The characteristic green minerals of a greenschist – epidote, actinolite, and chlorite – are hydrous, but all the minerals in "pristine" basalts are nominally anhydrous. So where does the water come from? One scenario is that it was introduced during hydrothermal metamorphism on the seafloor, such that the original basaltic minerals were converted to hydrous minerals like zeolites and sheet silicates. Subsequent burial and heating of this rock resulted in devolatilization reactions that ultimately produced the observed greenschist mineral assemblage. On the other hand, the basalt could have escaped hydrothermal metamorphism. In this case, water moving through the crust would have to have flowed into the rock and been added at some point during regional burial and heating. Of the two scenarios, the first is probably the most common. What if a metabasalt manages to make it to 500 ºC and 0.7 GPa without any significant hydration along the way? In this case, it won't develop into a greenschist! Due to the lack of water, the characteristic green hydrous minerals won't crystallize and the rock will still probably contain relict igneous minerals.

16.5 PETROGENETIC GRIDS

The boundaries between the various facies in Figure 16.6 are determined by important metamorphic reactions. In general, a number of reactions must occur at transitions between facies to convert one mineral assemblage to another. For example, formation of amphibolite from greenschist requires complex reactions involving the breakdown of chlorite, epidote, actinolite, and albite to form hornblende, plagioclase, and garnet. These reactions are unlikely to all occur at precisely the same pressures and temperatures. The discovery of rocks that formed in the transitional regions of *P–T* space between the facies led to the division of the main facies into *subfacies*. As knowledge of metamorphic rocks grew, the number of reactions that had to be accounted for in this way also grew. It became apparent that subfacies would soon become too numerous to be useful. Clearly, the only solution was to create pressure–temperature diagrams that contain all of the relevant reactions. With modern experimental and theoretical data, these diagrams can now be constructed.

To make such a diagram, one needs to define the chemical system and the phases of interest. A simple example is the system CaO–MgO–SiO_2–H_2O, which includes many minerals relevant for the metamorphism of ultramafic rocks. We will consider forsterite olivine (Mg_2SiO_4), enstatite ($Mg_2Si_2O_6$), anthophyllite (amphibole, $Mg_7Si_8O_{22}(OH)_2$), talc ($Mg_3Si_4O_{10}(OH)_2$), antigorite (serpentine group; $Mg_{48}Si_{34}O_{85}(OH)_{62}$), brucite ($Mg(OH)_2$), diopside ($CaMgSi_2O_6$), and tremolite ($Ca_2Mg_5Si_8O_{22}(OH)_2$). Pure water with an activity of one is assumed to be present. The pressure–temperature conditions for reactions among these minerals are known from either experimental studies or calculations based on thermodynamic data derived from experiments. The equilibrium *P–T* conditions for the reactions can be easily calculated using thermodynamic programs such as THERMOCALC (Holland and Powell, 1998; www.earthsci.unimelb.edu.au/tpg/thermocalc) or winTWQ (Berman, 1991, 2007; http://gsc.nrcan.gc.ca/sw/twq_e.php).

When the reactions are plotted on a *P–T* diagram for a given activity of water, they form a network of lines known as a *petrogenetic grid* (or a *P–T projection*; Fig. 16.8(A)). The grid is divided into areas that are bounded by two or more reactions. Consider the dehydration reaction: 20 brucite + antigorite → 34 forsterite + 51 H_2O. Brucite and antigorite are stable on the low-temperature side of the reaction, and forsterite is stable on the high temperature side. On the low-temperature side, the phase rule ($f = c + 2 - \phi$; Eqs. (8.12), (10.11)) says that for three components and three phases (including water), there are two degrees of freedom and the equilibrium is divariant. In this case, the two degrees of freedom correspond to temperature and pressure. Within the stability field of antigorite + brucite, therefore, either pressure or temperature can be varied without changing the mineral assemblage. The high-temperature side is also divariant, but here note that only two components (e.g. Mg_2SiO_4, H_2O) are needed to describe the system of two phases (forsterite and water). Along the reaction line itself, there are three components, and four phases (including water), so there is one degree of freedom and the reaction is univariant. Elsewhere on the diagram, multiple reactions (involving five phases) intersect at a unique invariant point with no degrees of freedom.

The power of the petrogenetic grid is that it can facilitate estimation of the pressures and temperatures of metamorphism. For example, ultramafic rocks containing antigorite and brucite in a regional metamorphic terrane probably did not reach temperatures much above about 450–500 ºC. A rock containing talc was probably metamorphosed at temperatures in excess of 550 ºC. A rock that contains the invariant point assemblage of Figure 16.8(A) constrains the temperature and pressure to about 700 ºC and 0.65–0.7 GPa.

Of course, there are several important caveats to consider when analyzing a petrogenetic grid such as Figure 16.8(A). First, variations in the activity of water can have a dramatic effect on the reaction curves. In general, dehydration reactions are shifted to lower temperatures as water activity decreases (Fig. 16.8(B)). Second, the natural minerals are solid solutions. The substitution of iron, for example, into mineral structures changes their thermodynamic properties and, thus, the positions of reactions. Finally, the grid was drawn assuming thermodynamic equilibrium. Many more reactions can be written among the phases in the model ultramafic rock, but these *metastable* reactions are thermodynamically less stable than the ones shown. If thermodynamic equilibrium was not achieved due to sluggish rates of reaction, then metastable reactions could have operated to produce nonequilibrium (or *disequilibrium*) mineral assemblages. Nonetheless, if all these caveats are taken into consideration and accounted for, petrogenetic grids can be extremely valuable for deciphering the pressures and temperatures of metamorphism. It turns out that reactions in many natural ultramafic rocks are reasonably well approximated by Figure 16.8(A).

Univariant dehydration reactions, like the reaction of antigorite to produce talc, forsterite, and water, generally have a curved shape on pressure–temperature diagrams – an important consideration for interpretation of petrogenetic grids (Fig. 16.8). Water is on the high-temperature (high-entropy) side of the reaction, so the ΔS of dehydration is positive. At low pressures, the volume change of reaction is positive and very large, so the slope of the reaction curve in P–T space ($dP/dT \quad \Delta S/\Delta V$; Eq. (8.3)) is shallow and positive when the water activity is 1. Because of the compressibility of the vapor phase, however, the ΔV of the reaction decreases sharply with increasing pressure, with the result that the slope of the reaction curve increases, becoming steep but still positive above 0.2 GPa. If the activity of water is less than 1, for example due to the presence of CO_2 in the fluid phase, the reaction is driven to the right in order to try and build up the water pressure, and thus the stability of the reactant minerals is pushed to lower temperatures (Fig. 16.8(B)). As the water activity decreases from 1, the slope of the dehydration reaction steepens and can become negative as it pivots about its intersection with the temperature axis (Figs. 16.8, 11.11).

Solid–solid reactions do not produce or consume fluids. Examples include anthophyllite = talc + enstatite (Fig. 16.8), or the albite = jadeite + quartz and calcite = aragonite reactions shown on Figure 16.6. These reactions lack the pronounced curvature of dehydration reactions. As discussed above, this curvature is mostly the result of the variation in the molar volume and compressibility of water as functions of pressure and temperature. Solids, on the other hand, undergo much smaller volume changes than water does. Thus, the $\Delta S/\Delta V$ values and dP/dT slopes of solid–solid reactions remain fairly constant over considerable ranges of pressure and temperature. Of course, both ΔS and ΔV for solid–solid reactions do vary somewhat as pressure and temperature change. This can produce subtle curvature, as is evident for the reaction albite = jadeite + quartz in Figure 16.6. As we will see in later chapters, solid–solid reactions are very useful for petrogenetic grid analysis because their pressure–temperature relations are not affected by variations in the activity of water. Note, for example, how the position of the anthophyllite = talc + enstatite equilibrium does not change as the activity of water is varied (Fig. 16.8).

16.6 ULTRAHIGH-PRESSURE AND ULTRAHIGH-TEMPERATURE METAMORPHISM

One of the most exciting developments in metamorphic petrology in recent years is the recognition of extreme pressures and temperatures of regional metamorphism (Fig. 16.6). Ultrahigh-pressure (UHP) eclogites and related metamorphic rocks develop deep within subduction zones where the high-pressure silica polymorph coesite is stable at pressures in excess of about 2.7 GPa (Fig. 16.9(A)). These pressures correspond to depths in excess of ~90 km; in fact, some UHP rocks are inferred to have formed at more than 200 km depth. In some cases, pressures were so high that even diamond was stable. UHP rocks have been recognized in a number of orogenic belts, including the Sulu-Dabie terrane of western China, the Western Gneiss region of Norway, and Alpe Arami in the western Alps (Fig. 16.10). Interestingly, UHP rocks have not been discovered in North America or Australia. The pressure–temperature conditions of UHP metamorphism and the processes that exhume UHP rocks from great depths to the surface are the subjects of much current research (Maruyama and Liou, 1998; Chopin, 2003).

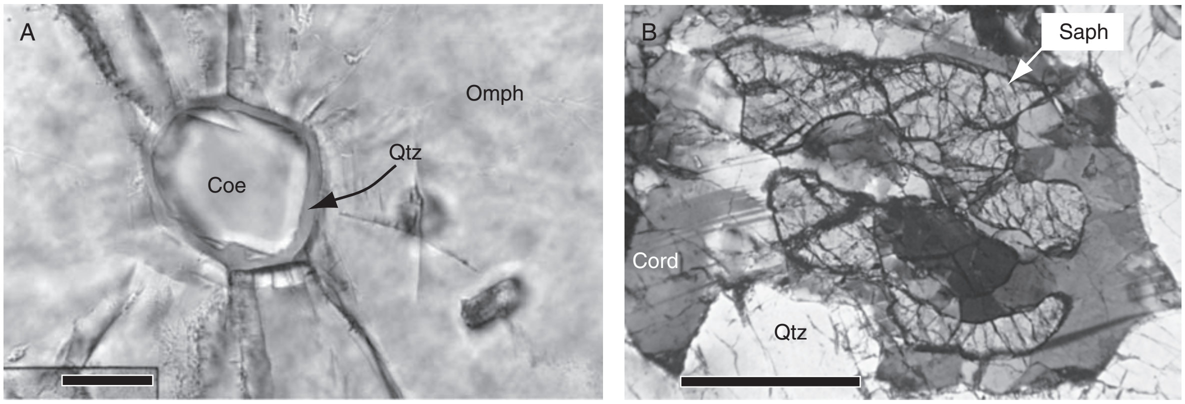


Fig. 16.9 Photomicrographs of characteristic minerals in UHP and UHT rocks. (**A**) Inclusion of coesite in omphacite pyroxene (Omph) from UHP eclogite, central Erzgebirge, Germany. Plane-polarized light; scale bar = 40 μm. The coesite (Coe) was partially replaced by quartz (Qtz) which developed during decompression as the rocks were brought to the surface. Quartz has a much larger molar volume than coesite, so the phase change resulted in a considerable volume increase which, in turn, produced the radial cracks in the omphacite extending away from the inclusion. (From Massone, 2001; published by permission of the *European Journal of Mineralogy* (www.schweizerbart.de.) (**B**) Sapphirine (Saph) rimmed by retrograde cordierite (Cord) in UHT granulite from Edwards Island, Antarctica. Crossed-polarized light; scale bar = 0.5 mm.

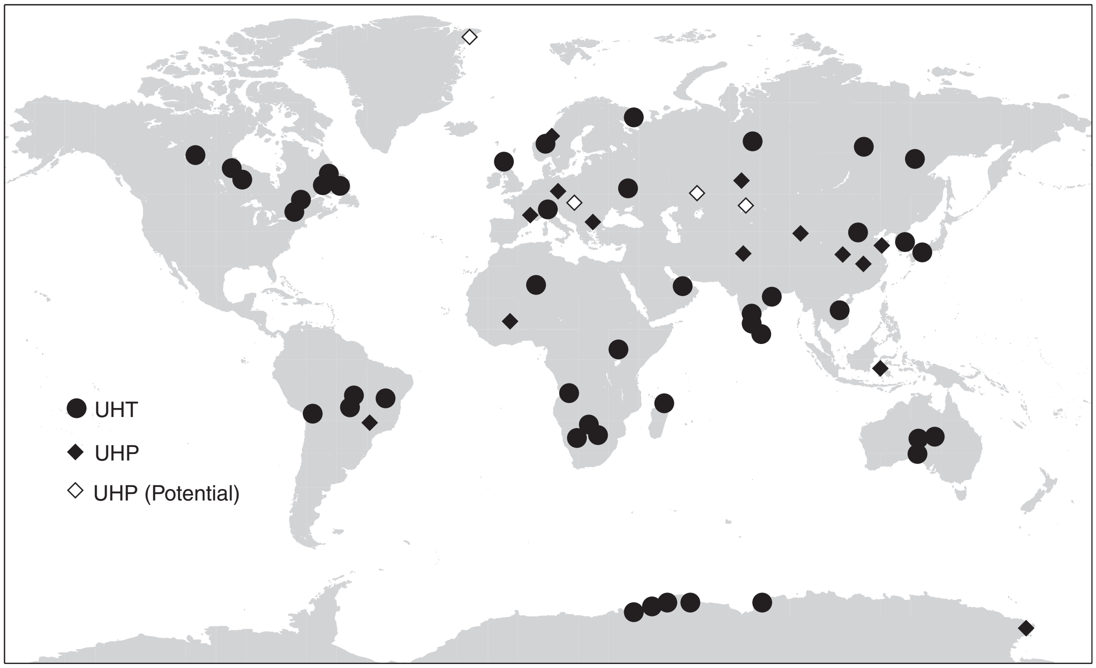


Fig. 16.10 Map of UHP and UHT localities (modified after Maruyama and Liou, 1998; Chopin, 2003; Kelsey, 2008).

Ultrahigh-temperature (UHT) granulite metamorphism occurs at temperatures of about 900–1000 °C or greater. Magnesium and aluminum-rich rocks contain coexisting sapphirine ($Mg_4Al_8Si_2O_{20}$) and quartz indicative of UHT metamorphism, as found, for example, on Edwards Island, Enderby Land, Antarctica (Fig. 16.9(B); Fig. 16.10). The tectonic and petrologic environments that produced these extreme temperatures are matters of active research (Harley, 2004; Brown, 2006; Kelsey, 2008).

UHP and UHT rocks have significantly different age distributions and form in different tectonic settings. Therefore, a global map of localities shows that the two are not found together (Fig. 16.10). Although Phanerozoic UHT granulites are known, many are Precambrian, ranging in age from Archean to Neoproterozoic. This accounts for the clustering of UHT localities within Precambrian shields in Canada, Australia, Antarctica, and elsewhere. Although most UHT rocks exposed at the surface are quite old, UHT metamorphism may even be occurring today in tectonic settings such as rift zones and continental back-arc regions (Brown, 2006). Known UHP rocks are Late Proterozoic age or younger. UHP metamorphism may have occurred earlier in Earth's history but, if so, the rocks have not been discovered yet, or they are not exposed. Exhumed UHP metamorphic rocks are associated with collisional mountain belts where continental subduction has occurred, and appear to be absent from Pacific-type subduction plate boundaries. We emphasize that new discoveries of UHP and UHT rocks are likely, so interpretations will no doubt change and evolve in the coming years.

16.7 PROBLEMS

16.1 If a shale containing 5.0 wt% water is metamorphosed to an entirely anhydrous set of minerals (granulite facies), calculate the volume of fluid formed, relative to the volume of metamorphic rock, at 800 °C and 0.7 GPa if the density of water under these conditions is 770 kg m^{-3} and that of the rock is 2900 kg m^{-3}.

16.2 A section of crust is made up of a 20-km-thick sequence of metasedimentary rocks that extend from 15 to 35 km depth and that lose, on average, 3 wt% water during prograde metamorphism.

(a) Assuming that the water flows vertically and upward, calculate the total flux of fluid released at the top of the metasedimentary sequence in units of m^3 fluid per m^2 rock area. This quantity is an example of a time-integrated fluid flux (Chapter 21). Take the densities of the rock and water to be 2800 and 950 kg m^{-3}, respectively.

(b) If the rock sequence undergoing dehydration in part (a) underlies an orogenic belt 200 × 50 km in areal extent, calculate the total volume of fluid (in m^3) released at the top of the metasedimentary sequence over this area.

(c) The Great Lakes of the Midwestern United States and Canada constitute the world's largest system of fresh, liquid surface water. Water volumes range from 484 km^3 for Lake Erie, to 12 100 km^3 for Lake Superior. How does the volume of fluid calculated in part (b) compare with the volumes of the Great Lakes?

16.3 **(a)** Calculate the equilibrium temperature for the dehydration reaction: 20 brucite + antigorite → 34 forsterite + 51H_2O at 0.5 GPa, assuming pure solids and unit activity of water. For many reactions, $\ln K$ (Eq. (9.56)) is nearly linear when plotted against $1/T$. The $\ln K$ for the reaction at 0.5 GPa can be described by the linear relationship: $\ln K = 275.1 - 1.9 \times 10^5/T$ (T is in kelvin). Note that this relation takes account of the fugacity term (Eq. (8.9)), so K is simply equal to $(a_{H_2O})^{51}$.

(b) Calculate how the position of the equilibrium shifts at 0.5 GPa for water activities of 0.75, 0.5, and 0.25.

16.4 The Al_2SiO_5 triple point lies at a pressure of about 0.4 GPa. Use the distribution of Al_2SiO_5 polymorphs in Figure 16.1(B) to draw a line on the map that separates the metamorphic belt into two regions: one that was metamorphosed at pressures $\leq$0.4 GPa, and another that was metamorphosed at $\geq$0.4 GPa. What can you conclude regarding the pressures of Barrovian and Buchan metamorphism in Scotland?

16.5 **(a)** A metamorphosed ultramafic rock is heated from 300 to 700 ºC at 0.3 GPa and unit activity of water. From Fig. 16.8(A), what will the sequence of equilibrium reactions be if the rock initially contains only antigorite?

(b) Repeat part (a), but this time for a rock that contains antigorite and a small amount of brucite.

(c) Repeat part (a), but this time for a rock that contains a large amount of brucite and little antigorite.

(d) Repeat part (a), but this time for a rock composed of pure brucite. From the results of parts (a)–(d), does the initial bulk mineralogical composition of the rock affect the sequence of prograde reactions?

(e) The equilibrium curves for the reactions brucite = periclase + H_2O and 2 forsterite + 2 tremolite = 4 diopside + 5 enstatite + 2H_2O intersect at about 0.85 GPa and 810 ºC. Does this intersection constitute an invariant point? Explain.

(f) How would Fig. 16.8(A) change if the chemical system was $MgO–SiO_2–H_2O$?

16.6 Typical contact aureoles, such as those bordering the Onawa pluton (Fig. 16.1(A)) are several kilometers wide. Consider a hypothetical, spherical granitic intrusion having a radius of 5 km. It is emplaced rapidly at 750 ºC into country rocks initially at 300 ºC. Assuming that the pluton and the country rocks have the same thermal diffusivity, k_T, we can use the spherical geometry analog of Eq. (5.20) to solve for temperature as a function of time. The spherical solution is well known and is given below (e.g. Crank, 1975). T_0 is the initial temperature of the magma, a is the radius of the sphere, and r is the distance from the center of the sphere. Note that the solution is for an initial country rock temperature of 0 °C, so T_0 must be adjusted as described in Section 5.3. Try using units of meters and years; for this case, $k_T = 10^{-6}$ m^2 s^{-1} = 31.5 m^2 a^{-1}.

$$T = 0.5\,T_0\left(\operatorname{erf}\frac{a-r}{2\sqrt{k_T t}} + \operatorname{erf}\frac{a+r}{2\sqrt{k_T t}}\right) - \frac{T_0}{r}\sqrt{\frac{k_T t}{\pi}}\left[\exp\left(-(a-r)^2/4k_T t\right) - \exp\left(-(a+r)^2/4k_T t\right)\right]$$

(a) Consider lateral heat conduction away from the pluton. Make a plot of temperature (y-axis) versus time (1 to 10^6 years) for the country rock at a distance of 0.001 km from the pluton–country rock contact, and for 1 km, 2 km, and 5 km from the contact. Is the timing of "peak" (maximum) temperature conditions the same regardless of distance from the contact? Explain.

(b) The calculations in part (a) incorporate many simplifications. List and discuss at least five major simplifications.

17 Deformation and textures of metamorphic rocks

17.1 INTRODUCTION

The mineralogical composition of a metamorphic rock can be used to determine pressures, temperatures, and fluid compositions during metamorphism. But this is only part of the record these rocks preserve. Their textures are also a source of valuable information and in most cases are diagnostic of a particular type of metamorphism – contact versus regional, for example. Textures are important in working out the timing of metamorphic events, especially in regionally metamorphosed rocks. Indeed, the interpretation of metamorphic textures is an important part of structural geology and some books have been devoted entirely to this topic. Only a brief introduction can be given here, but for more detailed coverage reference can be made to texts by Turner and Weiss (1963), Spry (1969), Hobbs *et al*. (1976), Suppe (1985), Passchier and Trouw (1996) and Vernon and Clarke (2008).

Many of the textures in metamorphic rocks result from the recrystallization of existing minerals or the growth of new ones. The general principles governing these processes were dealt with in Chapter 12, but little attention was given to the effects of deformation, other than to indicate that strain energies provide an additional driving force for recrystallization. Regional metamorphism is normally accompanied, at some stage in its development, by strong deformation, which results in the formation of a number of characteristic textures. It is these textures with which this chapter is mainly concerned. *Brittle* deformation involves fracturing at the grain scale or larger, whereas *ductile* or *plastic* deformation proceeds without fracturing. Some of the textural terms have already been used in previous chapters, but it is worth systematically reviewing these in light of what they can reveal about the deformation history of a rock.

Textures in regionally metamorphosed rocks can develop prior to, during, or after the main period of deformation. Rocks may even undergo more than one period of metamorphism, in which case textures may be complicated by relict ones from earlier episodes. Radiometric age determinations from the cores and rims of zoned crystals such as garnet, monazite, and zircon can provide absolute ages for multiple metamorphic events. Normally, however, only relative ages are determined from interpretations of both the textures of the rock and the outcrop-scale structures such as folds, faults, and deformed primary features. Once unraveled, the relative and absolute timing information, in conjunction with the interpreted pressure, temperature, and fluid composition, allows us to estimate pressure–temperature–time paths for rocks during metamorphism. This topic is dealt with in Chapter 22.

Metamorphism normally involves the growth of grains. These may form as new minerals produced by metamorphic reactions, or they may result from the recrystallization of old ones; less commonly, the process simply involves grain breakage. We will first examine the textures developed by the growth of new minerals in the presence of a stress field, leaving to later the textures developed by recrystallization and grain breakage.

17.2 METAMORPHIC FOLIATION

The most prominent textural feature exhibited by regionally metamorphosed rocks is a planar fabric, which is referred to generally as *foliation*. It is produced mainly by the parallel or subparallel arrangement of platy and elongate minerals. Minerals in contact metamorphic rocks, on the other hand, tend to have random orientations. The resulting rock, known as a *hornfels*, is characterized by extreme toughness, which is caused by the interlocking randomly oriented crystals.

When fine-grained pelitic sedimentary rocks undergo regional metamorphism, muscovite and chlorite are among the first minerals to form. These grow with a preferred orientation, which produces a prominent foliation referred to as *slaty cleavage* (Fig. 17.1). The micaceous minerals in slate are too fine-grained to be visible with the unaided eye. As they coarsen, the foliation planes develop a sheen and generally become slightly less regular. This rock is referred to as *phyllite*. Individual crystals are still too small to be visible to the unaided eye in phyllite. When they do become visible, the rock is called a *schist*. Schists can also form from mafic igneous rocks; these typically contain abundant chlorite, epidote, and/or amphibole (Fig. 16.7(B)). In sedimentary rocks of silty and sandy composition and in intermediate-to-felsic igneous rocks, fewer micaceous or needly minerals are formed during metamorphism. Those that are present tend to be concentrated into layers that parallel the foliation defined by the micas or amphiboles. Quartz and feldspars tend to concentrate in the alternate layers. Such rocks are referred to as *gneiss*. No hard-and-fast rule separates schists from gneisses. If micaceous or needly minerals predominate, the rock is a schist; if the granular minerals

Fig. 17.1 Slaty cleavage parallel to axial plane of fold in interlayered slate and limestone. Foliation planes are marked by concentrations of clay minerals and graphite. Note refraction of slaty cleavage near competent carbonate vein, which has also been folded. Width of field is 1 cm.

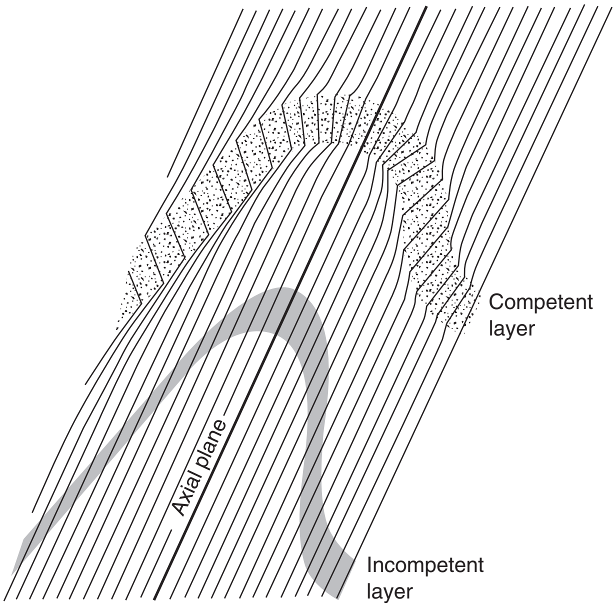


Fig. 17.2 In folded rocks, metamorphic foliation commonly parallels the axial plane of folds. Competent layers of rock, however, tend to buckle during folding, which rotates axial plane foliation into fan-like patterns across fold axes.

predominate, it is a gneiss. In general, schists tend to pulverize when hit with a geological hammer, whereas gneisses fracture.

When first developed in the lowest grades of metamorphism, foliation can be shown to coincide with the plane of maximum extension, as defined by deformed primary features, such as fossils and ooids in limestone, pebbles in quartzite, and reduction spots in slates. The reduction spots are particularly useful because they are initially spherical, and their physical properties, apart from color, are identical to those of the surrounding rock. Consequently, they form ideal passive strain markers. During deformation they are deformed into flattened ellipsoids, with their plane of maximum flattening coinciding with the slaty cleavage.

Foliation may develop parallel to shear planes, but this does not appear to be common. This is not to say that shearing does not occur along foliation planes. Indeed, once formed, a foliation is difficult to eliminate from a rock, and the foliation provides planes of weakness on which later movement can occur. As will be seen below, there is considerable evidence for shear on foliation planes in metamorphic rocks.

In folded rocks, foliation parallels or nearly parallels the axial plane of folds (Fig. 17.1). In slaty and schistose rocks, the foliation is almost everywhere parallel to the axial plane. Where more competent, quartz- or calcite-rich layers are interlayered with slate or schist, the foliation may fan across or converge toward fold axes (Fig. 17.2). This produces a refraction of the foliation planes as they cross from one layer into another. In such cases, all the foliation planes were probably parallel to the axial plane when first developed, but competent layers of rock buckle during folding and so rotate the foliation planes. Within folds, then, the foliation again generally parallels the plane of maximum flattening.

Several different mechanisms can cause minerals to become oriented during metamorphism. Minerals that form plates, blades, or needles have growth rates that clearly vary with crystallographic direction. Nuclei of these minerals that are oriented with their directions of maximum growth rate perpendicular to the maximum principal compressive stress can be expected to grow more rapidly than those with other orientations. Thus, even if nuclei form with random orientations, the unequal growth rates result in the development of a preferred orientation, with the majority of platy and elongate minerals lying in the plane normal to the maximum compressive stress (Fig. 17.3). In addition, pure shear tends to rotate platy and elongate minerals into this same plane. This is particularly true of clay minerals, which, during compression and dewatering, develop a strong preferred orientation. The growth of micaceous minerals from these clays during metamorphism is therefore likely to result in the nucleation of grains that already have a strong preferred orientation.

Dissolution–precipitation reactions can also cause grains to become flattened in the plane of foliation. Quartz grains commonly deform in this fashion. Tectonic compression creates stresses at points of contact between quartz grains, or between quartz grains and other minerals like chlorite and micas. As noted in Sections 12.6 and 16.2, stress increases the

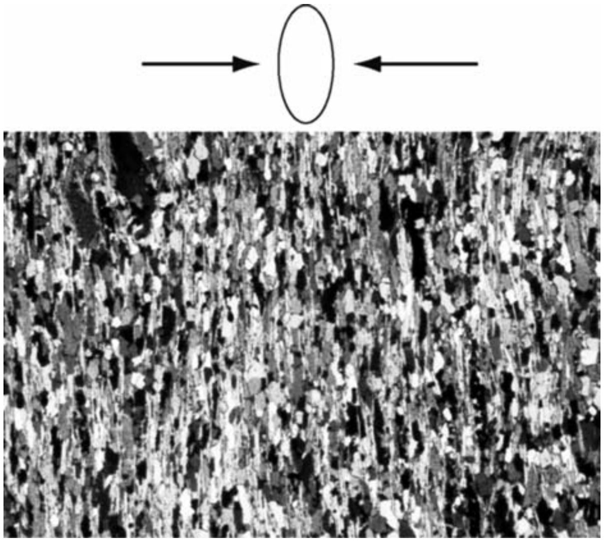


Fig. 17.3 Schistosity defined mainly by alignment of muscovite and chlorite, but quartz grains are also flattened parallel to foliation. Some biotite grains (dark, stippled) are not aligned parallel to foliation and must have grown later when the stress conditions were different. The foliation plane is normal to the direction of maximum compression as shown by the strain ellipse. Width of field is 3 mm; crossed-polarized light.

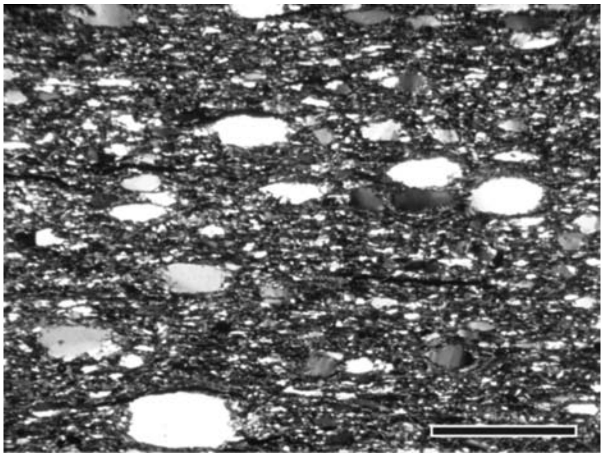


Fig. 17.4 Quartz grains in metasandstone flattened parallel to the foliation (horizontal) by pressure solution processes. Rock matrix between quartz grains consists mostly of muscovite and chlorite. Note that quartz grains are neither fractured nor deformed internally by intracrystalline deformation, indicating that they changed shape by dissolution–precipitation processes. Greenschist facies, Barrovian zones, northeastern Scotland. Crossed-polarized light. Scale bar is 1 mm.

solubility of minerals and, consequently, the quartz grains tend to dissolve at points of contact by *pressure solution*. The dissolved material migrates through pore fluids away from the high-stress contacts to the sides of the crystals normal to the maximum compressive stress where solubility is lower, a process referred to as *solution transfer*. These zones can then become sites for local reprecipitation; alternatively, the dissolved material may precipitate in the matrix, or leave the rock volume altogether. Both dissolution and local precipitation on grain margins act to flatten the grains in the plane of the foliation perpendicular to the maximum compressive stress (Fig. 17.4). Pressure solution appears to be most effective at low to moderate metamorphic grades. It may affect other minerals as well, including carbonate minerals and the feldspars. Furthermore, as we will discuss below, it is probably important in producing differentiated metamorphic fabrics (Fig. 16.5(B)).

The amount of dissolution–precipitation that occurs will be limited by porosity and the availability of fluids. In most cases, the rock being deformed already has had its porosity reduced to a minimum, and therefore no large pores are present into which dissolved material can be transported. If tension fractures open, they provide sites for redeposition, but these are soon filled and become veins. Material dissolved by the stress soon saturates the extremely small amount of pore fluid remaining in the rock, thus preventing further dissolution. If stresses continue to build and the pore fluid cannot be changed, deformation of the rock must occur by some process other than dissolution; this could be by a process as slow as dislocation creep or as rapid as thrust faulting. However, if the intergranular fluid is flushed out of the rock with fluid that is undersaturated in the phase being dissolved, the stress on the rock is released by dissolution.

Fluids that could cause such flushing are likely to be common in orogenic belts as products of metamorphic dehydration reactions. Under some circumstances, these fluids may travel along grain boundaries, but as flow and dissolution continue, undulous interconnecting channels develop along which the flow is concentrated. Once established, the channels become the main sites for dissolution. In hand specimens, they are marked by a buildup of insoluble residues. Closely spaced channels slowly migrate toward each other and become one as the intervening rock is dissolved away. This results in the development of *solution cleavage* (Geiser and Sansone, 1981), which is a prominent feature of many carbonate rocks in orogenic belts. Figure 17.5 shows an example of such cleavage in a limestone composed largely of foraminifera. The tests or shells of this type of foraminifera have a flat elliptical cross section. Where the tests abut against the solution cleavage, which is marked by zones of dark insoluble residue, they have square truncated terminations as a result of dissolution.

When flattening occurs in rocks that contain rigid grains, such as cubes of pyrite in slate or garnet crystals in schist (Fig. 17.6), zones of relatively low pressure develop on the sides of the crystals that are normal to the maximum compressive stress. These zones become the locus of precipitation of the mineral that is most soluble under pressure in the surrounding rock; this is normally quartz. This produces what are referred to as *pressure shadows*. Because most of these taper to a point away from the rigid grain, they resemble a goatee and thus are also referred to as *beards* (Fig. 17.6).

Within foliation planes, elongate minerals such as amphiboles may be aligned so as to produce a *lineation*. Linear aggregates of equant grains, elliptical reduction spots

Fig. 17.5 Foraminifera limestone with solution cleavage marked by seems of clay-rich insoluble residue. These forams normally have rounded terminations; their abrupt termination against the clay is clear evidence of dissolution. The width of field is 10 cm.

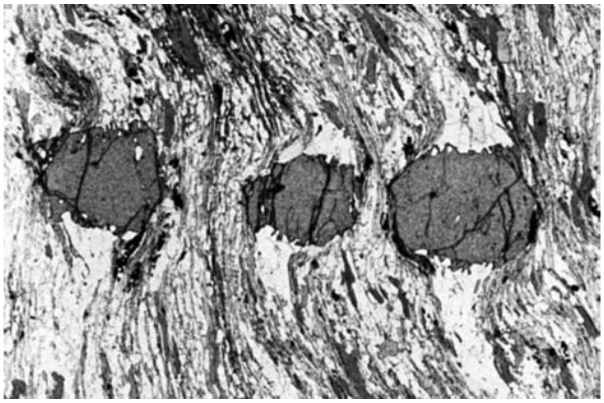


Fig. 17.6 Rigid grains, such as these garnet crystals in schist, are commonly flanked by tapering wedges of quartz, which because of their shape are named beards. Quartz is thought to migrate to regions of low pressure, so beards are also referred to as pressure shadows. Note quartz grains in the beards are intergrown with the garnet in the rims of these crystals. Because the cores of the garnet crystals contain no quartz inclusions, the beards probably developed late in the growth of the garnet. The garnets are slightly elliptical in cross section; this is a growth phenomenon and was not caused by deformation of the crystals. Bolton schist, Connecticut. Width of field is 7 mm; plane-polarized light.

in slate, and stretched pebbles in quartzite (Fig. 17.7) are other features capable of defining lineations. Some rocks have lineation but no foliation. In folded rocks, lineations lie in the axial-plane foliation, where they either parallel fold axes (*b* lineation) or are perpendicular to them (*a* lineation). In zones of strong shear, such as near thrust faults, lineations may parallel the tectonic transport direction (*a* lineation).

Foliation and lineation are described as *penetrative* structures because they are pervasive throughout the rock. Slaty cleavage, for example, is visible under the microscope down to the level of individual grains. A fault, by contrast, is *nonpenetrative* because it forms a discrete feature. Penetrative structures are useful because they provide, from a hand specimen or thin section, information on the orientation of larger nonpenetrative structures such as axial planes and plunges of folds that may not be exposed.

The foliation plane in many schists is folded into small crenulations which have a wavelength of a few millimeters to about a centimeter. These small folds define a *crenulation lineation*. These lines actually mark the intersection of a second, more widely spaced schistosity with the first schistosity. In cross section, the early schistosity is seen to be bent into sigmoid curves by the later, nonpenetrative *crenulation cleavage* or *schistosity* (Fig. 17.8). Within the bends the amount of quartz is substantially less than in the surrounding rock. This results in the crenulation cleavage being marked by mica-rich layers. This new compositional layering must not be mistaken for bedding. Crenulation cleavage can result from successive episodes of metamorphism, but in many cases it is produced by a single episode during which the orientation of the rock changes with respect to the stress field, possibly as a result of folding.

Nowhere are the effects of dissolution–reprecipitation more noticeable than in the development of crenulation cleavage. The folding of the early schistosity by the crenulation cleavage was initially interpreted as resulting from a slip parallel to the later cleavage; for this reason it has also been referred to as *strain-slip cleavage*. Inspection of Figure 17.8, however, reveals that the crenulations are not the result of simple folding of the earlier schistosity. The early foliation is defined by layers that are alternately quartz-rich and muscovite-rich. Where these layers begin to get folded by the crenulations, the layers of quartz decrease in thickness, until they completely vanish in the main foliation plane of the crenulation cleavage. The crenulation folds are therefore produced by the shortening of the rock normal to the crenulation cleavage planes as a result of pressure solution of quartz. The concentration of muscovite in the crenulation cleavage is formed simply as an insoluble residue.

Figure 17.9 illustrates how the geometry of crenulation cleavage develops as a result of selective dissolution of quartz (Problem 17.1). As we will discuss in Chapter 21, fluid capable of dissolving quartz is not likely to exist along all grain boundaries because of wetting problems, and as a result dissolution is most rapid where channels develop, and these become the new schistosity planes. The geometry of the crenulation cleavage does not indicate what happens to the dissolved quartz. It might, for example, be completely removed from the rock, but it could also be added to the already existing quartz layers, making them thicker. If the silica was transported out of the rock mass, the fluid fluxes required can be enormous (Chapter 21). Regardless of the ultimate fate of the dissolved quartz, the transport of material plays an important role in producing crenulation cleavage and, possibly, in determining rock composition.

While platy and elongate minerals become aligned during deformation to form a foliated fabric, other minerals that exhibit no morphological anisotropy may become

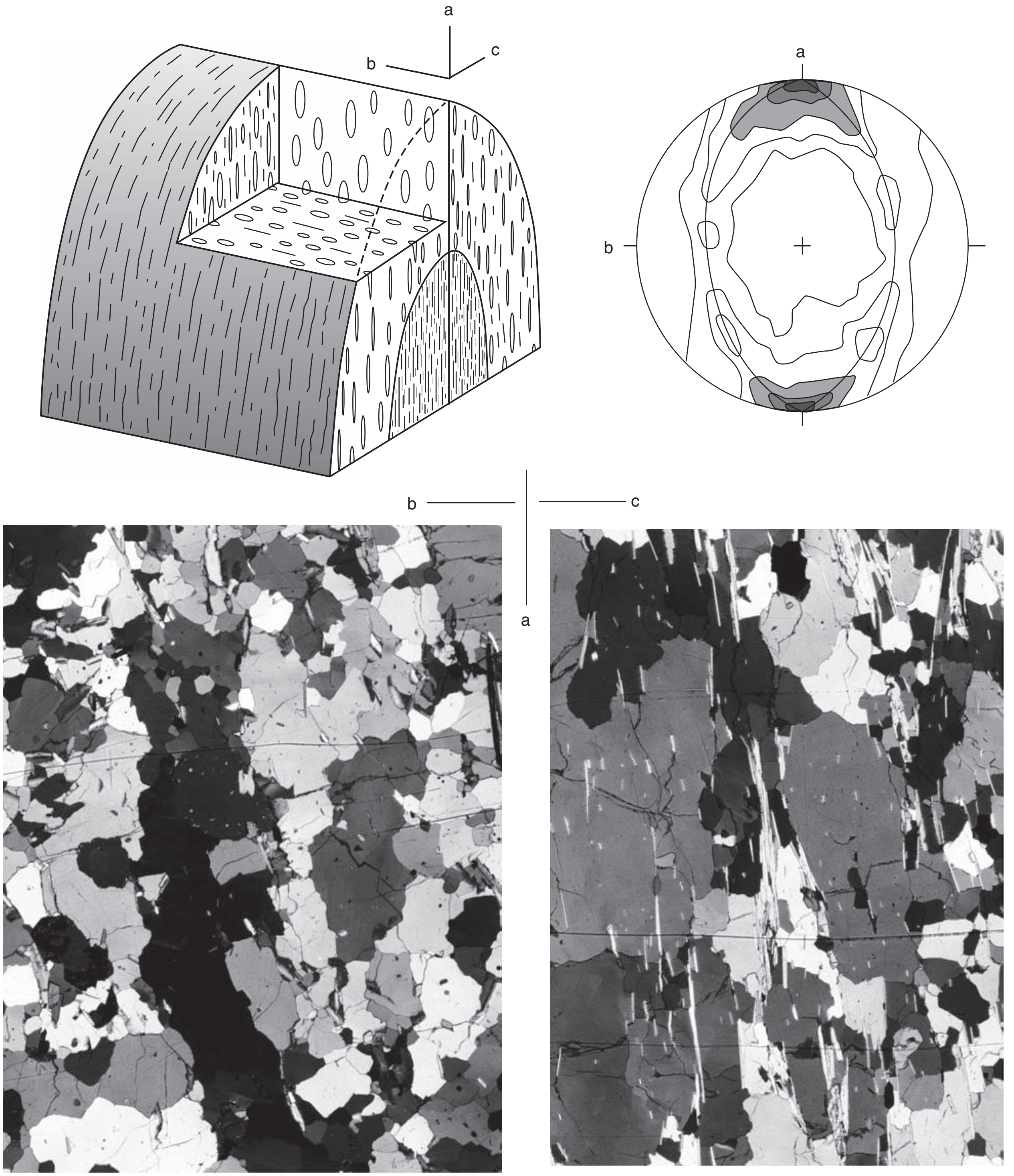


Fig. 17.7 Folded quartzite containing quartz pebbles which are flattened in the foliation plane (*a–b*) and elongated parallel to *a*. Photomicrographs illustrate appearance of rock in sections parallel to the *a–b* and *a–c* planes (width of each field, 5 mm; crossed polars). Insertion of a first order red interference filter causes grains to turn all blue or all yellow, indicating a strong preferred crystallographic orientation. This is shown in the stereographic plot of optic axes of quartz grains, most of which plot along two great circles at 45° to the foliation plane (*a–b*), which intersect to form a strong maximum parallel to *a*, the direction of pebble alignment. Clough quartzite, Connecticut.

crystallographically aligned; quartz and calcite are two common examples. Such preferred orientation is not as readily apparent as the morphological alignment but is nonetheless an important part of the metamorphic reorganization of a rock. The quartz grains in the quartzite illustrated in Figure 17.7 have a strong preferred crystallographic orientation, which is made evident under the microscope when a first-order red interference filter is inserted under crossed

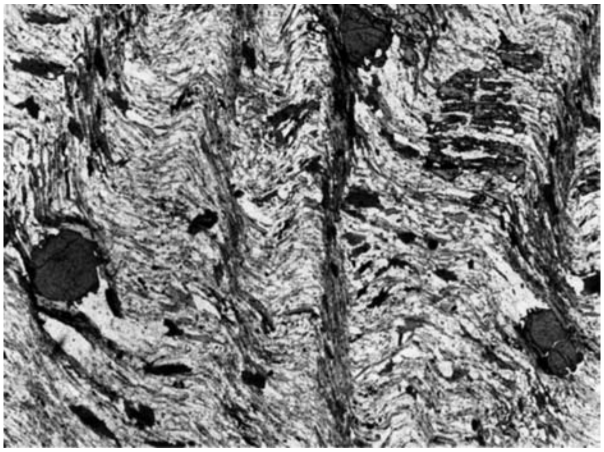


Fig. 17.8 Deformation of an early schistosity (top left to lower right) by a later one gives rise to crenulation cleavage. The crenulation cleavage is marked by a decrease in the abundance of quartz in the crenulation planes. Offset of the earlier schistosity by the crenulation cleavage is most likely the result of dissolution of quartz from within the crenulation zones rather than of shear. The width of field is 15 mm.

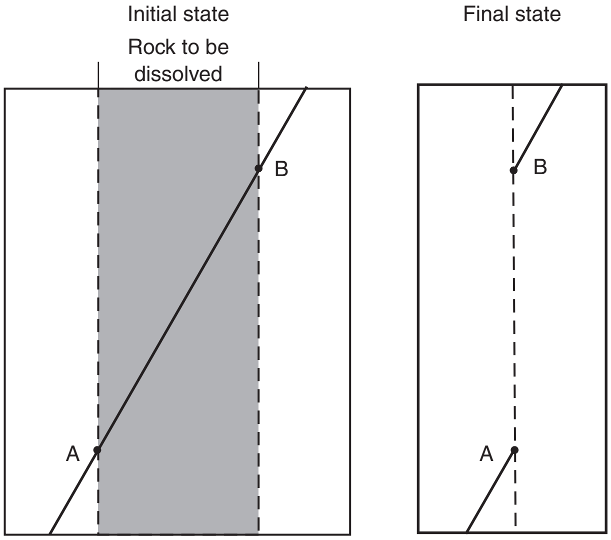


Fig. 17.9 Geometry of crenulation cleavage developed by dissolution of rock composed of mica and quartz. Within the shaded area, all quartz is removed by solution, leaving mica as an insoluble residue in the newly formed crenulation zone. The early schistosity appears offset between points A and B on the crenulation cleavage plane, but the displacement of the rock has only been in a direction normal to this cleavage plane.

polars and all of the quartz grains turn blue (or all turn yellow). The grains, therefore, cannot be randomly oriented, for there would then be equal numbers of blue and yellow grains.

Precise determination of preferred crystallographic orientation requires time-consuming measurements, which are normally made optically. The orientation of some easily identifiable crystallographic direction is determined using a universal microscope stage. In quartz and calcite, the optic axis (*c*-axis) provides such a direction. A large number of grains are measured in thin section, and the results are plotted in an equal-area stereographic projection. The density of points in the plot is then contoured to give a statistical measure of the preferred orientation of the grains. In recent years, scanning electron microscopes equipped with electron-backscatter diffraction equipment have been used to determine crystallographic preferred orientations in minerals, including isotropic ones like garnet, which cannot be measured optically.

Many different patterns of preferred orientation have been found. Plots of *c*-axes of quartz and calcite grains in stereographic projections form girdles or bands along which there may be one or more maxima. These patterns are classified according to the symmetry they exhibit – either orthorhombic, monoclinic, or triclinic. The pattern shown by the quartzite of Figure 17.7 is orthorhombic. The *c*-axes define two girdles at 90° to each other and at 45° to the foliation plane. The two girdles intersect to produce a maximum that coincides with the lineation of the stretched quartz pebbles.

Why mineral grains that show no morphological anisotropy and, in the case of quartz, little anisotropy of any physical properties should undergo reorientation during deformation and recrystallization is still not entirely understood. There is probably no one single cause, but the mechanisms by which rocks undergo plastic deformation by translation gliding, twin gliding, and recrystallization must play a role. Both gliding mechanisms are related to crystallographic directions, and for them to play roles in deformation they must be appropriately oriented with respect to the principal stress directions or be rotated into such orientations. Recrystallization in a stress field may also bring about preferred orientations, if certain orientations of nuclei effect a greater lowering of free energy than others.

The intracrystalline deformation process of *translation gliding* takes place in crystals along specific planes known as *slip planes*. These are normally characterized by a close-packed structure. The *slip direction* also tends to be one of either close packing or rows of similar ions or ions of similar charge. Gliding can be thought of as the simple shear of one part of the crystal past another (Fig. 17.10). Slip does not occur along the entire length of a slip plane at one time; this would result in a fracture and the grain would lose cohesion (brittle failure). Instead, a defect is generated in the structure (see beneath the ⊥ on the slip plane in Fig. 17.10), which migrates along the slip plane and so transmits the displacement through the crystal. The morphology of the crystal is modified by this gliding, but the crystallographic orientation remains unchanged. The effectiveness of translation gliding depends on the ease with which dislocations are formed and move through the structure.

Twin gliding can also be thought of as resulting from simple shear, with each successive layer of atoms above a twin plane gliding past the one below. In contrast to translation gliding, twin gliding results in rotation of the structure (Fig. 17.11). These twins, which are referred to as *deformation twins*, may obey the same laws as, and be indistinguishable from, growth twins or transformation twins. In some cases, however, a distinction can be made. Many deformation twins do not extend all the way through a crystal but instead form wedge-shaped polysynthetic lamellae that extend in from grain boundaries or are concentrated in parts of grains

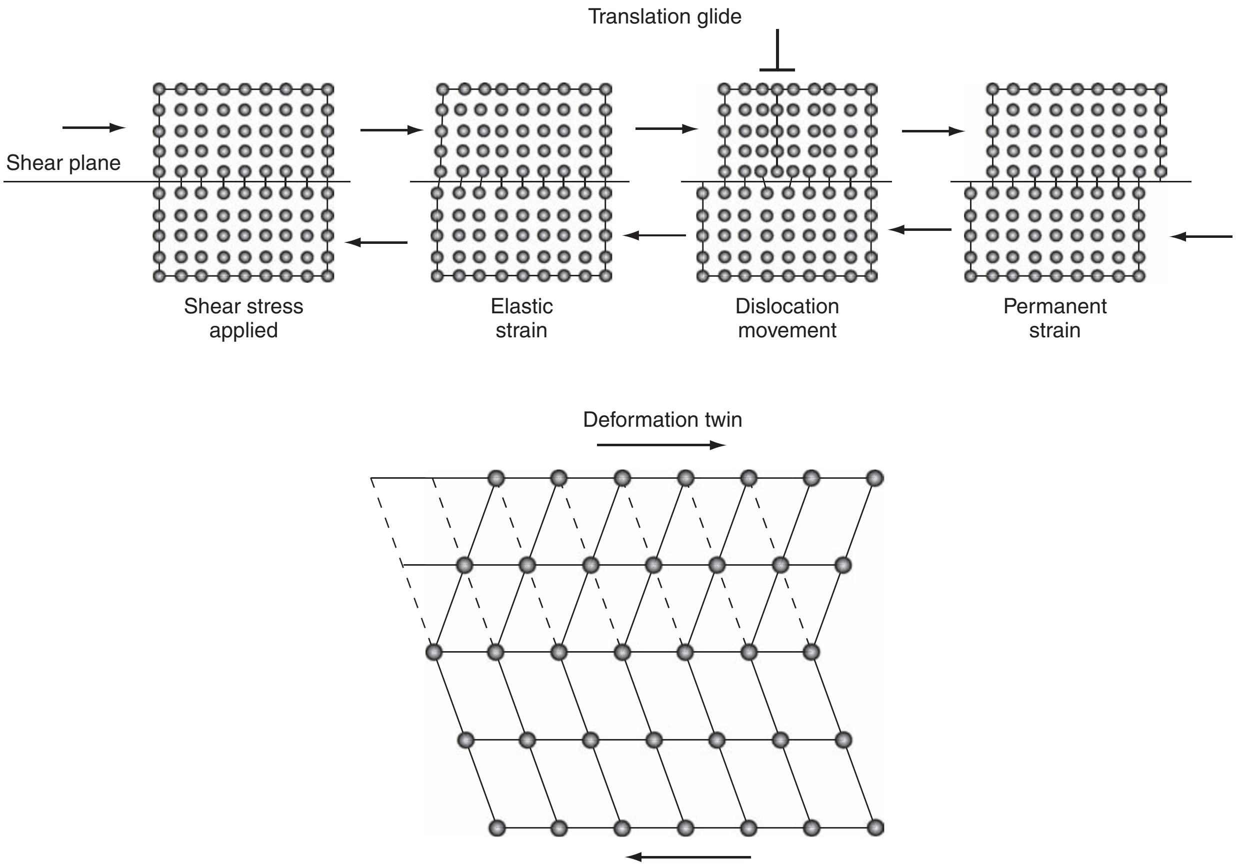


Fig. 17.10 Crystalline material deformed by translation gliding and twin gliding. In translation gliding a dislocation (see below ⊥ on slip plane) migrates along a slip plane through the structure. In twin gliding each successive layer of atoms above the twin plane glides past the one below. Twin gliding changes the crystallographic orientation, whereas translation gliding does not.

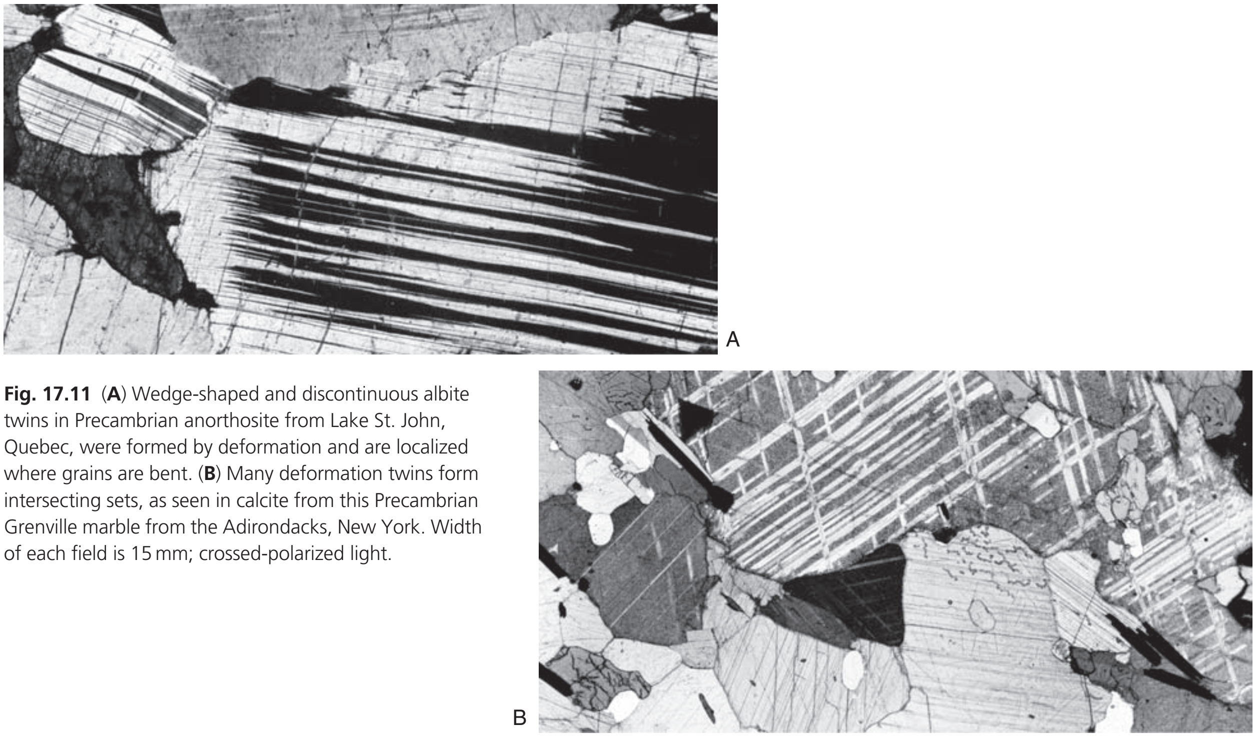


Fig. 17.11 (**A**) Wedge-shaped and discontinuous albite twins in Precambrian anorthosite from Lake St. John, Quebec, were formed by deformation and are localized where grains are bent. (**B**) Many deformation twins form intersecting sets, as seen in calcite from this Precambrian Grenville marble from the Adirondacks, New York. Width of each field is 15 mm; crossed-polarized light.

that are bent (Fig. 17.11(A)). Deformation twins also commonly form intersecting sets of lamellae (Fig. 17.11(B)). Deformation twins are common in calcite, dolomite, plagioclase, cordierite, and diopside.

In Chapter 12, we saw that deformation promotes recrystallization by introducing strain energy. We also saw that recrystallization attempts to minimize the free energy of a rock by reducing the surface area of grains and by eliminating boundaries that have high curvature; this results in grain coarsening (Section 12.6). The surface energy on grain boundaries also determines the dihedral angle that develops between grains during recrystallization. Clearly, a mineral grain in a metamorphic rock cannot be treated independently of its neighboring grains when considering size and shape because of the interfacial energies involved. Similarly, these energies may play a role in determining the orientation of juxtaposed grains. The surface energies of grain boundaries on anisotropic minerals vary with crystallographic orientation. Thus, two grains of equal size and shape may not have the same surface energy if their crystal structures are oriented differently with respect to the structure of surrounding grains. Also, if the grains are under stress, anisotropy of the elastic properties will result in differences in the free energy of the grains. This could also result in grains of certain orientation being favored during recrystallization over those of other orientations.

In summary, then, the foliation visible in hand specimen is only the outward expression of the almost total internal reconstitution of the fabric of a regionally metamorphosed rock. Although the alignment of platy and elongate minerals is the most pronounced aspect of this fabric, crystallographic alignment of all minerals may be strong. The fabric, whether measured in the outcrop or under the microscope, can be used to interpret regional structures because the plane of foliation normally corresponds to the plane of maximum extension, which during early stages of deformation probably is normal to the maximum compressive stress.

17.3 PORPHYROBLASTS

Next to foliation, the most characteristic feature of many metamorphic rocks is the presence of certain minerals whose grain size is significantly larger than that of the rest of the rock. These large grains are known as *porphyroblasts*, and the rock is said to have a *porphyroblastic texture*. The suffix *-blastic* indicates that the texture is of metamorphic origin. When used as a prefix, blasto- indicates that the texture is inherited from an earlier rock. For example, blastoporphyritic would apply to metamorphosed porphyritic igneous rock. Many of the metamorphic index minerals in pelitic rocks form porphyroblasts; for example, garnet, staurolite, cordierite, Al_2SiO_5 polymorphs, and chloritoid (cf. Fig. 16.4). The fact that these minerals form porphyroblasts aids in their recognition in the field.

The grain size of rocks depends largely on the number of nuclei formed, but factors such as the relative times of nucleation and the abundances of the various minerals must also play roles (Bell *et al.*, 1986). The tendency for large grains to grow at the expense of small ones might also be expected to contribute to the growth of porphyroblasts. However, this mechanism, while capable of eliminating small grains, does not lead to the development of very large ones. This is because the excess pressures that drive this process are produced by the curvature on grain boundaries, and these are significant only on small grains (Eq. (12.44)). As a result, porphyroblasts are not found in recrystallized monomineralic rocks, such as quartzite and marble. In multimineralic rocks, the number of nuclei formed of each mineral is unlikely to be the same, and thus variation in the grain size among the different minerals is to be expected.

The number of nuclei formed of a mineral depends on the distance over which diffusion is able to transport the nutrients and the nature and size of the critical nucleus (Section 12.2). How far material is transported depends very strongly on whether or not an intergranular fluid is present. Nucleation is likely to be heterogeneous, with some preexisting mineral providing the substrate on which the new one grows. Prediction of the actual site or sites of nucleation is difficult. The greater the critical radius of the nuclei, the less chance there is for the nuclei to form, and the fewer there will be. The critical radius is proportional to the interfacial energy on the crystal (Eq. (12.2)). Consequently, minerals that rank high in the *crystalloblastic series* (Table 12.1) are expected to have large critical radii and thus are likely to form porphyroblasts.

Some minerals high in the crystalloblastic series do not commonly form porphyroblasts, such as magnetite and rutile, but these have simple structures and compositions, which probably allows nucleation to occur more readily. Furthermore, many silicate minerals contain substantial Fe and Ti, so the amount of “free” Fe and Ti available to make oxides is often small and crystals cannot grow to large size. Some minerals low in the crystalloblastic series, such as cordierite and feldspar, also form porphyroblasts. These porphyroblasts tend to be ovoid in shape rather than being bounded by crystal faces (Fig. 16.4(F)). Perhaps the complexity of the structures of these minerals is what prevents large numbers of nuclei forming.

When porphyroblasts grow, they make room for themselves by either replacing material (involving transport and exchange of material from one place to another) or by physically pushing material aside. Although some porphyroblasts appear to have pushed apart foliation planes during growth, these can equally well be interpreted as foliation planes collapsing around the porphyroblast during or after growth. Indeed, when trails of inclusions are preserved in porphyroblasts that mark the position of the original foliation planes, the latter interpretation is normally found to be correct. Thus, during growth of the porphyroblast the spacing between foliation planes decreases, perhaps simply because of flattening of the rock or removal of quartz by solution, and as a result the foliation planes now wrap around the porphyroblast. In such cases the porphyroblast is commonly bordered by pressure shadows (Fig. 17.6). Of course, the growth of new minerals is accompanied by the disappearance of others, and this forms part of the transport and exchange, as discussed in Chapter 21.

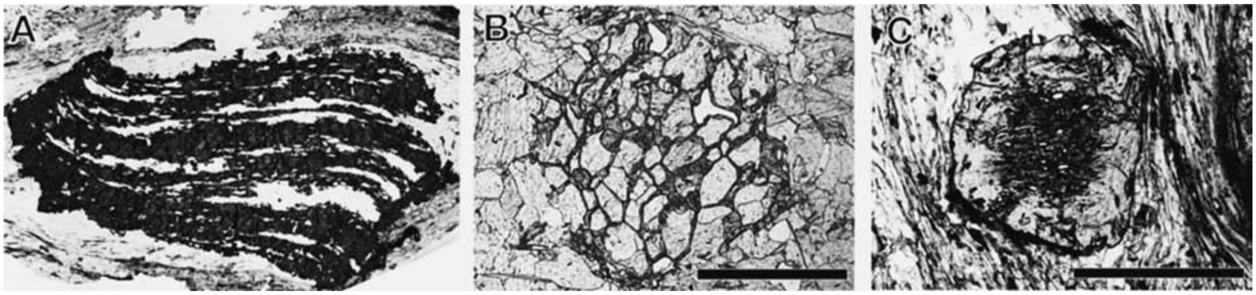


Fig. 17.12 Examples of inclusions within garnet crystals. (**A**) Large porphyroblast of garnet containing many layers of quartz, which are preserved from the metamorphic compositional layering of the rock that the garnet grew over. Width of field is 15 mm. (**B**) Extremely inclusion-rich "sieve" texture in garnet from blueschist–eclogite facies metacarbonate rock, Tinos, Greece. Inclusions are mostly calcite (transformed from aragonite) and quartz. Scale bar = 1 mm. (**C**) Garnet porphyroblast with inclusion-rich core and inclusion-poor rim. Core grew over an original foliation (horizontal in photo); continued deformation produced the observed foliation, which is at an angle of roughly 90° to relict foliation in core. Scale bar = 1 mm.

Many porphyroblasts incorporate inclusions of other minerals during their growth, giving rise to a *poikiloblastic* or *sieve* texture. In some porphyroblasts, inclusions may be concentrated into crystallographic planes, as in the variety of andalusite known as chiastolite (Fig. 16.4(C)). In most, however, the inclusions are distributed throughout the porphyroblast, occupying the same positions that they did prior to the growth of the porphyroblast (Fig. 16.4). Their distribution commonly reflects compositional differences inherited from sedimentary bedding or earlier foliation planes (Fig. 17.12(A)). The inclusions are commonly of one or a few minerals. Garnet and staurolite, for example, typically contain quartz inclusions; iron–titanium oxides, monazite, zircon, and graphitic matter are commonly included as well. Depending on the metamorphic history, inclusions may represent by-products of the reactions that produced the porphyroblasts, reactants that were present in excess, or "inert" minerals that were simply overgrown by the porphyroblasts. In addition to solids, fluid inclusions are also found in some porphyroblasts, and these can provide valuable information regarding metamorphic fluid compositions and flow histories (Touret, 1985; Whitney *et al*., 1996).

The presence of inclusions in porphyroblasts indicates that minimizing surface free energies during metamorphism is not always a high priority process. Some porphyroblasts are composed almost entirely of inclusions, with the host mineral forming only a thin crystallographically continuous intergranular network (Fig. 17.12(B)). The fact that a porphyroblast can grow as a thin film along grain boundaries can be interpreted as indicating that the dihedral angle between the porphyroblast and the other minerals was essentially zero at the time of growth; that is, the porphyroblast wetted the other minerals (see Section 12.7).

The abundance of inclusions can vary from core to rim in a porphyroblast. In such cases, cores are typically inclusion rich and rims are inclusion poor (Fig. 17.12(C)); the opposite situation is also found but is less common. The reasons for these changes in inclusion abundance are still not fully understood. One possibility is changes in fluid composition. We have seen from experiments (Watson and Brenan, 1987) that the composition of an intergranular fluid can significantly affect the dihedral angle between minerals. It is not unreasonable to suppose that during the growth of a porphyroblast the composition of a fluid phase may have changed, or the fluid phase may have disappeared altogether. Another possibility is that the mineralogical composition of the rock changed as the porphyroblasts grew. For example, the inclusion mineral may be consumed by metamorphic reactions, so less and less of it would be available to be included during porphyroblast growth, ultimately yielding inclusion-poor rims.

Variable growth rates may also be a factor in producing poikiloblastic crystals with inclusion-poor rims. If growth was originally rapid and then slowed toward the end, diffusion may have had insufficient time at first to remove the inclusions. Furthermore, if temperature increases during growth, as would normally be the case, then diffusion would be faster (Chapter 5) and more effective at higher temperatures when rims were crystallizing, even if growth rates didn't vary much. Although the conventional wisdom is that rapid porphyroblast growth results in the incorporation of more inclusions than slower growth, more research encompassing a broad range of metamorphic conditions is needed to test this concept further.

17.4 INTERPRETATION OF PORPHYROBLAST–INCLUSION RELATIONS

Trails of inclusions within porphyroblasts provide valuable information regarding the relative timing of metamorphic and tectonic events, as well as processes of mineral growth and deformation. The relative timing of mineral growth is usually referenced to a deformational structure, commonly a foliation, or a group of related structures. A variety of terminology conventions can be found in the literature; ours is based on Passchier and Trouw (1996).

Pretectonic crystals grow before regional deformation (Fig. 17.13). Consider andalusite or cordierite crystals produced during contact metamorphism in the upper crust. They may contain inclusions that are crystallographically controlled (e.g. chiastolite variety of andalusite; Fig. 16.4(C)). Inclusions may also follow existing bedding planes that the crystals grew over. Alternatively, in layers with no discernable bedding or other layering, inclusions will be randomly oriented (Fig. 16.4(F)). If the rocks were then buried and subjected to regional metamorphism and deformation, foliations would develop around the porphyroblasts. If the porphyroblasts are relatively rigid, as is the case for minerals like garnet and andalusite, then the foliation will bend and "collapse" around the crystals during flattening (Fig. 17.14(A)). Crystals can even undergo extension and boudinage in the plane of foliation (Fig. 17.14(B)). The key point is that the inclusion patterns preserved in the porphyroblasts are unrelated to deformation or the deformational fabrics in the surrounding matrix. Commonly, the pretectonic minerals are not thermodynamically stable at regional metamorphic pressure–temperature–fluid composition conditions, so they can react and be pseudomorphed to varying degrees by new minerals (e.g. Evans and Davidson, 1999).

Syntectonic crystals grow while regional deformation proceeds. Figure 17.15 illustrates garnet porphyroblasts containing quartz-rich inclusion trails. The most probable interpretation is that the garnets rotated with respect to the foliation as they grew during deformation. The inclusions in the center of the garnets parallel an early foliation, which must therefore have existed prior to the growth of the garnet. As the garnets grew, the schistosity became shear planes that caused the garnets to rotate relative to the schistosity in a counterclockwise direction (Fig. 17.15). The layers of quartz in the surrounding schist continued to be included in the garnet, but once included they were rotated by the garnet. By the time the garnet finished growing, the original foliation planes in the core of the porphyroblasts had been rotated through 180°. Other garnets in this same rock unit from southeastern Vermont exhibit rotations exceeding 600° (Rosenfeld, 1968).

Inclusion patterns such as those of Figure 17.15 in which the relative rotation angle exceeds 180° are commonly referred to as *snowball structures*. Interestingly, well-developed "snowballs" appear to be restricted to garnet porphyroblasts. The layers within these porphyroblasts are actually very different from those in a snowball. Layering in a snowball coils inward to the nucleus, whereas that in a porphyroblast can be traced in from one side, through the core, and back out the other side (Rosenfeld, 1970). The pattern is similar to that formed when a strand of spaghetti is wound from its middle onto a fork. Indeed, this analogy may be useful if difficulty is experienced in working out the sense of shear on a porphyroblast; simply picture rotating a fork so as to coil spaghetti in the same sense as the layering in the porphyroblast.

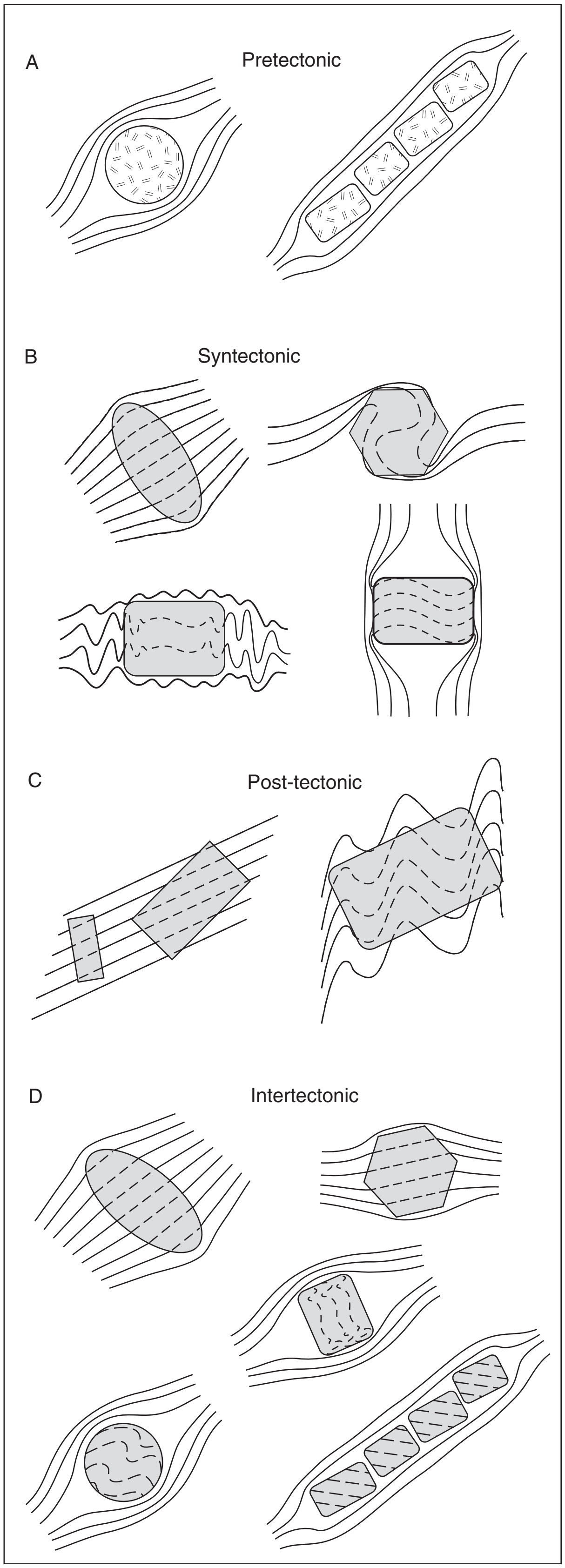


Fig. 17.13 Examples of matrix–porphyroblast relations. Matrix foliations denoted by solid lines, inclusion trails in porphyroblasts denoted by dashed lines. (**A**) Pretectonic porphyroblasts which grew over randomly oriented grains in matrix prior to development of foliation. Grain at right has been broken into four pieces and extended. (**B**) Syntectonic crystals that grew during deformation. (**C**) Post-tectonic porphyroblasts that grew over existing foliations after deformation. (**D**) Intertectonic porphyroblasts that grew during or after an earlier deformation event (or events) predating the fabric now evident in the matrix. See text for further discussion. (Modified after Spry, 1969; Passchier and Trouw, 1996.)

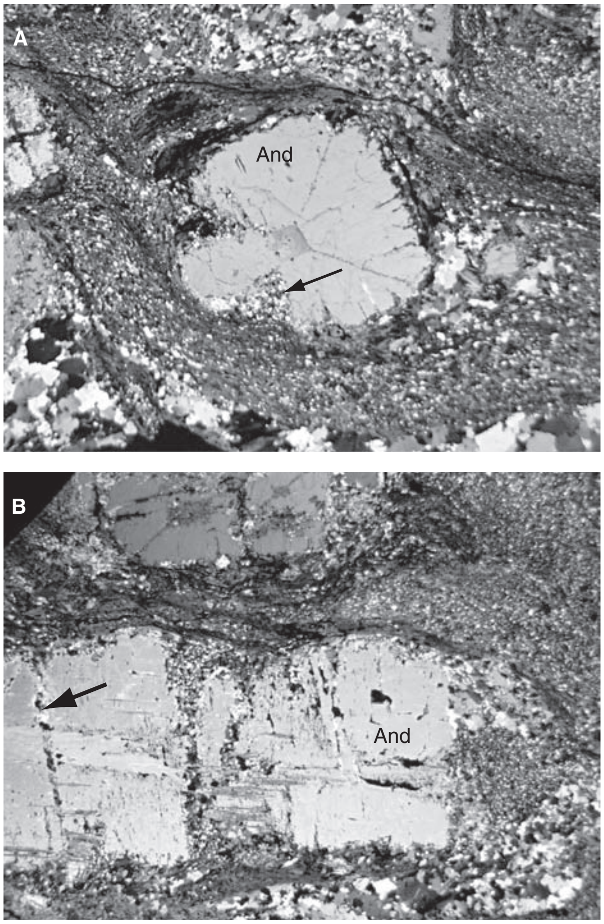


Fig. 17.14 Photomicrographs of pretectonic andalusite (And) porphyroblasts that developed during contact metamorphism of the Chiwaukum Schist adjacent to the Mount Stuart Batholith, Washington State (crossed polarizers). (**A**) Andalusite (chiastolite) contains crystallographically controlled inclusions (Chapter 12), but contains no planar inclusion trails that would indicate growth over a matrix foliation. Matrix foliation developed during later regional metamorphism, and wraps around the porphyroblast. Note that a small area of relict, undeformed matrix is preserved in the triangularly shaped region between two of the "arms" of the andalusite crystal (arrow). Field of view is 2 cm. (**B**) Pretectonic andalusite that is being fractured (e.g. arrow) and extended parallel to the regional metamorphic foliation that wraps around the crystal. Field of view is 2.5 cm.

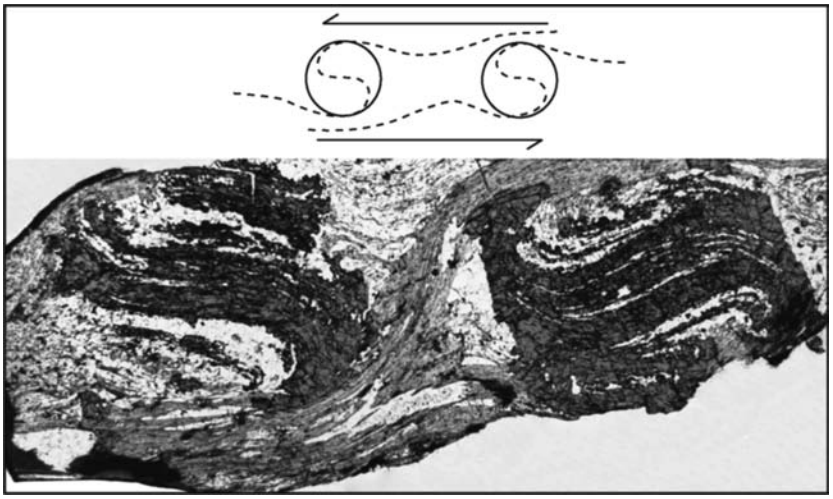
Snowball structures are not the only indicators of syntectonic growth. In fact, although spectacular, snowball structures are relatively uncommon. More typical are porphyroblasts whose curved inclusion trails indicate much less than 180° of apparent rotation. These crystals are also syntectonic, but one needs to be very careful about the structural interpretation because they can form in several different deformational regimes, not just the shear and rotation model envisioned for Fig. 17.15 (Passchier and Trouw, 1996). Additional syntectonic examples include rigid porphyroblasts that grow and record the collapse of foliation around them during flattening, and porphyroblasts that grow over developing folds (Fig. 17.13). Fold patterns preserved as inclusion trails within porphyroblasts are known as *helicitic folds* (Fig. 17.13).

Post-tectonic crystals grow after deformation. Common examples are porphyroblasts that grow over existing deformational features like foliations or folds (Figs. 17.13, 17.16). Inclusion patterns in the porphyroblasts are continuous with, and have the same orientations as, fabrics in the surrounding rock matrix. Furthermore, foliations are not deflected around the porphyroblasts.

Some porphyroblast inclusion textures do not fit neatly into the above three classifications. For example, syntectonic cores are often mantled by post-tectonic rims. The porphyroblast shown in Figure 17.17 is another example. Here, planar layers of quartz inclusions can be traced through the crystal; they almost certainly represent growth over a differentiated cleavage fabric that consisted of alternating sheet silicate-rich and quartzofeldspathic bands. The differentiated fabric in the surrounding rock, however, makes an angle of roughly 45° with the planes of inclusions preserved in the porphyroblast. In addition, the matrix foliation is deflected around the porphyroblast.

This one thin section records a considerable amount of history. First, a metamorphic foliation developed prior to the appearance of garnet. This was probably an axial plane schistosity that developed normal to the maximum compressive stress. The metamorphic grade increased to the point that garnet formed and overgrew this existing foliation; the planar

Fig. 17.15 Photomicrograph illustrating syntectonic "snowball" garnet. Layers of quartz in garnet porphyroblasts from southeastern Vermont (Rosenfeld, 1968) have inclusions indicating a counterclockwise rotation of 180°. Because garnets were growing during rotation, layers become folded during inclusion. Width of field is 30 mm.



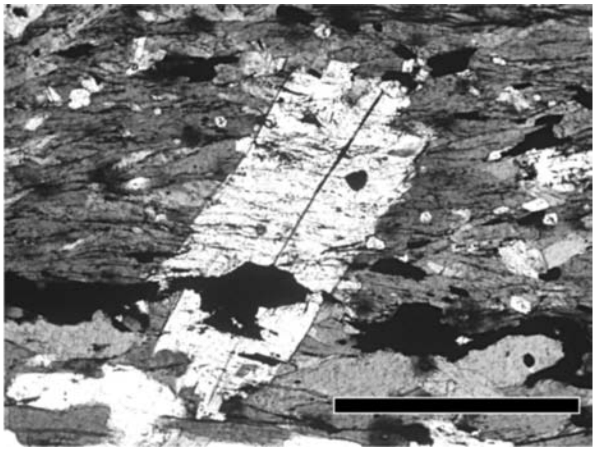


Fig. 17.16 Post-tectonic chlorite crystal that grew over existing biotite-rich foliation, Chiwaukum Schist, Washington State. Scale bar = 1 mm.

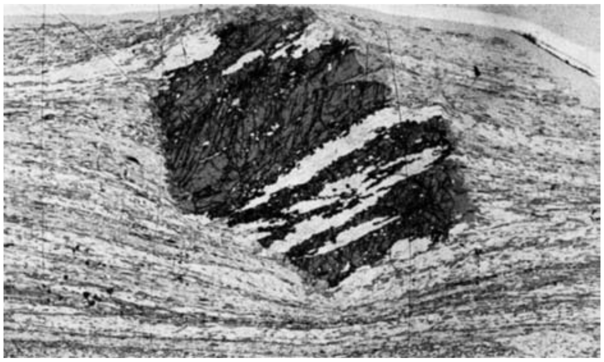


Fig. 17.17 Intertectonic garnet, southeastern Vermont (Rosenfeld, 1968). See text for discussion. Width of field is 25 mm.

inclusion trails indicate that the foliation was not being deformed as the garnet grew. But the history after the garnet ceased growing becomes a little trickier to deal with. In recognition of these difficulties, Passchier and Trouw (1996) introduced the term *intertectonic* for crystals that grew over a deformational foliation that are now contained in a rock matrix that has a different foliation orientation which is not recorded within the porphyroblast (Figs. 17.13, 17.17).

A likely intertectonic scenario is that after the garnet finished growing there was a hiatus in tectonic activity, followed by a new deformational event. This event produced the observed difference in orientation between the garnet inclusion trails and the matrix foliation, as well as the bending of the foliation around the porphyroblast. One possibility is that the garnet was rotated by a component of sinistral shear operating parallel to the foliation. Another possibility is that the stress field for the later event was oriented at an angle of ~45º from the original foliation. In response to this stress, a new foliation developed and was bent around the porphyroblast, without significant shear along foliation planes. Determining the correct intertectonic scenario would require detailed structural analysis (Passchier and Trouw, 1996). However, in either case, the temperature–pressure–fluid composition conditions for the later event were probably not much different than when the garnet grew, because the garnet is not being replaced by new prograde or retrograde minerals. The thin section textures do not tell us the duration of a hiatus, but one could estimate the timing using, for example, Sm/Nd dating of garnet together with U/Pb dating of monazite that grew or recrystallized in the matrix (see Chapter 13).

Treatment of natural metamorphic and deformational phenomena ultimately requires knowledge of the rates of porphyroblast growth. These rates are extremely important to determine because they can quantify the timescales of tectono-metamorphic processes, but they are difficult to assess. One approach focuses on the diffusion rates of constituents needed for porphyroblast growth. For example, most porphyroblasts contain aluminum. Taking Carmichael's (1969) estimate of 0.3 mm for the limit of aluminum mobility in the rocks of the Whetstone Lake area of southeastern Ontario together with a diffusion coefficient of $10^{-19}\,\mathrm{m}^2\,\mathrm{s}^{-1}$ (see Chapters 5, 21), aluminous porphyroblasts would require of the order of 10^4 years to grow ($x \approx 2\sqrt{Dt}$; Section 5.6). This is a relatively short geological timescale estimate; factors such as the presence or absence of a fluid-filled, interconnected porosity, as well as longer or shorter length scales of aluminum transport, could decrease or increase such estimates considerably.

Another approach is age dating. Christensen *et al*. (1989) obtained a range of Rb/Sr absolute ages on garnet porphyroblasts from the Ottauquechee Formation of southeast Vermont, which was metamorphosed during the Devonian Acadian orogeny. These porphyroblasts, which are up to 3 cm in diameter and are poikiloblastic, have an age span of 9 Ma from their core to rim, indicating a very slow growth rate. On the other hand, Baxter *et al*. (2002), using Sm/Nd dating of multiple garnet growth generations, found that the "peak" (maximum temperature conditions) of Barrovian metamorphism in northeastern Scotland lasted only a few million years, or even less (Chapter 22). It is likely that rates of growth vary widely depending on factors including the rates of burial and heating and the presence and composition of fluids; mineral growth kinetics will be an important area of active research for many years to come.

The above discussion highlights many important relationships between mineral growth, deformation, and patterns of inclusion trails within porphyroblasts. A simple model of syntectonic garnet growth that can produce snowball structures is very useful for exploring some of these relationships (Fig. 17.18). We first assume that the growth of garnet is diffusion controlled (Eq. (12.7)); that is, the radius of a porphyroblast at any time t is given by $r^2 = Kt$, where K is a growth constant. We will also assume that the rate of shearing, and hence the rate of rotation of the porphyroblast, is constant. If the angle between the foliation plane trapped in the center of the crystal and that in the surrounding rock is ϕ (Fig. 17.18), then a constant rotation rate gives $d\phi/dt$ = constant, which on integration gives $\phi = Ct$, where C is a time constant. We will further assume that a marker layer initially passes through the center of the crystal and at all times remains parallel to the initial foliation. As the crystal rotates,

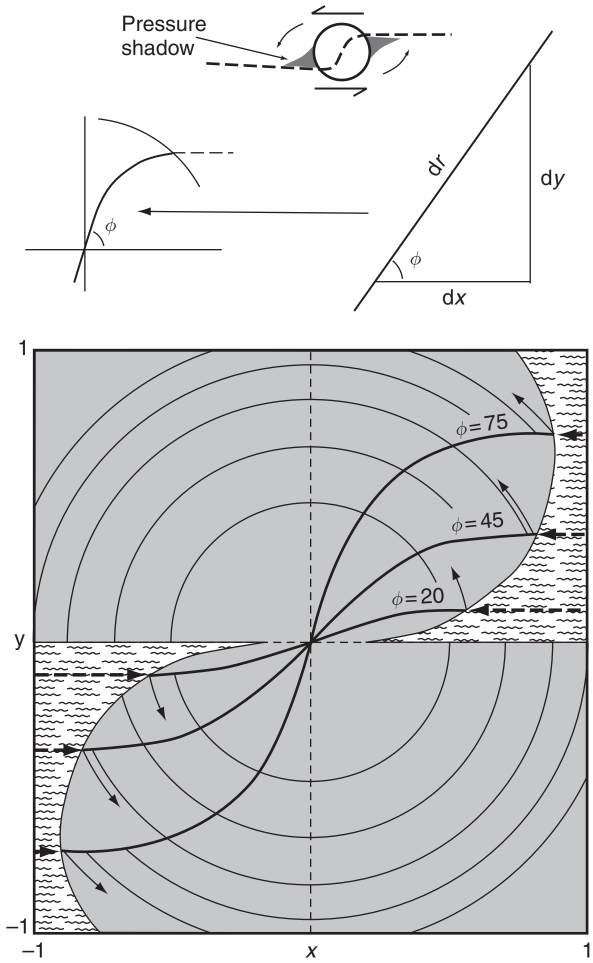


Fig. 17.18 Model for the incorporation of an inclusion layer in a constantly rotating spherical porphyroblast whose growth is diffusion-controlled. The included layer at the center of the crystal rotates through an angle ϕ. The slope at any point on the layer in the crystal is given by the angle θ. The graph shows the shape of the layer at three stages of growth, which are given in terms of the angle ϕ. See text for discussion.

the layer in the surrounding rock must be moved by the rotating garnet. The space created by this movement would be filled by quartz (pressure shadow) or by schist flowing in from the other side of the garnet (Fig. 17.18).

According to this model, the marker layer is always included with a zero degree slope at the time of inclusion in the garnet. Once included, the angle through which the layer is rotated is a measure of the length of time it has been included. The total growth time of the porphyroblast is given by the angle of the foliation at the center of the crystal ($t=\phi/C$). The growth time to any radius r within the porphyroblast is $t_r=(\phi-\theta)/C$, and because the growth rate is diffusion-controlled ($r^2=Kt$), we can write

$$r^2=\frac{(\phi-\theta)K}{C} \tag{17.1}$$

Because K and C are constants, and we are interested in only the shape of the deformed marker layer, not its absolute dimensions, we will set the constants equal to unity. Equation (17.1) relates the slope of the deformed layer (θ) to its distance from the center of the porphyroblast (r), which has undergone a total rotation of ϕ at its center.

To plot a graph of the layer in terms of x and y coordinates, we make use of the fact that a small increment in radius, dr, can be related to x and y (Fig. 17.18) by

$$(\mathrm{d}x)^2+(\mathrm{d}y)^2=(\mathrm{d}r)^2 \tag{17.2}$$

But $\mathrm{d}y=\mathrm{d}r\sin\theta$, which, when substituted in Eq. (17.2), gives

$$(\mathrm{d}x)^2=(1-\sin^2\theta)(\mathrm{d}r)^2=\cos^2\theta(\mathrm{d}r)^2 \tag{17.3}$$

From Eq. (17.1), $(\mathrm{d}r)^2=-\mathrm{d}\theta$, so

$$(\mathrm{d}x)^2=-\cos^2\theta\,\mathrm{d}\theta \tag{17.4}$$

which, upon integration between the limits of $\theta=\phi$ (center of crystal) and $\theta=0$ (rim of crystal), gives

$$x^2=-\frac{\theta}{2}-\frac{1}{4}\sin 2\theta+\frac{\phi}{2}+\frac{1}{4}\sin 2\phi \tag{17.5}$$

or

$$x=\left(-\frac{\theta}{2}-\frac{1}{4}\sin 2\theta+\frac{\phi}{2}+\frac{1}{4}\sin 2\phi\right)^{1/2} \tag{17.6}$$

The y coordinate is similarly obtained by substituting $\mathrm{d}x=\mathrm{d}r\cos\theta$ into Eq. (17.2), which gives

$$y=\left(-\frac{\theta}{2}+\frac{1}{4}\sin 2\theta+\frac{\phi}{2}-\frac{1}{4}\sin 2\phi\right)^{1/2} \tag{17.7}$$

Graphs of the deformed marker layer are shown in Figure 17.18 for three different angles of rotation that correspond to different ages in the growth of the crystal. The marker layer in the surrounding schist is shown by heavy dashed lines with arrows indicating the points of inclusion in the porphyroblast at the three stages of growth. Because the model crystal growth is diffusion controlled, the radius of the porphyroblast increases most rapidly at first, and thus there is least time for rotation at this stage. The marker layer is therefore relatively straight in the core of the crystal. With time, the growth rate decreases, and thus there is progressively more rotation for any given increment in radius, causing the marker layer to become more bent toward the rim. The pattern seen in the porphyroblasts of Figure 17.15 could therefore be consistent with this particular growth model. If crystal growth were controlled by a different mechanism, the pattern formed by the inclusions would be different (Problem 17.4).

17.5 METAMORPHIC TEXTURES IN SHEAR ZONES

Despite the general tendency for grain size to increase during prograde metamorphism, a number of important rocks are formed by a reduction in grain size. These rocks typically have a strong foliation and may be compositionally layered. They occur most commonly in zones of intense shear deformation and may be associated with faults, in particular thrust faults. There is little doubt that their fine grain size is related to the deformation. A clear and well-illustrated treatment of these rocks is given by Higgins (1971).

Many of the textural terms used to describe these rocks reflect early interpretations in which the fine grain size was thought to result from grain breakage. As a group, they are referred to as being *cataclastic*, a term defined by the American Geological Institute Glossary of Geology as "pertaining to a texture found in metamorphic rocks in which brittle minerals have been broken and flattened in a direction at a right angle to the pressure stress." It is now clear, however, that much of the reduction in grain size results from plastic deformation and recrystallization during which the rock does not lose cohesion.

The transition from brittle to ductile behavior in rocks depends on confining pressure, temperature, strain rate, fluid phase, and mineralogical composition (Suppe, 1985). In continental rocks, the transition is controlled largely by the presence of quartz, which is one of the first of the common minerals to become ductile with increasing depth (Tullis and Yund, 1987). Because the brittle–ductile transition occurs at different temperatures with different minerals, a polymineralic rock may exhibit both forms of deformation. At the elevated temperatures and pressures of middle and lower crustal metamorphism, deformation is predominantly ductile for small to moderate strain rates ($<\sim 10^{-14}$ s^{-1}). However, in the lowest-grade rocks, and even in high-grade rocks exposed to high strain rates, extensive brittle deformation can occur. An important consequence of such deformation is vein formation, a topic we will return to in Chapter 21.

The grain size of rocks undergoing strong deformation is reduced regardless of whether the deformation is brittle or ductile. With brittle behavior, grains are fractured. In shallow fault zones, *breccias* develop, and if there is sufficient movement and comminution, rock can be reduced to an extremely fine powder known as *fault gouge*. With plastic deformation, grain size is reduced by the generation and migration of defects and by recrystallization. In Chapter 12 we saw that the number of sites for nucleation and growth of new unstrained grains increases as the amount of strain in a rock increases. Consequently, rocks that have undergone high degrees of strain become fine grained if they recrystallize. During progressive metamorphism, rising temperatures promote recrystallization. But deformation itself causes heating and in extreme cases results in melting (see the discussion of pseudotachylite below). Recrystallization is therefore likely to occur in most zones of intense deformation.

In a polymineralic rock, some minerals become strained more easily than others, and these are likely to recrystallize more rapidly. Growth rates during recrystallization also differ from mineral to mineral, because recrystallization depends on the migration of grain boundaries, which, in turn, is determined by diffusion rates. In the brittle–ductile transition zone, ductile minerals may deform and recrystallize around the more brittle ones. Consequently, the various minerals in a rock have their grain sizes reduced to different extents by deformation. Regardless of the mode of deformation, the larger residual grains are referred to as *porphyroclasts*.

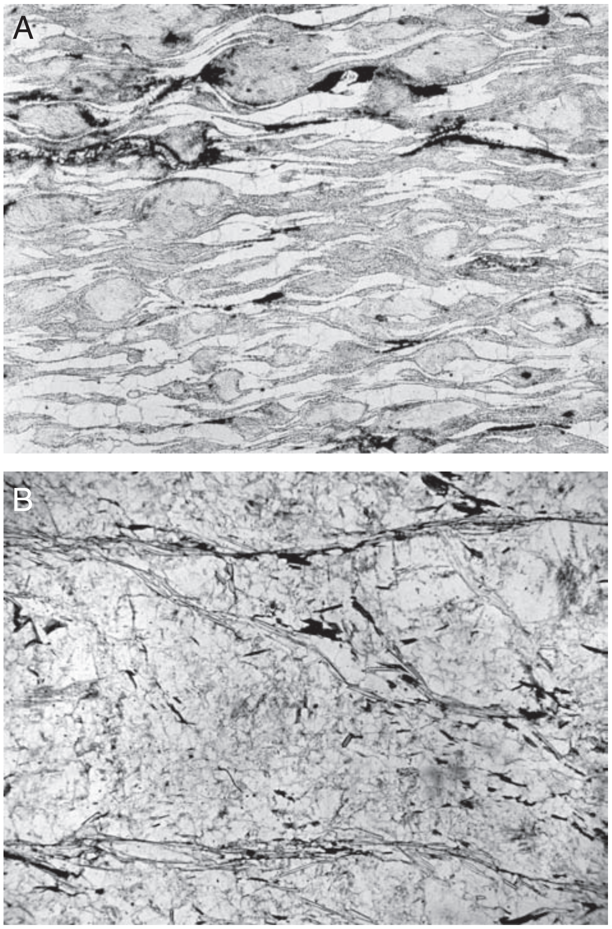


Fig. 17.19 (**A**) Strongly deformed augen (flaser) granite gneiss, Grenville Township, Quebec. Porphyroclasts of deformed perthitic feldspar are surrounded by extremely fine-grained feldspar and lenses of recrystallized quartz. Width of field is 3 mm; plane-polarized light. (**B**) C-type shear band cleavage in deformed granite, south-central Connecticut. Shear bands (horizontal in photo) cut earlier foliation that trends from the upper left to the lower right. Sinistral shear sense (Fig. 17.21). Section cut normal to the foliation and parallel to the stretching lineation. Width of field is 5.7 mm; plane-polarized light.

In many medium- to high-grade metamorphic quartzofeldspathic rocks, quartz grains deform rapidly into flattened sheets or lenses (Fig. 17.19(A)), whereas feldspar grains may retain their original shape or show only small amounts of strain (Simpson, 1985). The quartz in the sheets recrystallizes to unstrained grains. Such rocks are known as *flaser gneiss*. With increased deformation, feldspar also undergoes recrystallization. This starts with the rims of grains recrystallizing to very fine aggregates of unstrained grains that commonly have a strong preferred crystallographic orientation (Fig. 17.19(A)). Commonly, the cores remain as large strained porphyroclasts within aggregates of fine-grained feldspar and quartz sheets. Because of the eye-like shape of relict feldspar grains, this rock is referred to as *augen gneiss* after the German word for an eye (*auge*). The term *phyllonite* is used for those rocks in which the grain size has been reduced to that which typifies a phyllite.

In zones of intense shearing, as for example near thrust faults, the original grains of a rock can be stretched into long thin streaks. With such large amounts of strain, recrystallization takes place easily, and because of the large number of nucleation sites, the resulting rock is extremely fine grained and usually dark colored. Because the reduction in grain size was originally thought to have formed by the grinding or milling of material in shear zones, these rocks were named *mylonites* from the Greek word for a mill (*mylon*). In addition to being fine grained, they are characterized by what is referred to as a *fluxion* or flow structure, a penetrative foliation which may or may not be accompanied by compositional layering and which is commonly visible only under the microscope.

Mylonites can be classified as protomylonite, mylonite, and ultramylonite, depending on the degree to which grain size is reduced (Higgins, 1971). *Protomylonite* contains more than 50% megascopically visible porphyroclasts. This rock may resemble a metaconglomerate if the porphyroclasts are large enough. In *mylonite*, porphyroclasts constitute from 10% to 50% of the rock and are generally larger than 0.2 mm. *Ultramylonite* contains less than 10% porphyroclasts, which are smaller than 0.2 mm, and most are reduced to fine-grained streaks. The rock generally becomes darker as grain size decreases.

Figure 17.20 illustrates this progression in rocks from the Honey Hill thrust zone of southeastern Connecticut. The protolith is a coarse-grained, massive gabbro. Toward the fault zone its pyroxene is converted to amphibole, which recrystallizes more rapidly than the plagioclase. The resulting protomylonite contains porphyroclasts and lenses of plagioclase in fine-grained amphibole (Fig. 17.20(A)). Nearer the fault, the size of the porphyroclasts decreases, and the foliation of the mylonite is folded (Fig. 17.20(B)). In the fault zone itself, the rock is reduced to an ultramylonite, in which almost no porphyroclasts remain, but a strong compositional layering of amphibole and plagioclase attests to their former existence. Extremely flattened folds are evident in this layering (upper left of Fig. 17.20(C)). The rock is so fine grained (examine light-colored layers in Fig. 17.20(C)) that it resembles flint in hand specimen. The term *flinty-crush-rock* has, in fact, been used for these rocks in the Scottish Highlands.

Although some of the layering seen in ultramylonites is inherited from compositional differences in the protolith, completely homogeneous rocks can become layered during deformation. Apparently, differences in mechanical properties of minerals can result in cataclastic metamorphic differentiation. The layering and fine grain size of ultramylonites have led to their misidentification as volcanic rocks and siliceous sedimentary rocks. In hand specimen, the distinction may be difficult to make.

The textures present in shear zones often indicate the sense of shear and are, therefore, important for structural analysis (Fig. 17.21). The deformation and offset of preexisting lithologic layering or foliations can reveal the sense of shear in outcrop. In some cases, thin section examination is necessary. Fortunately, a large number of shear sense

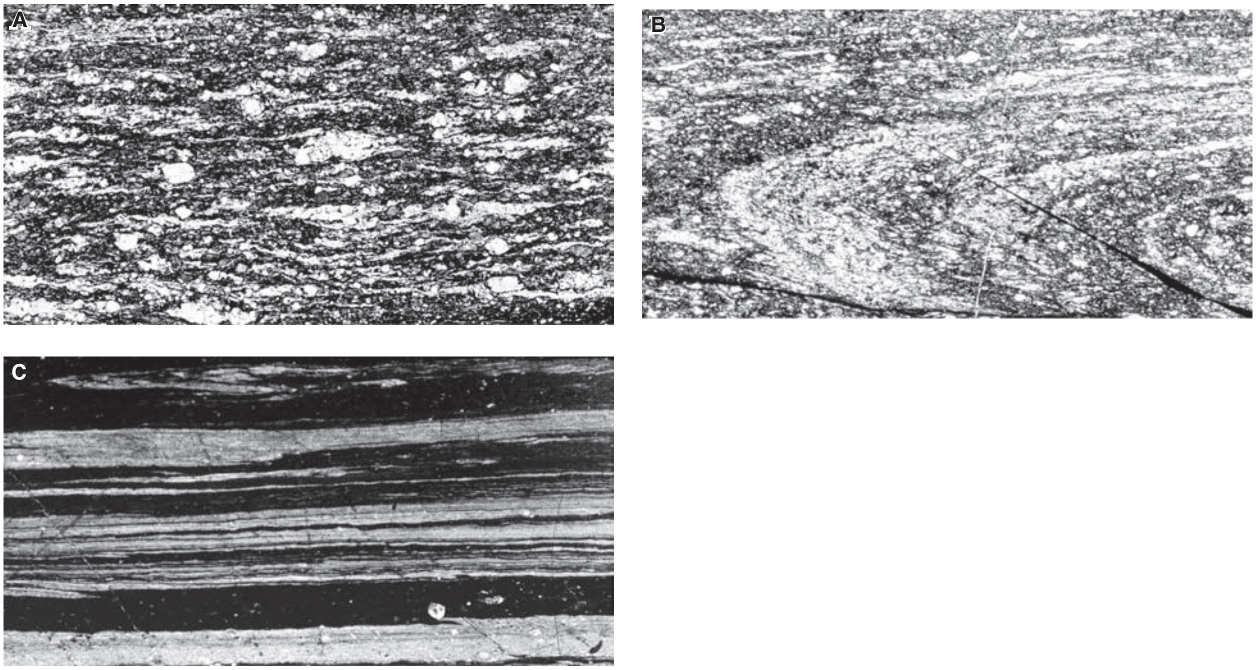


Fig. 17.20 Stages in the development of mylonite from a coarse-grained gabbro in the Honey Hill thrust fault of southeastern Connecticut. (**A**) *Protomylonite*: on approaching the fault, pyroxene is converted to amphibole (dark), which recrystallizes rapidly to fine grains. Plagioclase resists deformation and survives as porphyroclasts surrounded by lenses of fine-grained recrystallized plagioclase, which is streaked out to form a fluxion structure. (**B**) *Mylonite*: in this section plagioclase porphyroclasts are greatly reduced in size, and fluxion layering is folded. (**C**) *Ultramylonite*: in the thrust fault itself the grain size is extremely fine, and porphyroclasts are visible only under the microscope. Layering is much more tightly folded than in part (B) indicating the higher degree of strain. Width of each field is 16 mm; plane-polarized light.

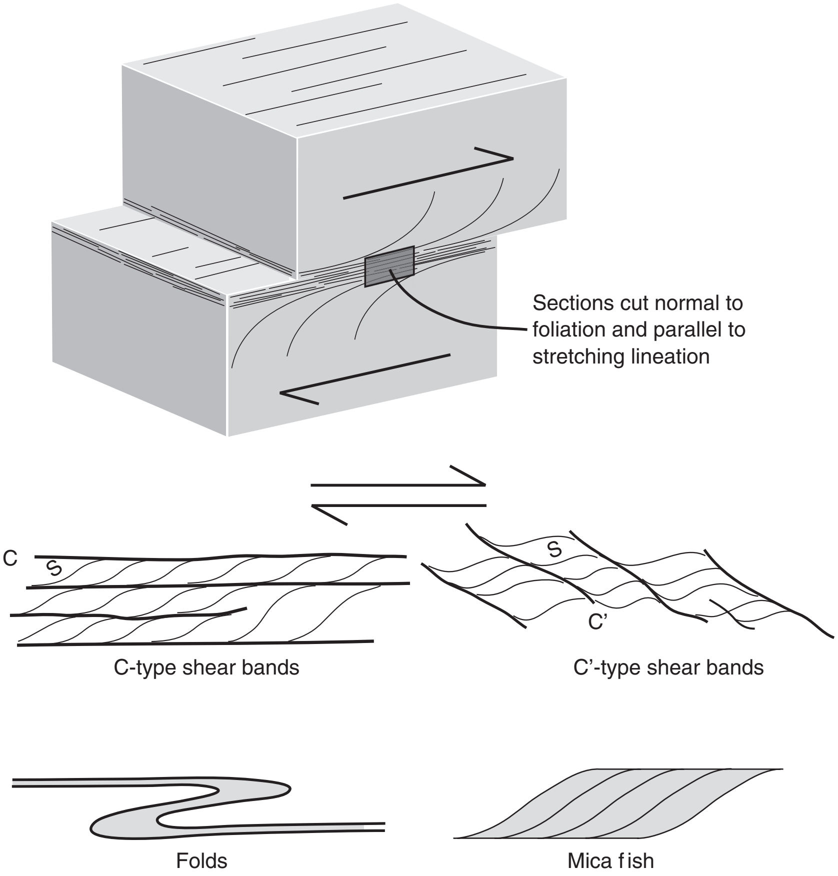


Fig. 17.21 Examples of shear sense indicators for sections cut normal to the foliation and parallel to the stretching lineation in shear zones. The S foliation planes (from French *schistosité*) are transected by C or C′ shear bands (from the French *cisaillement*, denoting shear). Dextral sense of shear for all examples. (Modified after Passchier and Trouw, 1996; see this text and references therein for detailed examples and discussion.)

indicators have been discovered; the terminology and examples used here follow Passchier and Trouw (1996). In order to evaluate shear sense, thin sections should be oriented and cut perpendicular to foliation and parallel to the stretching lineation (this lineation is parallel to the shear direction).

Within the shear zone, foliation or other layering is commonly cut by a series of small, subparallel shear zones or bands (Problem 17.8). Such *shear band cleavages* fall into two broad varieties. *C′-type* shear band cleavages have shear bands that are oblique to the margins of the shear zone (15–35º) and to preexisting cleavage planes (Fig. 17.21). In contrast, *C-type* shear band cleavages (also called C/S, S-C, or C-S fabric) are characterized by cleavage planes oblique to the shear zone margins that are transected by shear bands that are subparallel to the margins (Figs. 17.19(A), 17.21). The shear bands tend to be straighter and more continuous than those found in the C′-type. The development of shear band cleavages is not fully understood, but is dependent on rock type. C′-type shear band cleavages tend to develop in mica-rich mylonites or phyllonites, whereas the C-type is usually found in less micaceous lithologies, such as deformed granitic rocks. There are many other shear sense indicators besides shear band cleavages, such as strongly asymmetric mica "fish" (Fig. 17.21). See Passchier and Trouw (1996) for a comprehensive review.

The temperature of rocks can be raised by deformation, but the effect is usually small. If the maximum deviatoric stress that a metamorphic rock can withstand is 0.03 GPa, and the rock is strained at a representative rate of 10^{-14} s^{-1}, the rate of working is 3×10^{-7} Pa s^{-1}. But 1 Pa = 1 J m^{-3}. So the rate of working is therefore 3×10^{-7} J m^{-3} s^{-1} or 0.3 μW m^{-3}. If the rock has a heat capacity of 0.9 kJ kg^{-1} K^{-1} and a density of 2.8 Mg m^{-3}, its heat content would be 2.5 MJ K^{-1} m^{-3}. If the deformation lasted for 1 Ma (3×10^{13} s), and the heat was completely retained by the rock, the temperature rise would be

$$\Delta T = \frac{3 \times 10^{-7}\left(\mathrm{J\ m^{-3}\ s^{-1}}\right) \times 3 \times 10^{13}(\mathrm{s})}{2.5 \times 10^{6}\left(\mathrm{J\ K^{-1}\ m^{-3}}\right)} = 3.6\ \mathrm{K}$$

On a regional scale, therefore, deformation at this strain rate is not likely to cause significant heating. Larger bulk strain rates of 10^{-13} s^{-1} or more are possible, however, and these could produce greater heating effects. Furthermore, if deformation is concentrated into narrow zones, deviatoric stresses in excess of 0.1 GPa may cause appreciable heating (Graham and England, 1976). Indeed, in extreme cases temperatures may be raised to the melting point of rock and the liquid formed injected into fractures where it is quenched rapidly to a dark aphanitic or glassy rock known as *pseudotachylite* because of its resemblance to volcanic glass (tachylite).

Pseudotachylite forms narrow dikelets that are rarely more than 1 cm wide (Fig. 17.22). In the field, they resemble rapidly quenched diabase. In thin section, however, their mylonitic origin is clear. They contain many small fragments

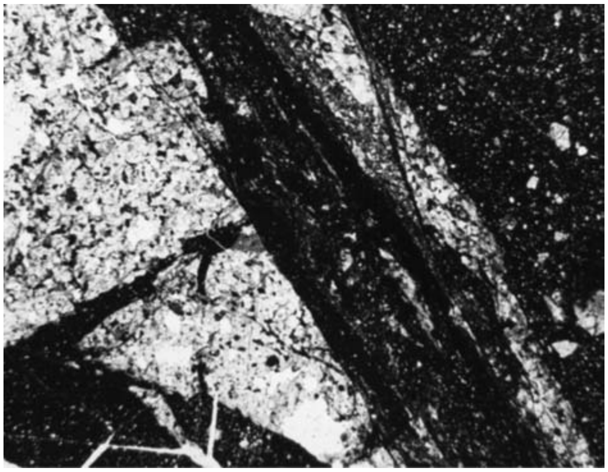


Fig. 17.22 Pseudotachylite cutting quartz monzonite, Paillé Lake, Quebec. Frictionally fused rock was injected into fractures formed by brittle failure and quenched to dark aphanitic material carrying abundant rock fragments (Philpotts, 1964). Width of field is 1 cm; plane-polarized light.

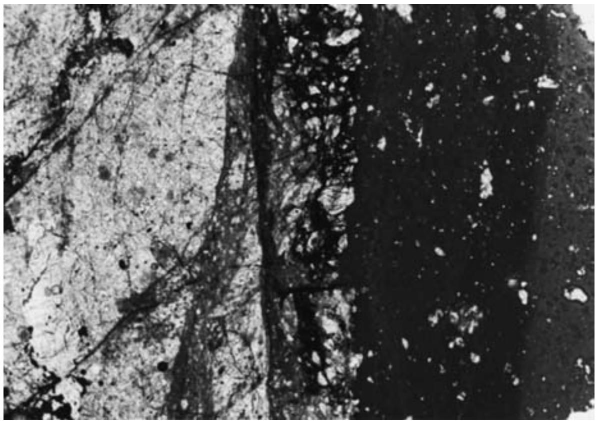


Fig. 17.23 Mylonite and pseudotachylite, Paillé Lake, Quebec (Philpotts, 1964). The pseudotachylite is a deep brown isotropic glass, which towards the margin of the dike and around feldspar fragments has become a deep red devitrified glass. Width of field is 1 cm; plane-polarized light.

of the surrounding rocks embedded in a dark aphanitic or glassy groundmass. They may or may not have a fluxion structure. In Figure 17.22, for example, the pseudotachylite extending through the center of the photograph between the two pieces of country rock has a foliation, whereas that forming the dikelets on the left and the main body on the right are devoid of flow structure; in the latter areas fragments appear to be freely suspended in the opaque groundmass.

The term *pseudotachylite* was introduced by Shand (1917) to describe dark aphanitic rock that forms networks of veins cutting the granite around the Vredefort Ring structure, South Africa (Figs. 15.17, 15.32). Because these veins did not appear to be associated with faults and most occur as isolated networks, he concluded that they must have been formed by shock and that fusion was caused by “incandescent gases.” The Vredefort Ring structure is now recognized to be a meteorite impact structure (see Section 15.9), in which case pseudotachylite is, indeed, a shock-induced glass (Dietz, 1961). Despite the meteorite impact origin of the original pseudotachylite, most other occurrences are associated with faults (Philpotts, 1964).

Although deformation in fault zones containing pseudotachylite may be largely ductile, the dilational fractures filled by pseudotachylite indicate that occasionally brittle fracture does occur when strain rates exceed the levels that can be coped with by ductile flow. The fact that ultramylonite can be fused to form pseudotachylite through frictional heating indicates that strain rates can be extremely high. This is further supported by the fact that pseudotachylite has essentially the same composition as the rock in which it occurs. This indicates that the heating must have been sufficiently intense to cause total rather than partial melting. In addition, quartz fragments in pseudotachylite may be surrounded by extremely silica-rich glass that has fusion temperatures well above those normally encountered in crustal magmatic systems. Moreover, these high temperatures must be generated in rocks that are at relatively low temperatures in order for the pseudotachylite to quench to a glass. Such short but intense thermal pulses can only be produced by the conversion of mechanical energy to heat. Other means of transferring heat in or out of these zones would be too slow to allow glass to form.

Pseudotachylite and ultramylonite can occur in zones where there is repeated movement. Inspection of Figure 17.23 reveals evidence of several periods of deformation, with some marked by mylonite and others by pseudotachylite. At times, deformation on this fault was ductile, but at others it was clearly brittle when pseudotachylite was formed. It seems likely that glassy pseudotachylite is the petrologic record of fault movements associated with large earthquakes (McKenzie and Brune, 1972).

17.6 PROBLEMS

17.1 **(a)** From the geometry of the crenulation cleavage in Figure 17.9, calculate the volume percent of the initial rock that must have been dissolved in order to create a crenulation cleavage that offsets the earlier schistosity by 1 mm if the two fabrics intersect at 30° and the crenulation cleavage planes are spaced 5 mm apart.

(b) If the mica defining the crenulation cleavage planes is the insoluble residue left after quartz has been dissolved, its thickness should be a measure of the amount of quartz removed. If the rock on either side of the crenulation cleavage planes comprises 30% muscovite and 70% quartz, and the rock in the crenulation cleavage planes is 100% muscovite, calculate what volume percent of the initial rock must have been removed by solution if the crenulation cleavage planes are 0.5 mm thick and are spaced 5 mm apart.

(c) Is it possible for the crenulation cleavage in part (b) to develop with no volume change? Explain. Assume that the muscovite in the crenulation cleavage planes represents insoluble residue.

17.2 How many different ages of mylonite and pseudotachylite are evident in Figure 17.23, and what are their relative ages?

17.3 If pseudotachylite is formed on a fault surface during an earthquake, part of the dissipated energy goes to fuse rock. If the pseudotachylite is 1 cm thick and was formed by total fusion, which required the temperature of the rock to be raised from 300 °C to 1300 °C, how much energy per square meter of fault surface would be consumed in producing the pseudotachylite? The rock has a heat capacity of 0.8 kJ kg^{-1} $°C^{-1}$, a latent heat of fusion of 418 kJ kg^{-1}, and a density of 2.8 Mg m^{-3}. Do not ignore the 10% volume increase on melting. What effect might the volume increase have on faulting?

17.4 The object of this problem is to calculate the pattern produced by a layer passing through the center of a spherical porphyroblast that rotates at a constant rate during the growth of the crystal, which is controlled by a phase boundary reaction (Eq. (12.8)). Because we are interested only in the pattern, not in the absolute dimensions, all constants can be set equal to unity. Plot x–y graphs of the trace of the marker layer for times when the marker layer at the center of the crystal forms angles of 3 and 1.9 radians from the initial foliation plane. By comparing the shapes of your calculated graphs with those in Figure 17.18 and the deformed layers in the garnet porphyroblasts of Figure 17.15, do you think that a phase boundary reaction or diffusion is most likely to have controlled the growth of these garnets? The procedure followed in solving this problem is essentially the same as that used in Eqs. (17.1) to (17.7).

17.5 A staurolite porphyroblast can be found in the upper right corner of Figure 17.8. Layers of quartz pass through this crystal and are also affected by the crenulation cleavage. Note that garnet porphyroblasts in this rock have quartz beards that bend into the crenulation cleavage. What conclusions can you draw concerning the relative ages of the two fabrics and the garnet and staurolite porphyroblasts?

17.6 Careful mapping of rotated garnet crystals in southeastern Vermont by Rosenfeld (1968) has revealed much about the timing of Acadian metamorphism relative to deformation in that area. The garnets, which have been rotated about two different axes during growth, indicate the following possible sequence of events (Fig. 17.24). First, uplift in the east caused the Silurian–Devonian Waits River Formation to flow downward to the west between the underlying and overlying rocks. This formation, which consists of relatively dense graphitic calcareous and noncalcareous pelitic schists, was already in the garnet zone at this time, as evidenced by the growth of snowball garnets. The second period of deformation, which may have been almost continuous with the first, was caused by the rise of the Green Mountain anticlinorium and the Chester Dome. Garnets continued to grow during this period but were rotated about different axes. Rotation was caused by the rise of the dome and the eastward flow of material off the anticlinorium. List the evidence preserved in the rotated garnets that supports this history. Early rotation directions are indicated by the inner circles and the later rotation direction by the outer circles.

17.7 **(a)** As shown in Figure 17.13, the distinctions between intertectonic and syntectonic inclusion patterns in porphyroblasts can be quite subtle. Explain why the porphyroblasts in Figure 17.13 are classified as syntectonic or intertectonic.

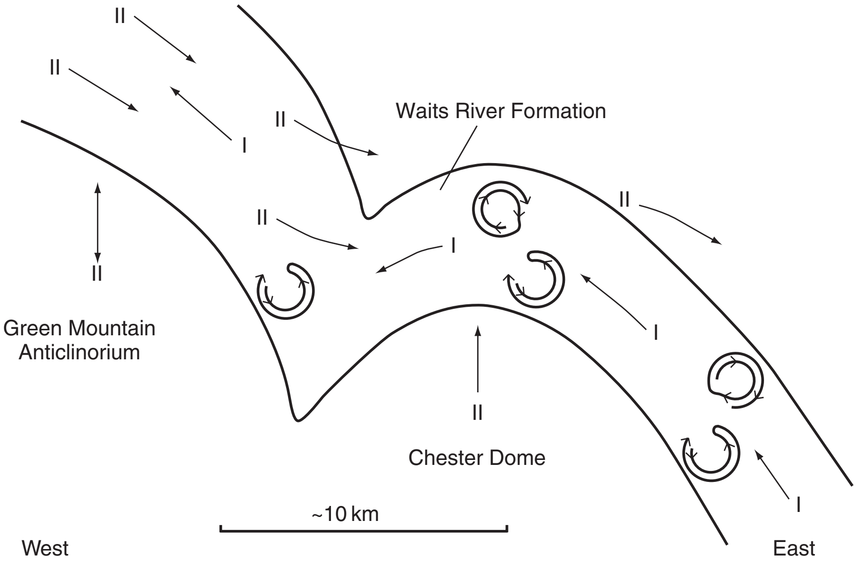


Fig. 17.24 For Problem 17.6.

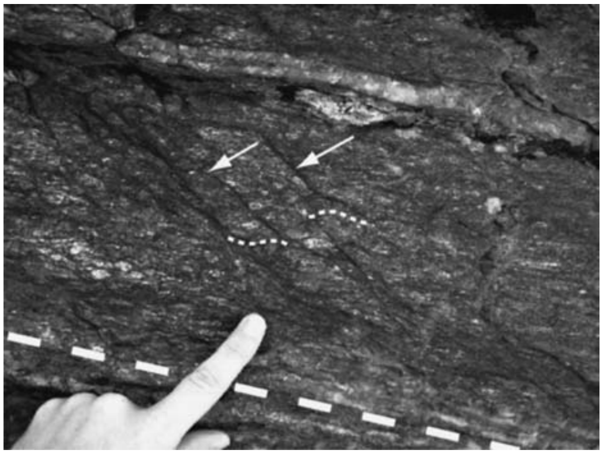


Fig. 17.25 For Problem 17.8. (Photo by Jeffrey R. Rahl.)

(b) Can you envision circumstances in which the top two intertectonic porphyroblasts of Figure 17.13(D) could actually be syntectonic?

17.8 The field photo in Figure 17.25 illustrates shear band cleavage in deformed schists from the Simplon Pass, Switzerland. The orientation of the shear zone is shown with the thick dashed line, and example shear bands and preexisting foliation planes are highlighted with arrows and dotted lines, respectively. Is this a C′-type or C-type shear band cleavage fabric? What is the sense of shear?

18 Graphical analysis of metamorphic mineral assemblages

18.1 INTRODUCTION

Thus far we have described in general terms fundamental topics in metamorphism including mineral reactions, metamorphic grade, and deformation textures. In this chapter, we turn to a more in-depth discussion of metamorphic reactions and mineral assemblages. As we saw in Chapter 16, mineral assemblages vary with metamorphic grade at the regional scale. But different rock layers often comprise different mineral assemblages at outcrop or even hand specimen scales. For example, one metasedimentary rock layer might contain quartz, muscovite, and sillimanite, whereas an adjacent layer contains only quartz and muscovite. What are the reasons for this mineralogical difference, and how can we depict the mineral assemblages in a thermodynamically rigorous way?

P–T projections (or petrogenetic grids, e.g. Fig. 16.8) show reactions for all bulk compositions in a system of interest. They don't show directly the mineral assemblages expected for a given bulk composition (although these assemblages can generally be deduced from such diagrams). What is needed are diagrams that depict mineral assemblages and that take account of mineral solid solutions and rock bulk compositions. These *compatibility diagrams* have a rich history and are the focus of this chapter. Most of the commonly used compatibility diagrams were developed when little thermodynamic or reaction rate data were available, open-system processes were difficult to quantify, and radiometric age dating was in its infancy. Nonetheless, compatibility diagrams have proven to be remarkably effective for portraying phase relations in many metamorphic rock types as functions of grade and chemical composition.

We start by examining a simple model field area that has undergone only one episode of metamorphism. This will allow us to see how field data are collected, how they are analyzed, and what metamorphic conditions they record. Compatibility diagrams and *P–T* projections are used together to show how metamorphic histories can be reconstructed. Later in the chapter, the common types of compatibility diagrams are examined.

18.2 MODEL METAMORPHIC TERRANE

Figure 18.1 is a map of an idealized metamorphic terrane underlain by folded and faulted metamorphosed sedimentary rocks. Two different types of geological boundary are shown in this map. First, there are the primary boundaries between different rock bodies, in this case sedimentary bedding planes separating mappable formations. In other areas these boundaries might also include igneous contacts. Second, there are boundaries (isograds) between different metamorphic zones that have undergone different grades of metamorphism.

Two different types of protolith form the metamorphic rocks shown in the map of Figure 18.1. Extending from west to east across the map area is a folded and faulted bed of quartzite (stippled in Fig. 18.1) formed from a feldspathic sandstone (arkose). It is overlain and underlain by muscovite schists (lined in Fig. 18.1), which were initially shales (pelites). The compositions of the protoliths are sufficiently different that the rock types are easily distinguished at all metamorphic grades, the metasandstone being composed predominantly of quartz and the metapelite of muscovite.

In tracing these rock units through the map area, we find systematic changes in their mineralogy, which allow us to define metamorphic zones. As will be evident from the discussion below, these mineralogical changes indicate increasing intensities of metamorphism from west to east.

Let us first examine the changes in the metapelitic schist. In the western corner of the area, this rock consists of muscovite, quartz, and the Al_2SiO_5 polymorph andalusite. Progressing eastward, andalusite disappears from the metapelite, and the polymorph sillimanite takes its place. The metamorphic boundary separating these two zones can therefore be named the sillimanite isograd. Farther to the east, quartz vanishes from the metapelite, and potassium feldspar appears. The boundary between these zones marks another isograd. It would be tempting to name it the K-feldspar isograd, but we will postpone naming it until the mineralogical changes in the other rock unit are examined.

The metasandstone in the western part of the area is composed of quartz, minor potassium feldspar, and muscovite. On following this unit to the east, we reach the point where the sillimanite isograd is identified in the metapelitic rock. But because the metasandstone has no andalusite to be converted to sillimanite, this isograd cannot be mapped in this unit. We see, therefore, that isograds are discontinuous boundaries that occur only in rocks of appropriate composition. For an isograd to be useful, then, it should occur in a rock that

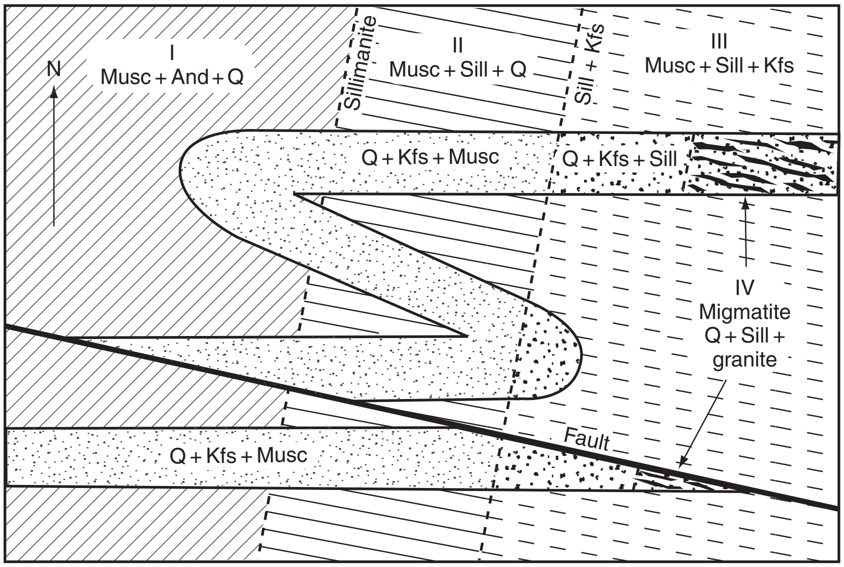


Fig. 18.1 Idealized geologic map of a model metamorphic terrane containing two protoliths, sandstone (stippled) and shale (lined). These rocks were folded and faulted prior to metamorphism. The shale was converted to a muscovite schist and the sandstone to a quartzite during metamorphism. The intensity of metamorphism increases from west to east, with four metamorphic zones being developed (Roman numerals), each with its own characteristic mineral assemblages, as shown on the map. The boundaries between zones, which are called isograds, mark the intersection of reaction surfaces in P–T–a_{H_2O} space with the erosion surface. The isograds are labeled with the new mineral or mineral assemblage appearing on the high-temperature side of the isograd.

has a wide distribution. Metapelitic rocks are particularly useful in this respect, as they contain minerals that take part in many reactions.

Where the metasandstone reaches the second isograd in the metapelitic rock a mineralogical change does take place, with sillimanite appearing and muscovite disappearing. Recall that in the metapelite, potassium feldspar appears and quartz disappears at this isograd. As will be evident from our analysis of the mineral assemblages in these zones, both sillimanite and potassium feldspar are created at this isograd, but in the metasandstone, which already contains potassium feldspar below the isograd, we recognize only sillimanite as a new mineral, and in the metapelite, which already contains sillimanite below the isograd, only potassium feldspar is recognized as a new mineral. The isograd is named the sillimanite–K-feldspar isograd, because both minerals are produced by the metamorphic reaction that occurred in both rock types along this line.

In the extreme eastern part of the area, the metasandstone develops patches and layers of coarse-grained granitic material parallel to the foliation of the rock, thus making it a migmatite. Here, temperature, pressure, and fluid composition during metamorphism reached the beginning of melting of granite. On the basis of mineralogy, then, we are able to divide the rocks of this area into four different zones, which are indicated on the map with Roman numerals. Zone I, in the western part, is recognized by the presence of andalusite with quartz and muscovite in the metapelitic schist. Zone II is also distinguished by the mineral assemblage in the metapelitic unit, with sillimanite taking the place of andalusite. The third zone, however, is marked by mineralogical changes in both rock types, quartz vanishing and K-feldspar appearing in the metapelite, and sillimanite appearing in the metasandstone. The fourth zone, marked by the presence of migmatite, is recognized only in the metasandstone. These divisions are purely descriptive and require no interpretation of the data. The resulting map showing the distribution of the primary rock units and metamorphic zones is the raw material that enables us to interpret the metamorphic history of the region.

18.3 INTERPRETATION AND REPRESENTATION OF MINERAL ASSEMBLAGES

The different mineral assemblages in the four metamorphic zones of Figure 18.1 indicate that during metamorphism the protoliths were exposed to different intensities of metamorphism, with those in zone IV reaching the melting point of granite. The mineral assemblages can be related through reactions. Some of these are easily deduced, such as the polymorphic transformation of andalusite to sillimanite between zones I and II. The sillimanite isograd can therefore be interpreted as the intersection of the andalusite → sillimanite reaction curve with the present erosion surface. Other reactions involve more minerals, such as the reaction taking place at the second isograd where sillimanite and potassium feldspar are produced at the expense of muscovite and quartz. These minerals can be related by the simple reaction

$$\begin{aligned} &\underset{\text{(muscovite)}}{\mathrm{KAl_2AlSi_3O_{10}(OH)_2}} + \underset{\text{(quartz)}}{\mathrm{SiO_2}} \\ &= \underset{\text{(K-feldspar)}}{\mathrm{KAlSi_3O_8}} + \underset{\text{(sillimanite)}}{\mathrm{Al_2SiO_5}} + \underset{\text{(fluid)}}{\mathrm{H_2O}} \end{aligned} \tag{18.1}$$

Identifying such reactions becomes more challenging with larger numbers of minerals and more widely varied rock types. It is necessary, therefore, to have techniques by which we can systematically analyze mineral assemblages and deduce what metamorphic reactions have taken place.

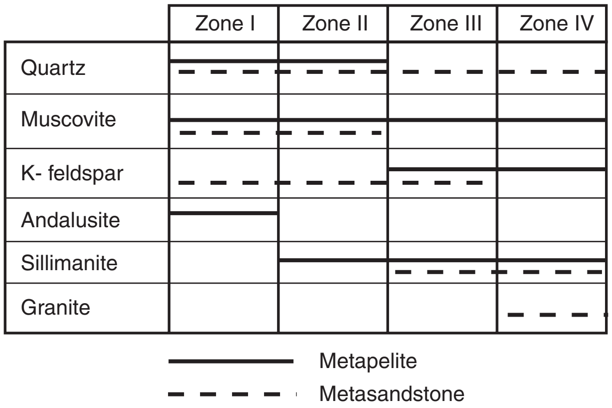


Fig. 18.2 Minerals present in the metapelite and metasandstone in the four different metamorphic zones of the area shown in Figure 18.1.

A straightforward way of systematizing the data on mineral assemblages in metamorphic zones is to use the type of diagram introduced by Miyashiro (1973). This is a simple bar graph with lines used to indicate the presence of a mineral. Figure 18.2 lists all of the minerals found in the model metamorphic terrane of Figure 18.1. Each of the four zones is represented by a column. A horizontal line in any of these columns indicates the presence of the mineral beside which the line is drawn. Because the mineral assemblages in the metapelite and metasandstone are different in each zone, they are distinguished in the graph with solid and dashed lines, respectively. In such a diagram, the changes in mineral assemblages in going from one zone to another are readily apparent. For example, in the metapelite, quartz is present in zones I and II but not zones III and IV, and conversely, potassium feldspar is not present in zones I and II but is in zones III and IV. At a glance, then, we see that the reaction that took place between zones II and III must have consumed quartz and produced potassium feldspar. This diagram, however, does not indicate all of the minerals involved in the reaction, only those that are either totally consumed or formed for the first time.

Differences in the mineral assemblages in the metasandstone and metapelite are related to differences in bulk composition. In zone I, for example, the metasandstone is not sufficiently aluminous to contain the andalusite found in the aluminous metapelite. In analyzing changes in mineral assemblages between zones, we must therefore be able to take into account the differences in bulk composition of the rocks. One way to do this is to use compositional diagrams similar to those used for phase diagrams in Chapter 10. When dealing with metamorphic rocks, these diagrams are normally plotted in terms of mole fractions rather than weight fractions. Because mole fractions are not conservative measures, the lever rule (Section 10.3) cannot be used to read off quantities from these diagrams. The use of mole fractions, however, does simplify the plotting of metamorphic minerals.

To construct such a diagram, we must first select appropriate components with which to describe all of the minerals in the map area. The minerals and their formulas are:

Quartz	SiO_2
Andalusite, sillimanite	Al_2SiO_5
Potassium feldspar	$KAlSi_3O_8$
Muscovite	$KAl_2AlSi_3O_{10}(OH)_2$

It will be recalled from Section 8.5 that components are defined as the minimum number of compositional building units required to describe all the phases present. Thus rather than selecting the five elements Si, Al, K, H, and O as components, we can reduce this number by noting that certain elements invariably go together. For example, the mineral formulas can be written in terms of the oxides SiO_2, $AlO_{3/2}$, $KO_{1/2}$, and $HO_{1/2}$ rather than the elements, because the first four elements are invariably combined with oxygen; that is, none of these elements occurs in its native state. This reduces the number of components from five to four. Note that the oxides are not written in their conventional way with whole numbers, but are expressed in terms of single cations; this greatly simplifies plotting minerals. One further simplification can be made with respect to the potassium component. Neither feldspar nor muscovite contain much $KO_{1/2}$, and if the oxide is used as one of the components, the mineral assemblages of interest will be compressed into the potassium-poor part of the diagram. If, instead, $KAlSi_3O_8$ is used as the potassium component, we can obviously make feldspar, and muscovite can be formed by combining $KAlSi_3O_8$ with $2AlO_{3/2}$ and $2HO_{1/2}$. The four components we will use to describe the minerals are, therefore, SiO_2, $AlO_{3/2}$, $KAlSi_3O_8$, and $HO_{1/2}$.

Four components can be plotted using a tetrahedron (Fig. 18.3(A)) with each apex representing one of the components (Section 10.22). The minerals in the map area can now easily be plotted in terms of mole fractions. Andalusite and sillimanite, for example, plot one-third of the way along the line from $AlO_{3/2}$ toward SiO_2, and muscovite plots on the $AlO_{3/2}$–$KAlSi_3O_8$–$HO_{1/2}$ face of the tetrahedron. Quartz and feldspar both plot at apexes.

If the minerals comprise an assemblage that was at equilibrium at the time of metamorphism, they can be joined by *tie lines* in the tetrahedral plot. In zone I, for example, quartz coexists with andalusite and muscovite in the metapelite, and with potassium feldspar and muscovite in the metasandstone. Lines can therefore be drawn from quartz to each of these minerals (Fig. 18.3(A)). Of course, if quartz coexists with andalusite and muscovite, andalusite must coexist with muscovite, and these minerals can also be joined with a tie line. The result is the *tie triangle*, quartz–andalusite–muscovite, which indicates that this is a stable assemblage in the pelitic unit of zone I. The other tie triangle that we know exists from the mineral assemblage in the metasandstone in zone I is quartz–muscovite–K-feldspar.

Inspection of Figure 18.3(A) indicates that in addition to the three-phase tie triangles, there are four-phase tie tetrahedra. For example, from the diagram it would appear possible to have

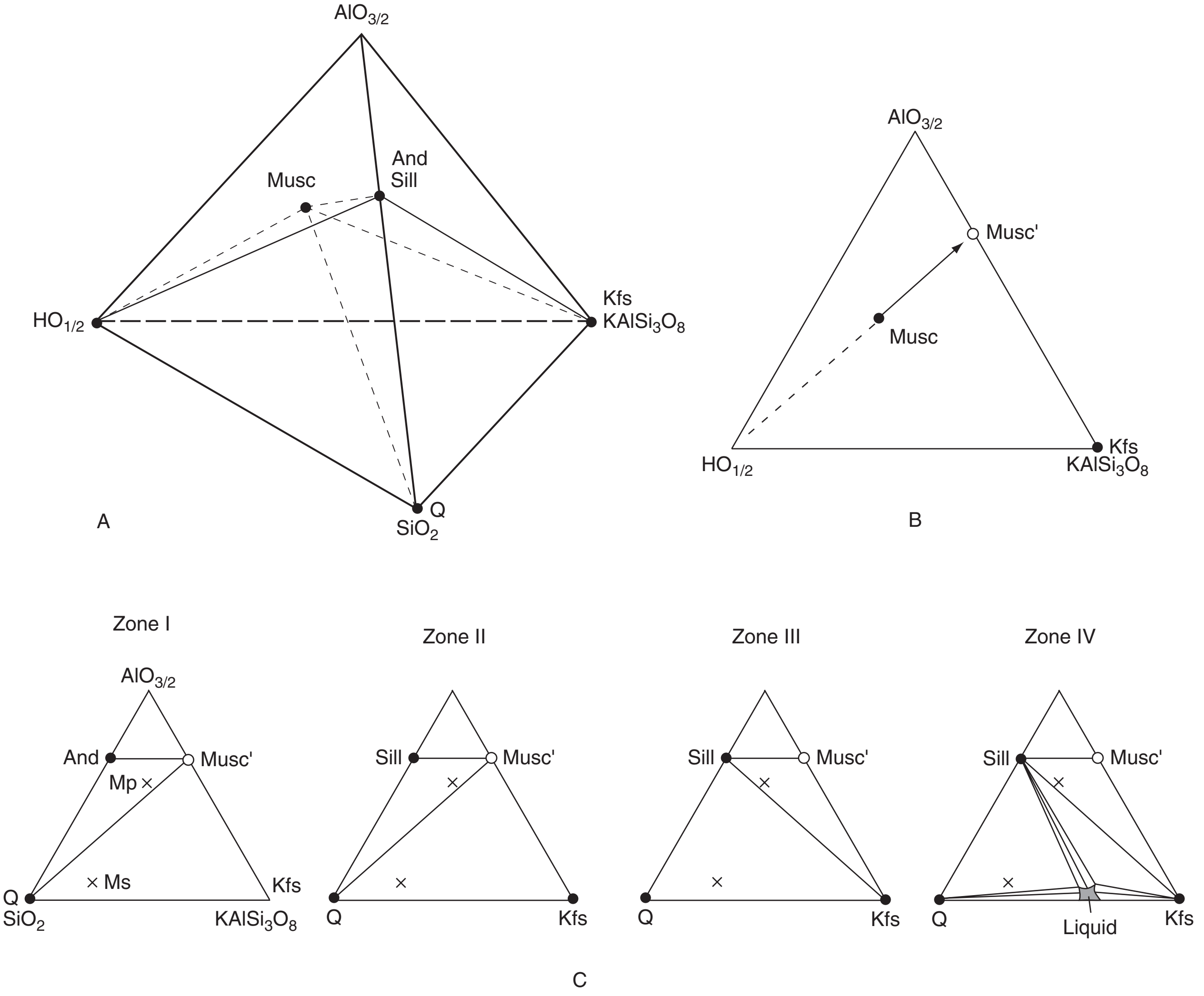


Fig. 18.3 (**A**) Tetrahedral mole fraction plot of the mineral assemblages in zone I of the model metamorphic terrane shown in Figure 18.1. Muscovite plots on the back face of the tetrahedron. (**B**) Projection of the composition of muscovite across the back face of the tetrahedron onto the water-free join $AlO_{3/2}$–$KAlSi_3O_8$. (**C**). Mineral facies diagrams for each of the metamorphic zones. Compositions within the tetrahedron of part (A) are projected onto the triangular face $AlO_{3/2}$–$KAlSi_3O_8$–SiO_2 from the water apex, assuming that the activity of water was buffered at some fixed value by the environment. The bulk composition of the metapelite is indicated by the point marked *Mp* and the metasandstone by *Ms*.

the coexistence of quartz–andalusite–muscovite–K-feldspar. This assemblage is not recorded from the area, but this is because no rocks have appropriate compositions. Two other possible four-phase assemblages are quartz–andalusite–muscovite–water, and quartz–K-feldspar–muscovite–water. If water were present in the rocks at the time of metamorphism, these two assemblages are the ones found in the metapelite and metasandstone respectively. Water is not retained as a separate phase in metamorphic rocks, except as small amounts trapped in fluid inclusions. We will assume for the moment, however, that water was present.

Each of the four-phase assemblages in Figure 18.3(A) forms a volume in the tetrahedron that does not overlap with the volumes occupied by the other phase assemblages. Thus, at the pressures and temperatures under which the mineral assemblages of zone I formed, any rock–water mixture having a bulk composition that plots in the volume quartz–andalusite–muscovite–water would have formed these phases in the appropriate proportions. This was the case for the metapelite. Similarly, a rock–water mixture having a bulk composition in the volume quartz–muscovite–K-feldspar–water would have formed these phases in the appropriate proportions. This was the case in the metasandstone. Clearly, andalusite could not have formed in the metasandstone of zone I because andalusite is separated from the bulk composition of that rock by the tie triangle quartz–muscovite–water. Diagrams like Figure 18.3(A) show what phases can coexist in thermodynamic equilibrium for a given bulk composition and, therefore, they are known as *compatibility diagrams*.

Before continuing, we must deal with water in a more satisfactory way than simply assuming that it was present during metamorphism. Fluids released during prograde metamorphism migrate through rocks. Consequently, the fluid phase in equilibrium with a rock may not have come from that particular rock, but may instead have been generated by metamorphic reactions taking place in completely different rocks at greater depth. This fluid phase could contain other components such as CO_2, chlorine, and/or dissolved metals, in which case the activity of water in the fluid phase could vary from one (pure water) to zero if the fluid phase were composed, for example, of pure CO_2. The activity of water in this mobile fluid phase is likely to be controlled by regional concentrations rather than by the local composition of any given rock. Nonetheless, the water activity may be lowered even if the fluid is generated locally in the rocks of interest. For example, metapelitic rocks may contain graphite, which can react with water at elevated temperatures to produce CO_2, CH_4, and other fluid species. Carbon is not part of the chemical system depicted in Figure 18.3(A), so the lowered water activity can be thought of as being "externally" controlled, even though the fluid is locally generated.

Because of water's activity variations and mobility, it is convenient to treat it in a fundamentally different way from the other components in the rocks. Rather than treating water as a compositional variable, we can treat it as a variable of the environment, just as we do temperature and pressure. This concept is already familiar to us because of daily weather forecasts, which give the temperature, pressure, and relative humidity. We are simply extending the application of this concept from the surface of the Earth to the environment in which metamorphic rocks are formed, and rather than using relative humidity we use activity of water (Eq. (9.34)).

In Section 10.5, the Gibbs phase rule (Eq. (10.11)) was derived for the case where temperature and pressure are the only intensive variables controlled by the environment; the remainder of the intensive variables, the chemical potentials of each of the i components, are related through Gibbs–Duhem equations to composition. The number 2 in the resulting phase rule ($f = c + 2 - \phi$) refers to temperature and pressure. If the activity, and consequently chemical potential, of water is to be treated as an environmentally controlled intensive variable, that is, its value is determined external to the mineral assemblages being considered, a modified phase rule can be written as

$$f = c + 3 - \phi \tag{18.2}$$

where c, the number of components, no longer includes water. By treating water in this way, we reduce the number of components from four to three; these are SiO_2, $AlO_{3/2}$, and $KAlSi_3O_8$. The graphical representation of mineral assemblages is now simplified, because three components can be represented with a triangle. Hydrous minerals such as muscovite are projected onto the water-free face of the tetrahedron (Fig. 18.3(B)). We must not forget, however, that such a plot represents the mineral assemblages for some value of temperature, pressure, and activity of water. Changes in any one of these variables could bring about a change in the mineral assemblages (Problem 18.1).

Figure 18.3(C) shows the ternary plots of the mineral assemblages in each of the four zones. Both rock types always consist of three minerals, which is to be expected from the modified phase rule (Eq. (18.2)). In the field, the three-phase assemblages occur over a considerable area, and therefore we can presume that they were stable over some range of temperature, pressure and fluid composition; that is, there were three degrees of freedom ($f = 3$). Because there are three components, there must, therefore, be three phases.

18.4 METAMORPHIC MINERAL ASSEMBLAGES

The ternary plots for each of the four zones (Fig. 18.3(C)) are divided into a number of triangles by the mineral assemblages in the two different rock types. Note that in any one zone, the triangles do not overlap. This indicates that given a particular composition, only one possible assemblage of minerals can form and be at chemical equilibrium under the conditions of temperature, pressure, and activity of water that existed in that zone. The metasandstone in zone I, for example, has a bulk composition that plots in the triangle quartz–muscovite–K-feldspar (Ms in Fig. 18.3(C)), and thus it must contain that mineral assemblage. The metapelitic rock, on the other hand, has a bulk composition that plots in the quartz–muscovite–andalusite triangle (Mp in Fig. 18.3(C)), and thus it must contain those minerals.

The set of nonoverlapping mineral assemblages in compositional space is classically defined as a *mineral facies*, and the diagram in which they are plotted is a *mineral facies diagram*. This diagram clearly indicates the relation between rock composition and the mineral assemblages that will form under a given set of environmentally controlled intensive variables. Recall from Chapter 16 the two fundamental tenets of the facies concept, both of which are predicated on chemical equilibrium:

1. For rocks metamorphosed under the same physical conditions, different mineral assemblages represent different bulk compositions.
2. For a given bulk composition, different mineral assemblages indicate metamorphism under different physical conditions. (Do Problem 18.1.)

Today, the term *facies* is most commonly used for the major metamorphic facies subdivisions shown in Figure 16.6. Thus, to avoid confusion, we will usually just refer to *mineral assemblages* or *equilibrium mineral assemblages* when describing phase relations like those in Figure 18.3(C), and avoid overuse of the term *facies*.

Each of the compatibility diagrams in Figure 18.3(C) contains a different mineral assemblage that is characteristic of that particular zone. The arrangement of tie triangles in zones I and II is similar, the only difference being the change from andalusite to sillimanite. The arrangement of tie triangles in zone III, however, is different from that in zone II. We now see that although the bulk compositions of the metasandstone and metapelite plot in the same positions in the ternary diagram, the tie triangles have switched, and thus the rocks contain new mineral assemblages. The nature of the reaction taking place at the boundary between zones II and III is simple to deduce from the compatibility diagrams. The change in mineral assemblage between zones II and III is brought about by the tie line between quartz and muscovite being broken and the one between sillimanite and K-feldspar forming. This indicates that at this boundary, quartz and muscovite were no longer stable together, and they reacted to form sillimanite and K-feldspar. Such a reaction is called a *tie-line switching reaction*. Each of the diagrams in Figure 18.3(C) is separated from the next by a single reaction, which indicates that conditions of temperature, pressure, or activity of water must have been different in each zone.

When the metamorphic facies concept is applied to rocks with more complex compositions than those of our model terrane, crossing tie lines may not be an indication of different metamorphic conditions but simply a product of the graphical analysis. Because the mineral assemblage diagrams of Figure 18.3(C) need only three components to describe all the mineral assemblages in the model terrane, they are rigorously correct. But when a rock requires more components, the simplifying assumptions made to portray the mineral assemblages graphically may invalidate the diagram for distinguishing the effects of bulk composition from those of metamorphic conditions on the mineral assemblage.

18.5 SIMPLE PETROGENETIC GRID

We will now return to the interpretation of the metamorphic record preserved in the rocks of the model terrane. The mineral facies diagrams allow us to systematize the relations between mineral assemblages and bulk composition and to work out the reactions relating one facies to another. The isograds are interpreted as the intersection of the P–T–a_{H_2O} reaction surfaces with the erosion surface. As we saw in Chapter 16, P–T–a_{H_2O} relations can be depicted on petrogenetic grids. The grid is divided into areas that are bounded by two or more reactions, and within any one area only one arrangement of mineral tie lines is possible; that is, each area contains a separate equilibrium mineral assemblage. By locating the mineral assemblages of the model terrane on the P–T diagram, the possible limits of pressure and temperature in the region during metamorphism can be determined for any given activity of water.

The four metamorphic zones are separated by three reactions: the polymorphic transformation of andalusite to sillimanite, the reaction of quartz with muscovite to form sillimanite and potassium feldspar, and the melting of quartz and potassium feldspar to form a granitic melt. The first reaction involves only a polymorphic transformation and therefore is not dependent on the composition of the fluid phase. But the other two reactions involve water and are consequently dependent on the activity of water in the fluid phase.

The Al_2SiO_5 polymorphs, andalusite, kyanite, and sillimanite, are extremely useful minerals to have in metamorphic rocks when trying to determine the P–T conditions of metamorphism. First, they are common constituents of metapelites. Second, their composition is almost exactly that of the simple formula; thus, substitutions do not significantly affect their stability ranges, except for Mn, which can expand the field of andalusite into those of sillimanite and kyanite (Strens, 1968). Finally, because the change from one to another is a simple polymorphic transformation the fluid phase has no effect on their relative stability relations. This is extremely important because often the activity of water during metamorphism is not precisely known. The P–T diagram for the Al_2SiO_5 polymorphs was calculated in Problem 8.3. At 500 °C and 0.38 GPa, all three polymorphs coexist at an invariant or triple point (Holdaway, 1971). Three univariant lines radiate from the invariant point dividing P–T space into three divariant fields, with kyanite occupying the high-pressure one, andalusite the low-pressure one, and sillimanite the high-temperature one. One caveat to keep in mind, however, is that reactions among the polymorphs can be kinetically sluggish. Consequently, once formed, one polymorph can persist into the stability field of another, an issue we will return to later in Chapter 21.

The reaction of quartz with muscovite to produce sillimanite and potassium feldspar is a typical metamorphic dehydration reaction. The reaction curve, which was calculated in Problem 8.8, is shown in Figure 18.4(A) for a water activity of 1. If the activity of water is less than 1, for example due to the presence of CO_2 in the fluid phase, the reaction is driven to the right in order to try and build up the water pressure, and thus the stability of quartz plus muscovite is pushed to lower temperatures (see Section 16.5). In Figure 18.4(B) the reaction is shown for a water activity of 0.3.

The beginning of melting curve of quartz and potassium feldspar is also sensitive to the activity of water because of the solubility of water in silicate melts. The melting curve in Figure 18.4 is actually for a granitic composition rather than for the quartz–K-feldspar eutectic, which is only poorly known. The quartz–K-feldspar eutectic would occur at slightly higher temperatures. The melting curve in Figure 18.4(A), which is for a water activity of 1, is almost a mirror image of the dehydration reaction, having a shallow negative slope at low pressures but steep negative one at high pressures. With decreasing water activity, the melting curve steepens as it pivots about its intersection with the temperature axis, eventually having a positive slope at low water activities (Fig. 11.11). Figure 18.4(B) shows the melting curve for a water activity of 0.3. The liquid field is shown as a single phase plotting at the quartz–feldspar eutectic in the facies diagrams because during metamorphism it was a homogeneous liquid, but on cooling it would have crystallized to a near-eutectic mixture of quartz and feldspar. Note that once melting occurs in a rock with the bulk composition of the metasandstone, all potassium feldspar enters the liquid, so that the rock still contains only three phases: quartz, sillimanite, and liquid.

With these reactions plotted on the P–T diagram, we now draw a compatibility diagram in each one of the areas on the

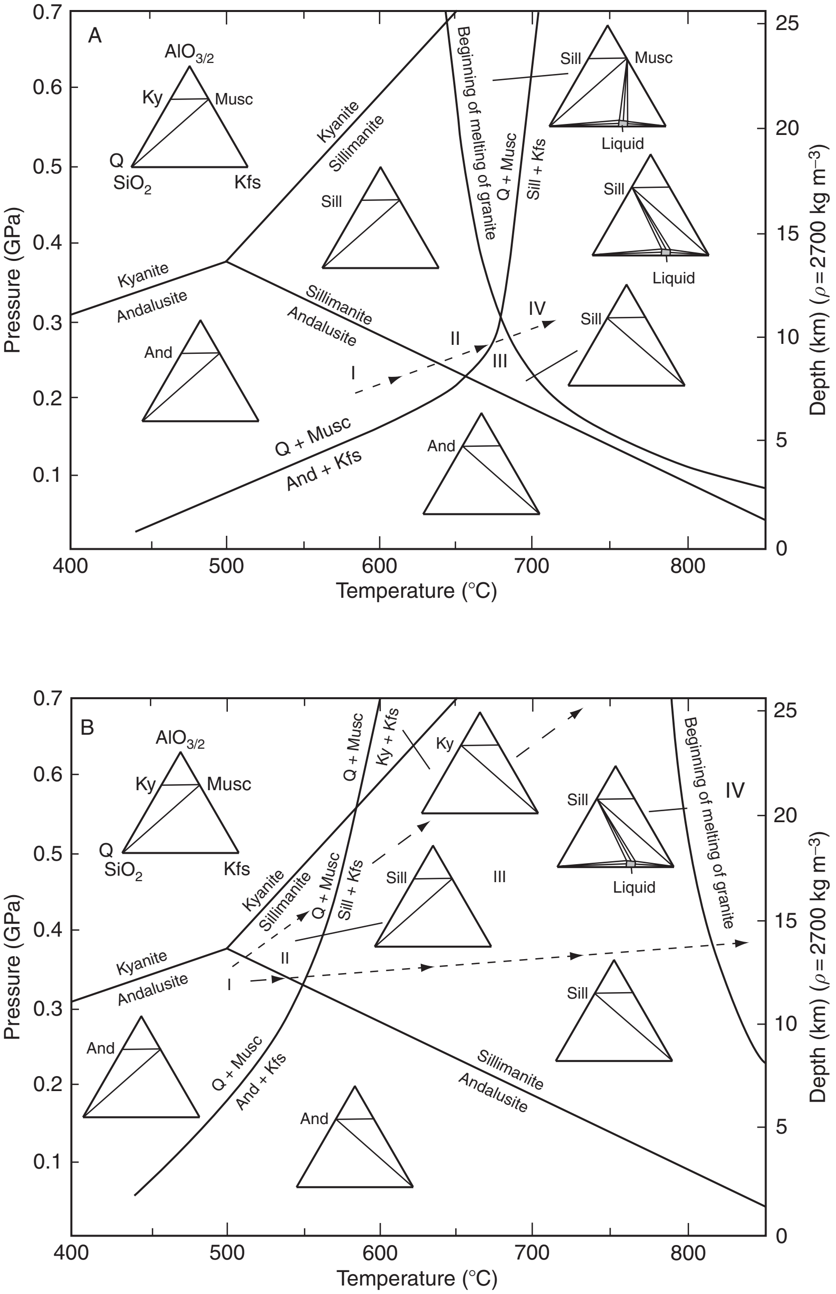


Fig. 18.4 Petrogenetic grids for rocks in the model terrane of Figure 18.1 shown for an activity of water of 1.0 in (**A**) and of 0.3 in (**B**). The possible ranges of conditions in the four zones are shown by heavy dashed lines. The stability fields of the Al_2SiO_5 polymorphs are from Holdaway (1971); the beginning of melting curves for granite are from Clemens and Wall (1981); and the stability of muscovite + quartz is from Kerrick (1972).

petrogenetic grid. It is then a simple task to identify which ones correspond to those found in the model terrane, and to determine the possible range of metamorphic *P*–*T* conditions. The *P*–*T* stability range of the mineral assemblages of zone I is limited by the reactions of andalusite to kyanite, andalusite to sillimanite, and quartz + muscovite to andalusite + K-feldspar. This sets an upper limit on the pressure in this zone of 0.38 GPa regardless of the activity of water. The upper temperature limit is given by the intersection of the dehydration reaction with the andalusite-to-sillimanite reaction. The position of this intersection depends on the activity of water. If the fluid phase is pure water ($a_{H_2O} = 1$), the maximum temperature for this zone would be 650 °C (Fig. 18.4(A)). This point of intersection also sets a lower limit on the pressure in this zone at the sillimanite isograd of 0.22 GPa. For activities of water less than 1, the intersection of these two reactions shifts to lower temperatures and higher pressures. For an activity of 0.3, for example, it occurs at 550 °C and 0.32 GPa (Fig. 18.4(B)).

The stability field of zone II is limited by the reactions of sillimanite to andalusite, sillimanite to kyanite, quartz + muscovite to sillimanite + K-feldspar, and, at high water activities, by the beginning of melting of granite. This is a large area on the *P*–*T* diagram, but at least under high water activities the possible range of conditions is limited by the position of zone III on the diagram. The isograd separating zones II and III is limited to the dehydration reaction between the sillimanite-to-andalusite reaction and the beginning of melting curve. This limits the conditions on this isograd to being between 655 °C at 0.22 GPa and 680 °C at 0.3 GPa if the activity of water is 1 (Fig. 18.4(A)). If the activity is less than 1, the shift in the dehydration reaction to lower temperatures causes the intersection with the beginning of melting curve (which shifts to higher temperatures) to be at very much higher pressures. At low enough water activities, the dehydration reaction intersects the sillimanite-to-kyanite reaction before intersecting the beginning of melting curve. At a water activity of 0.3, the upper stability limit of zone II is 0.55 GPa and 585 °C (Fig. 18.4(B)).

The upper stability limit of zone III, which is the beginning of melting of granite, is strongly dependent on the composition of the fluid phase. At high activities of water, the upper pressure limit of zone III is set by the intersection of the beginning of melting curve with the dehydration reaction (0.3 GPa and 680 °C at $a_{H_2O} = 1$). But as the activity of water decreases, this point of intersection rapidly moves to much higher pressures as the dehydration curve shifts to lower temperatures and the beginning of melting curve to higher temperatures. At low activities of water, the upper pressure limit of this zone is determined by the intersection of the beginning of melting curve with the sillimanite-to-kyanite reaction.

We are now in a position to estimate the metamorphic conditions recorded in the rocks of the model terrane. Pressure and temperature gradients must be smooth functions. Even though faulting or igneous intrusion might temporarily produce discontinuities, the flow of rock and diffusion of heat would soon eliminate such irregularities. So in addition to selecting regions in the petrogenetic grid that match the mineral assemblages from the field, temperatures and pressures must be chosen that allow for smooth variations from one zone to the next.

The deduced metamorphic conditions in the model terrane are most tightly constrained if the activity of water was 1. Zone I would then have formed at temperatures near 600 °C and 0.2 GPa, and zone IV at about 700 °C and slightly less than 0.3 GPa (dashed line in Fig. 18.4(A)). If the activity of water was less than 1, the temperature range represented by the four zones greatly increases and so also may the pressure range. For example, if the activity of water was 0.3, zone I would have formed below 500 °C and at a pressure of approximately 0.3 GPa, whereas zone IV would have formed near 800 °C and at pressures that could have been anywhere between 0.3 GPa and possibly 1 GPa. Clearly, without more knowledge of the activity of water, the conditions of metamorphism cannot be better defined with the present data. In Chapter 20 we will see that there are a number of common metamorphic reactions from which the activity of water can be determined.

Finally, having determined the possible range of metamorphic conditions in the model terrane, the question remains as to their significance. The simplest interpretation is that they represent a fossil geotherm. But if they do, they give extremely steep temperature gradients (dT/dz, shallow slope in Fig. 18.4). There can, of course, be no doubt that high temperatures were reached during metamorphism, but the interpretation that conditions in successive metamorphic zones represent geothermal gradients makes the implicit assumption that the mineral assemblages that record the information formed simultaneously. However, in Chapter 22 we will see that successive metamorphic zones can form at different times. Furthermore, regional deformation during or after metamorphism can distort isograd relations. When this happens, the spacing between isograds in the field today differs from the spacing when they developed.

18.6 *ACF* AND *AKF* DIAGRAMS

The model metamorphic terrane considered in previous sections was designed to be compositionally simple, so that the principles of analysis of mineral assemblages could be illustrated without encumbrance from complexities resulting from the graphical representation of systems containing more components. Most rocks, however, contain several more components than those used in Figure 18.3(C). The group of major oxides needed to describe the essential composition of most igneous and sedimentary rocks includes SiO_2, Al_2O_3, Fe_2O_3, FeO, MgO, CaO, Na_2O, K_2O, H_2O, and CO_2. Because graphical representations are limited to three dimensions, all of these oxides cannot be treated as separate components. However, by restricting the types of rocks that are plotted and by making a few approximations, the number of components can be reduced to three, which can therefore be represented in triangular diagrams. Two such diagrams are the common *ACF* and *AKF* plots, both of which were introduced by Eskola (1920) in his study of the rocks of the Orijarvi district of Finland.

Diagrams for plotting mineral assemblages can be greatly simplified if the compositional range of the rocks is restricted. For example, basaltic rocks contain very little potassium, and therefore this element need not be included. Such rocks can therefore be described essentially in terms of SiO_2, Al_2O_3, CaO, ferromagnesian components, Na_2O, and the volatile constituents H_2O, and CO_2. Moreover, in many ferromagnesian minerals FeO and MnO freely substitute for MgO. The number of components can therefore be reduced by treating these three oxides as a single component, which is designated *F* for ferromagnesian. A further simplification is to plot only that alumina which forms minerals other than alkali feldspar; in this way, alkali feldspar (mainly albite in metabasalt) can be ignored. To do this, the alumina is reduced by an amount equal to the number of moles of $Na_2O + K_2O$, in the rock. This simplification assumes that muscovite is not present, which is reasonable for basaltic rocks. Also, because small amounts of ferric iron substitute for alumina in minerals, Fe_2O_3 is added to the alumina. If we further assume that H_2O and CO_2 are environmentally controlled variables, they need not be considered as components. Basaltic rocks can therefore be expressed approximately in terms of the following four components, which are expressed in mole percentages:

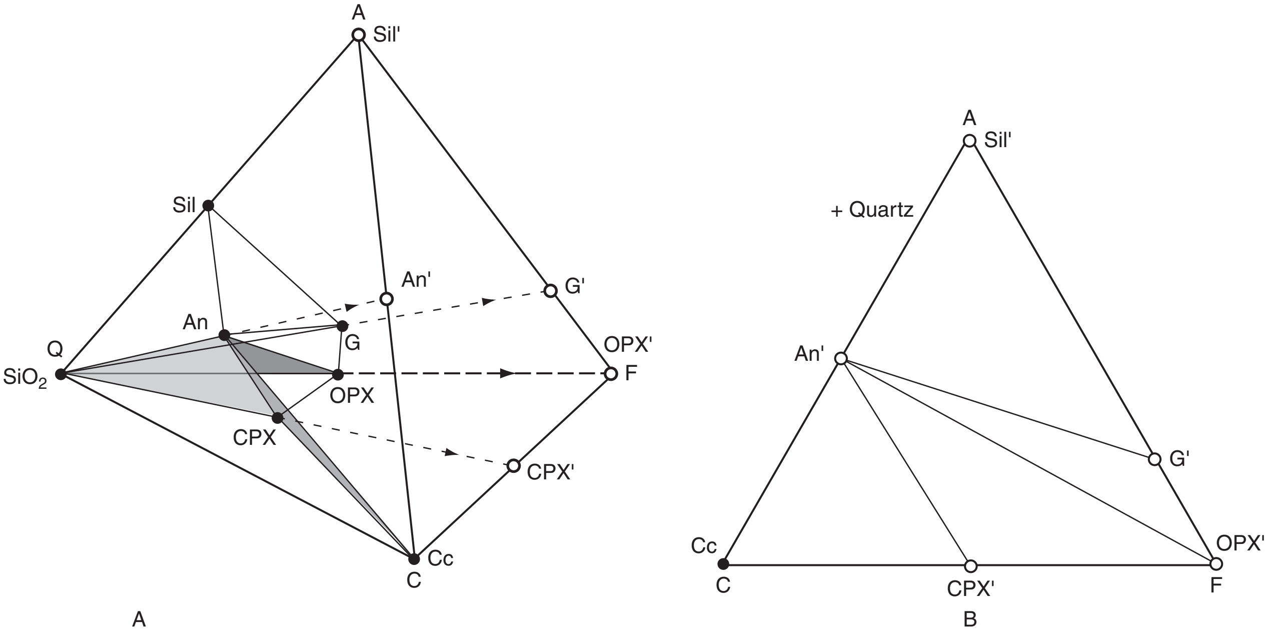


Fig. 18.5 *ACF* diagram for rocks containing quartz. (**A**) Tetrahedral plot of SiO_2–*A*–*C*–*F*, where $A = Al_2O_3 + Fe_2O_3–(Na_2O + K_2O)$, $C = CaO$, and $F = FeO + MgO + MnO$. (**B**) For rocks containing quartz, compositions can be projected from quartz onto the *ACF* face. Mineral assemblages can then be plotted on the triangular *ACF* triangle. Phases shown are quartz (Q), sillimanite (Sil), anorthite (An), clinopyroxene (CPX), orthopyroxene (OPX), garnet (G), and calcite (Cc). Projected compositions of phases are shown with a prime and open circle.

$$SiO_2$$
$$A = Al_2O_3 + Fe_2O_3 - (Na_2O + K_2O)$$
$$C = CaO$$
$$F = FeO + MgO + MnO$$

These four components can be plotted in a tetrahedron (Fig. 18.5(A)). But most metamorphosed basalts contain some amount of quartz. The mineral assemblages thus form tetrahedral volumes extending from the silica apex of the tetrahedron. If we restrict the plots to rocks containing quartz, the diagram can be further simplified by projecting all minerals and rock compositions from quartz onto the *ACF* face of the tetrahedron. Mineral assemblages appearing on the resulting triangular plot (Fig. 18.5(B)), which is the common *ACF* diagram, must all include quartz. A systematic approach to reducing the variables to a number that can be plotted is given below in Section 18.8.

Clearly, many assumptions and simplifications are made in reducing a rock to the three *ACF* components. Such plots, therefore, do not have the thermodynamic rigor of the triangular plot of Figure 18.3. Nonetheless, the *ACF* diagram is well suited for portraying generalized mineral assemblages in metamorphosed igneous rocks, especially those of basaltic composition, and metamorphosed limestones, dolostones, and impure carbonate rocks. It is not suitable for rocks formed from aluminous sediments or ultramafic rocks. When metamorphosed, metapelites contain a number of ferromagnesian minerals, such as staurolite, chloritoid, garnet, cordierite, and chlorite, that tend to be either iron-rich or magnesium-rich; that is, iron and magnesium do not substitute freely for each other and therefore cannot be treated as a single component. When plotted in the *ACF* diagram, these minerals commonly form four-phase rather than three-phase assemblages and may also have crossing tie lines. Such rocks require a completely different plot (Section 18.8). The *ACF* diagram is also unsuitable for many meta-ultramafic rocks.

To plot the composition of a rock in the *ACF* diagram, the chemical analysis, which is normally presented in terms of weight percent, is recast in mole proportions by dividing the weight percentage of each oxide by its molecular weight. The mole proportions of the *A*, *C*, and *F* components are then recalculated to 100%. If accurate modal data are available, corrections can be made to the mole proportion of CaO for any accessory apatite or sphene present in the rock, and to FeO for accessory ilmenite and magnetite (see norm calculation, Section 6.3). Normally, the amounts of these minerals are small, and they do not seriously affect the *ACF* plot.

The positions of the minerals in the *ACF* plot can be deduced from their formulas. For convenience, their compositions in terms of *A*, *C*, and *F* are given in Table 18.1.

Most rocks form, in addition to quartz, a three-phase assemblage in the *ACF* diagram (accessory minerals are neglected). Some minerals, such as hornblende, however, have such a wide range of composition that some bulk compositions may fall in two-phase or even single-phase fields in *ACF* diagrams.

The typical mineral assemblages formed under the range of pressures and temperatures encountered during metamorphism are shown in Figure 18.6. To illustrate how these vary, we will consider the mineralogical makeup of a typical basalt

Table 18.1 *Mineral compositions for* ACF *diagrams*

Mineral	*A*	*C*	*F*
Anthophyllite, cummingtonite, Orthopyroxene, talc, serpentine	0	0	100
actinolite, tremolite	0	28.5	71.5
Hornblende	<25	~ 28.5	71.5
Diopside, dolomite	0	50	50
Calcite, wollastonite	0	100	0
Grossularite, andradite	25	75	0
Vesuvianite	14	72	14
Epidotes	43	57	0
Anorthite	50	50	0
Al_2SiO_5, pyrophyllite	100	0	0
Staurolite	67	0	33
Chloritoid (Fe), cordierite (Mg)	50	0	50
Spesartine, almandine, pyrope	25	0	75
Chlorite	10–35	0	90–65

in each facies. Remember that in addition to the minerals in the *ACF* diagram, quartz is also present (Problem 18.2). In the pyroxene hornfels facies, a tholeiitic basalt would have essentially the same mineralogy it had when it crystallized from the magma: that is, plagioclase, clinopyroxene, and orthopyroxene (+ quartz); we will assume that the basalt has a composition near the center of this mineral tie triangle. At lower temperatures and under water-saturated conditions, hornblende is stable in the hornblende hornfels facies. The bulk composition of the basalt would plot in the two-phase region plagioclase + hornblende. At still lower temperatures, in the albite–epidote hornfels facies, calcic plagioclase is not stable; instead, it is replaced by albite + epidote, and the coexisting amphibole is actinolite rather than hornblende.

Another less commonly used plot is the *AKF* diagram. The principles involved in its construction are similar to those for the *ACF* diagram, except that it is used for rocks containing

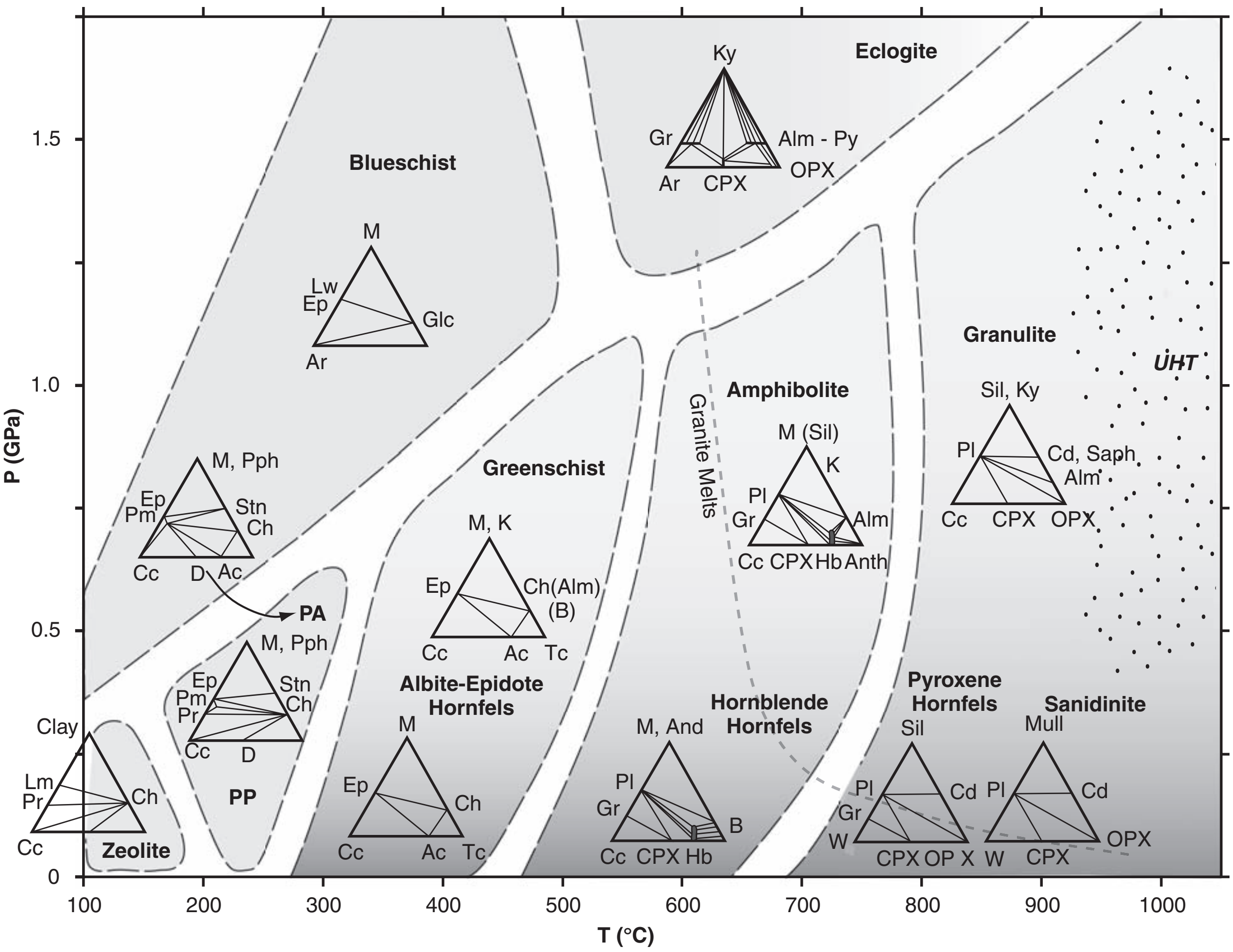


Fig. 18.6 *ACF* plots of common quartz-bearing mineral assemblages in the metamorphic facies. Boundaries and conditions are the same as in Figure 16.6. Minerals plotted include andalusite (And), kyanite (Ky), sillimanite (Sil), muscovite (M), mullite (Mull), grossularite (Gr), almandine (Alm), pyrope (Py), orthopyroxene (OPX), clinopyroxene (CPX), calcic plagioclase (Pl), epidote (Ep), lawsonite (Lw), laumontite (Lm), pumpellyite (Pm), prehnite (Pr), calcite (Cc), aragonite (Ar), dolomite (D), wollastonite (W), actinolite (Ac), hornblende (Hb), glaucophane (Glc), anthophyllite (Ath), talc (Tc), biotite (B), chlorite (Ch), cordierite (Cd), pyrophyllite (Pph), sapphirine (Saph), and stilpnomelane (Stn). The beginning of melting curve for water-saturated peraluminous granite is from Clemens and Wall (1981).

excess alumina and silica, which have the minerals muscovite, quartz, and plagioclase. Quartz and plagioclase are not plotted; instead, the alumina content is reduced by the amount that would enter feldspar. The three components, expressed in mole fractions, are

$$\begin{aligned}
&\mathrm{A} = \mathrm{Al_2O_3} - (\mathrm{CaO} + \mathrm{Na_2O} + \mathrm{K_2O})\\
&\mathrm{K} = \mathrm{K_2O}\\
&\mathrm{F} = \mathrm{FeO} + \mathrm{MgO} + \mathrm{MnO}
\end{aligned}$$

Serious problems arise from grouping MgO and FeO as one component, because many of the ferromagnesian minerals in the rocks for which this diagram is designed (metapelites) do not freely substitute MgO for FeO. For this reason a diagram in which MgO and FeO are treated separately is preferred (see Section 18.8).

The boundaries between the various metamorphic facies in Figure 18.6 are determined by important metamorphic reactions. These boundaries do not, however, produce a petrogenetic grid. Inspection of the petrogenetic grid created for the model metamorphic terrane in Figure 18.4 reveals that facies diagrams in adjoining areas of the grid differ by only one reaction. This is certainly not the case for adjoining areas in Figure 18.6. Compare, for example, the *ACF* compatibility diagrams for the greenschist and amphibolite facies. Many tie lines must be broken or switched in going from one to the other; that is, several reactions must take place to convert the assemblages of the one facies into those of the other (Problem 18.3). These reactions do not occur at precisely the same pressures and temperatures. We can therefore conclude that within the upper part of the greenschist facies or lower part of the amphibolite facies a number of reactions must occur to transform the assemblages of the greenschist facies shown in Figure 18.6 to those of the amphibolite facies.

18.7 REPRESENTATION OF SOLID SOLUTIONS

Solutions are encountered in many different forms in petrology. A magma, for example, is a liquid silicate solution; the fluid phase in a metamorphic reaction is a supercritical solution involving H_2O and CO_2; and many of the common metamorphic minerals are solid solutions involving substitution of Fe for Mg. Solid solutions constitute many of the essential minerals of a metamorphic rock, and in their compositions is preserved valuable information about the metamorphic history of a rock.

In constructing *ACF* and *AKF* diagrams, FeO and MgO are assumed to behave identically in order to minimize the number of components that have to be plotted. Although Fe does substitute for Mg in many metamorphic minerals, it is erroneous to believe that these elements have the same effect on mineral reactions, even when the solid solutions behave ideally. We have already seen in Chapter 10 that olivine, for example, forms an almost ideal solid solution between fayalite (Fe_2SiO_4) and forsterite (Mg_2SiO_4); and yet when olivine melts, the liquid is invariably richer in Fe than the coexisting solid (Fig. 10.11). Similarly, in metamorphic rocks, garnet, for example, is invariably more Fe-rich than coexisting chlorite. To account for such compositional differences, FeO and MgO must be treated as separate components.

The variance of mineral assemblages is increased by solid solution. Thus a metamorphic reaction that would occur at a specific temperature under a given pressure if it involved minerals of fixed composition takes place over a temperature range when the minerals form solid solutions. By analogy with igneous systems, these two cases are equivalent to reactions taking place at a peritectic, such as that of olivine + liquid = enstatite (Fig. 10.7), or along continuous reaction curves such as the liquidus and solidus in the olivine system (Fig. 10.11). We will see that metamorphic reactions can also be classified as being either discontinuous or continuous.

Although the added variance caused by solid solutions appears to reduce our ability to determine metamorphic conditions from mineral assemblages, the reverse is actually the case. If equilibrium is to be achieved, when a rock contains several minerals belonging to solid solution series, elements involved in substitutional exchange must partition themselves between these minerals in accordance with the thermodynamic laws governing solutions (Chapter 9). Because partition coefficients are functions of temperature and, in some cases, pressure, the compositions of some coexisting minerals can be used as geothermometers or geobarometers.

In no group of rocks are the effects of solid solution on mineral equilibria more evident than in those of pelitic composition. The rest of this chapter therefore concentrates on these rocks, but the principles dealt with are applicable to rocks of any composition.

18.8 GRAPHICAL REPRESENTATION OF MINERAL ASSEMBLAGES IN SYSTEMS OF FOUR OR MORE COMPONENTS

Rocks formed from mud are known as shales or pelites. They constitute approximately two-thirds of all sedimentary rocks. They are composed essentially of clay minerals and variable amounts of extremely fine-grained detrital material (largely quartz). The average shale contains 62% SiO_2, 17% Al_2O_3, 7% FeO^t, 3.7% K_2O, 2.4% MgO, 1.5% CaO, 1% Na_2O, and up to 5% H_2O; all other components, and there may be many, are normally present in minor or trace amounts. Despite their small compositional range, metapelites can contain a large number of minerals, many of which form assemblages with limited *P*–*T* stability ranges. These rocks are, therefore, particularly sensitive indicators of metamorphic conditions, and for this reason they have been studied more intensively than any other group of metamorphic rocks. Barrow (1893, 1912), for example, used the first appearance of chlorite, biotite, almandine garnet, staurolite, kyanite, and sillimanite in metapelites to map isograds in the southeast Scottish Highlands (Fig. 16.1(B)).

Most metapelites contain quartz, muscovite, and one or a few essential minerals from the list given in Table 18.2.

Table 18.2 *Common minerals in metapelitic rocks and their formulas*

Quartz and other silica polymorphs	SiO_2
Muscovite	$KAl_2(AlSi_3O_{10})(OH,F,Cl)_2$
Chlorite	$(Mg,Fe,Al)_6(Al,Si)_4O_{10}(OH)_8$
Biotite	$K(Mg,Fe)_3(AlSi_3O_{10})(OH,F,Cl)_2$
Almandine	$Fe_3Al_2Si_3O_{12}$
Spessartine	$Mn_3Al_2Si_3O_{12}$
Staurolite	$(Fe,Mg)_2Al_9Si_{3.75}O_{24}H_2$
Andalusite, Kyanite, Sillimanite	Al_2SiO_5
Mullite	$3Al_2O_3{\cdot}2SiO_2$
Chloritoid	$(Fe,Mg,Mn)Al_2SiO_5(OH)_2$
Cordierite	$(Mg,Fe)_2Al_4Si_5O_{18}$
Sapphirine	$(Mg,Fe)_4Al_8Si_2O_{20}$
Orthopyroxene	$(Mg,Fe)_2Si_2O_6$
Orthoclase	$(K,Na)AlSi_3O_8$
Plagioclase	$NaAlSi_3O_8$–$CaAl_2Si_2O_8$
Paragonite	$NaAl_2(AlSi_3O_{10})(OH)_2$

Clearly, a large number of components is necessary to describe these phases. If the mineral assemblages are to be graphically analyzed, the number of components must be reduced to a manageable number while retaining those that show significant changes in mineral assemblages. The important question is: which components need to be plotted, and which can be ignored? The necessary components are those whose variation in concentration causes changes in mineral assemblages; those that do not cause changes can be ignored. Four simple rules allow us to distinguish these two types of component. In applying these rules, we are following a procedure introduced by J. B. Thompson (1957) for the graphical analysis of metapelitic schists.

The variance of an assemblage of minerals is defined as the number of variables that can be independently changed without changing the assemblage (Eq. 10.11). Numerically this is equal to the total number of variables minus the number of independent equations relating the variables. For each phase present, a Gibbs–Duhem equation (Eq. (9.7)) can be written, which describes the relations between the various intensive variables in any phase in internal, homogeneous equilibrium (i.e. no compositional gradients). Thus for a phase α, we can write

$$S^{\alpha}\mathrm{d}T - V^{\alpha}\mathrm{d}P + \sum n_i^{\alpha}\mathrm{d}\mu_i = 0$$

The diagram in which such a phase is to be plotted will be for a given temperature and pressure. This leaves only the chemical potentials of the various components as variables. The following rules show which of these chemical potentials can be independently varied, and thus which ones play a role in determining mineral equilibria.

Rule 1

> *Components that occur as pure phases, such as SiO_2 in quartz, TiO_2 in rutile, or Fe_2O_3 in hematite, need not be plotted.*

The Gibbs–Duhem equation for any silicate, S, has the following form:

$$S^{\mathrm{S}}\mathrm{d}T - V^{\mathrm{S}}\mathrm{d}P + \sum n^{\mathrm{S}}_{\mathrm{SiO_2}}\mathrm{d}\mu_{\mathrm{SiO_2}} = 0$$

If quartz (Q) is present, then $\mu_{\mathrm{SiO_2}} = \mu^{\mathrm{Q}}_{\mathrm{SiO_2}}$, but $\mu^{\mathrm{Q}}_{\mathrm{SiO_2}}$ is simply the molar free energy of quartz, which is a function only of temperature and pressure. The chemical potential of silica is therefore not an independent compositional variable. Another way of expressing this is to say that the rock is saturated in quartz. Any addition or subtraction of SiO_2 from the rock simply varies the amount of quartz but does not change the chemical potential of SiO_2; variation in the amount of silica does not, therefore, affect the stability of the other minerals. A rock is saturated in all those components that occur as pure phases, and because variations in their amounts cannot change mineral assemblages, these components need not be plotted.

Rule 2

> *Components whose chemical potentials are controlled external to the system need not be plotted.*

This rule applies mostly to components that appear in the fluid phase, such as H_2O and CO_2, but it could include others if they are sufficiently mobile. The Gibbs–Duhem equation for a fluid containing H_2O is

$$S^f\mathrm{d}T - V^f\mathrm{d}P + n_i^f\mathrm{d}\mu_i + n^f_{\mathrm{H_2O}}\mathrm{d}\mu_{\mathrm{H_2O}} = 0$$

But $\mu^f_{\mathrm{H_2O}} = \mu^{f*}_{\mathrm{H_2O}} + RT\ln\left(\mathrm{f}/10^5\right)$. Now $\mu^{f*}_{\mathrm{H_2O}}$ is a function only of pressure and temperature, which is fixed for any given compatibility diagram. If the fugacity (f) of H_2O is buffered by an external reservoir, then, $\mu^f_{\mathrm{H_2O}}$ is not an independent variable. It is common in such cases to express $\mu^f_{\mathrm{H_2O}}$ in terms of activity:

$$\mu^f_{\mathrm{H_2O}} = \mu^{f*}_{\mathrm{H_2O}} + RT\ln\left(a^f_{\mathrm{H_2O}}\right)$$

from which it follows that

$$a^f_{\mathrm{H_2O}} = \exp\left(\frac{\left(\mu^f_{\mathrm{H_2O}} - \mu^{f*}_{\mathrm{H_2O}}\right)}{RT}\right) \tag{18.3}$$

Thus a pure H_2O fluid has $a^f_{\mathrm{H_2O}} = 1$, and if the fluid contains a mixture of components, $a^f_{\mathrm{H_2O}} < 1$. Although H_2O need not be treated as a component if the activity of water in the fluid phase is buffered at a particular value, we must remember that the graphical representation of the mineral assemblages will be for only that one particular activity of H_2O.

Rule 3

> *Components occurring in only one phase need not be plotted if we also ignore the phase in which the component occurs.*

Some components are important in only one phase, for example ZrO_2 in zircon, Na_2O in albite, CaO in plagioclase, and P_2O_5 in apatite. For each one of these phases a separate Gibbs Duhem equation can be written, but each equation introduces just one new variable, that of the chemical potential of the component that occurs just in that phase. From the standpoint of the phase rule, the additional variable is balanced with the additional equation, and the variance of the system is unchanged by including these phases and components. Rule 3 perhaps offers the greatest opportunity to reduce the number of components erroneously. Before a component is eliminated by this rule, you must be certain that the component occurs in only one phase. Na_2O, for example, might be present, not only in albite, but in the hard-to-detect paragonite component of muscovite, and CaO might occur in both plagioclase and garnet. If such second phases are present, the elimination of the component may result in assemblages that apparently violate the phase rule.

Rule 4

> *Components that are not sufficiently abundant to stabilize a mineral can be ignored.*

This rule applies to the trace components in rocks. As long as an element's concentration is so low that it does not affect the stability of minerals, it need not be treated as a separate component. Instead, its amount can be added to the major element for which the trace component substitutes. Thus Ni^{2+} and Co^{2+} can be added to Fe^{2+}, and Sr^{2+} can be added to Ca^{2+}. This grouping, however, must be done with care. Mn^{2+}, for example, is grouped with Fe^{2+}, but if garnet is present, Mn^{2+} preferentially enters this phase, thus stabilizing it. Zn^{2+}, on the other hand, preferentially enters staurolite, helping to stabilize that mineral. And Sr^{2+} is known to stabilize aragonite with respect to calcite. Many elements are normally present in such small amounts that their stabilizing effect is negligible. But one must remain alert to the possibility that minerals with elevated concentrations of these elements may be present, such that apparent violations of the phase rule may occur.

Let us now apply these rules to reducing the number of components in a metapelite. Most metapelitic schists are composed essentially of quartz + muscovite and up to three other minerals from Table 18.2. At the highest grades of metamorphism, muscovite is not stable and breaks down to form K-feldspar via Eq. (18.1). In addition to the essential minerals, most metapelites contain a number of accessory minerals, such as hematite, magnetite, ilmenite, rutile, pyrrhotite, pyrite, graphite, zircon, tourmaline, and apatite. The formulas of the essential minerals are given in Table 18.2.

The minerals in a typical metapelitic schist are composed of the following components: SiO_2, TiO_2, Al_2O_3, Fe_2O_3, FeO, MnO, ZnO, MgO, CaO, Na_2O, K_2O, ZrO_2, B_2O_3, H_2O, C, and S. This is a long list, but we will now proceed to reduce it to only four components.

Most pelitic rocks contain quartz so, according to rule 1, SiO_2 need not be plotted. Carbon can similarly be eliminated if it occurs as graphite, TiO_2 if it occurs as rutile, and Fe_2O_3 if it occurs as hematite. We eliminate H_2O as a component by rule 2, expressing it rather as an intensive environmental variable. It is important to emphasize, however, that the compatibility diagram will be restricted to a certain activity of H O.

Under rule 3, the following components are eliminated because each occurs in only one phase: CaO (plagioclase), Na_2O (plagioclase), ZrO_2 (zircon), B_2O_3 (tourmaline), and S (pyrrhotite). Note, however, that CaO can be problematic if it is important in both plagioclase and garnet. If the rocks lack rutile and hematite (rule 1 above), TiO_2 and Fe_2O_3 can be eliminated under rule 3 if the rocks contain ilmenite and magnetite. The elimination of TiO_2 and Fe_2O_3 is not well justified if some of the silicate minerals, like biotite, are rich in Ti or ferric iron.

The trace components MnO and ZnO are eliminated by adding their amounts to FeO, keeping in mind that any significant amount of these elements may cause problems in plotting rocks containing garnet and staurolite.

By these rules, the 16 components needed to describe all the minerals in a pelitic schist are reduced to only four, Al_2O_3–FeO–MgO–K_2O (*AFMK*), which can be plotted in a tetrahedron (Fig. 18.7(A)). Most of the minerals plot on the *AFM* face of the tetrahedron with the exception of muscovite, biotite, and K-feldspar. One further simplification can be made; that is, to project points plotting within the tetrahedron onto a face of the tetrahedron. This we do according to the following rule.

Rule 5

> *If a component occurs in two or more phases, and one phase is common to all assemblages, project the composition of all phases with this component from the composition of the common phase onto a face of the tetrahedron or some other convenient plane.*

Most metapelitic rocks contain muscovite. Thus, if we restrict the plot to rocks containing muscovite (+ quartz), compositions in the tetrahedron can be projected from the common phase, muscovite, onto the *AFM* face. This projection is shown graphically in Figure 18.7(A). The advantage of the projection is that it reduces the three-dimensional *AFMK* tetrahedron, which can be difficult to visualize, to a simpler two-dimensional diagram. Note, however, that software is now available to make stereographic images of three-dimensional diagrams in order to aid visualization (Spear *et al*., 1991; http://ees2.geo.rpi.edu/MetaPetaRen/Software/Software.html). To project a point quantitatively onto the *AFM* face, the K_2O content of the point must be reduced to zero by subtracting the composition of the point from which the projection is made – muscovite in this case ($K_2O{\cdot}3Al_2O_3$). The alumina content of the projected point is therefore obtained by reducing the initial value by the alumina removed with K_2O as muscovite, that is, $Al_2O_3 - 3K_2O$. To plot the projected point on the *AFM* face, the new alumina content plus the FeO and MgO are recalculated to a total of one. The alumina value then becomes

$$\text{alumina index} = \frac{Al_2O_3 - 3K_2O}{(Al_2O_3 - 3K_2O) + FeO + MgO} \qquad (18.4)$$

Biotite, for example, has a composition of $K_2O{\cdot}6(FeO, MgO){\cdot}Al_2O_3$ in the tetrahedron. Its alumina index is therefore $-2/(-2+6) = -0.5$. Consequently, biotite does not fall within the *AFM* triangle but plots below its base, as would be

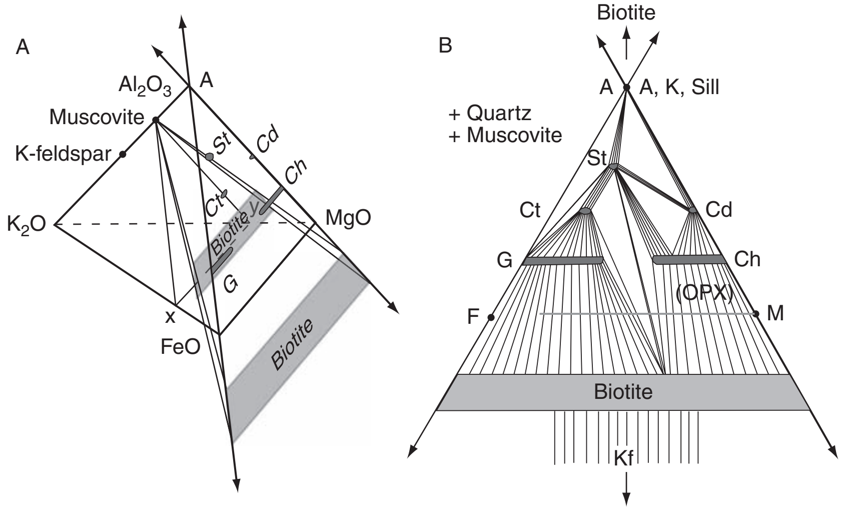


Fig. 18.7 Projection of mineral compositions at fixed water activity in quartz- and muscovite-bearing metapelites onto the *AFM* face of the Al_2O_3–FeO–MgO–K_2O tetrahedron. (**A**) Projection from muscovite in the tetrahedron. (**B**) Plotting positions of common minerals on the *AFM* face. Minerals plotted are Al_2SiO_5 polymorphs (A, K, Sill), staurolite (St), chloritoid (Ct), cordierite (Cd), garnet (G), chlorite (Ch), biotite, muscovite (M), and potassium feldspar (Kf). Plotting position of orthopyroxene (OPX) is also shown. (Modified after Thompson, 1957; published by permission of *American Mineralogist*.)

expected from the graphical projection from the tetrahedron (Fig. 18.7(A)). Points plotting in the subtetrahedron bounded by muscovite, biotite, and the two points *x* and *y* project onto the *AFM* plane below biotite at distances that approach $-\infty$ as compositions approach the plane muscovite–*x*–*y*, which is parallel to the *AFM* face. Compositions plotting in the subtetrahedron muscovite–*x*–*y*–K_2O, rather than being projected *from* muscovite, can be projected onto the *AFM* face only *through* muscovite (negative projection), intersecting the face above *A*. K-feldspar projects through muscovite to fall right on the *A* apex of the *AFM* triangle.

All of the essential minerals in metapelites can be plotted in the *AFM* projection (Fig. 18.7(B)), and their range of compositions in terms of FeO and MgO can be shown clearly. Staurolite allows little substitution of Mg for Fe and plots as a small area near the *AF* side of the diagram. Chloritoid allows a greater, but still limited substitution of Mg, with Mg-rich compositions forming cordierite instead. Garnet and chlorite have a wide range of compositions, but garnets are always Fe-rich relative to coexisting chlorite. The Mg-rich garnet, pyrope, is stable only at ultrahigh pressures. Stilpnomelane tends to be Fe-rich, whereas biotites have a complete range of Fe:Mg values; they also have variable alumina contents.

At the highest grades of metamorphism, quartz and muscovite react to form K-feldspar and an aluminum silicate (Eq. (18.1)). For such rocks, K-feldspar is the common phase, and so compositions in the tetrahedron are projected onto the *AFM* face from feldspar. Because most minerals lie in the *AFM* plane, the new projection does not change their positions. Biotite, however, plots at a slightly higher (less negative) value than in the muscovite projection (Problem 18.4).

18.9 VARIANCE IN METAPELITIC MINERAL ASSEMBLAGES

Figure 18.8 shows typical *AFM* diagrams for metapelitic rocks. We consider next the effect of solid solution on the variance, or degrees of freedom, of a mineral assemblage.

If the mineral assemblages depicted in Figure 18.8 occur in rocks over an extensive geographic region, the *AFM* diagram can be assumed to hold true over some range of *T*, *P*, and a_{H_2O}. Thus, from the viewpoint of the phase rule ($\phi + f = c + 3$, the numeral 3 referring to *T*, *P*, and a_{H_2O}), three intensive variables were able to vary without causing changes in the mineral assemblages. The number of components is only three because all others have been eliminated according to rules 1 to 4. Consequently, for the variance we have $f = 6 - \phi$.

The variance of a rock containing three minerals in an *AFM* diagram (plus quartz and muscovite) is 3. But three degrees of freedom are required to define the *T*, *P*, and a_{H_2O} of the diagram, leaving no degrees of freedom for varying mineral compositions. Thus in any three-phase assemblage, such as chloritoid + garnet + chlorite, the composition of each mineral is fixed for a given *T*, *P*, and a_{H_2O}. In this case, variations in the bulk composition of the rock result in variations in the proportions of minerals but not in their compositions. Note that minerals in some three-phase assemblages have fixed compositions regardless of the pressure and temperature (kyanite, for example). Others belong to solid solution series, and although their compositions are fixed for a given *T*, *P*, and a_{H_2O}, under different conditions their compositions are fixed at different values. This allows the composition of minerals in some three-phase assemblages to be used for geothermometry and geobarometry (Chapter 19).

If only two minerals are present in an *AFM* plot, the variance is 4. Again, three degrees of freedom are needed to specify the *T*, *P*, and a_{H_2O}, leaving one degree of freedom for composition. If the two minerals are chloritoid and chlorite, for example, both belong to Fe–Mg solid solution series, and chlorite even exhibits some variability in its Al content. However, only one degree of freedom exists among these compositional variables. Chlorite coexisting with chloritoid must contain its maximum amount of Al; that is, it must plot on the Al side of the chlorite field. If the one degree of freedom is used to specify the Fe/Mg ratio in the chlorite, the composition of coexisting chloritoid is then determined, with the composition being indicated by the chloritoid

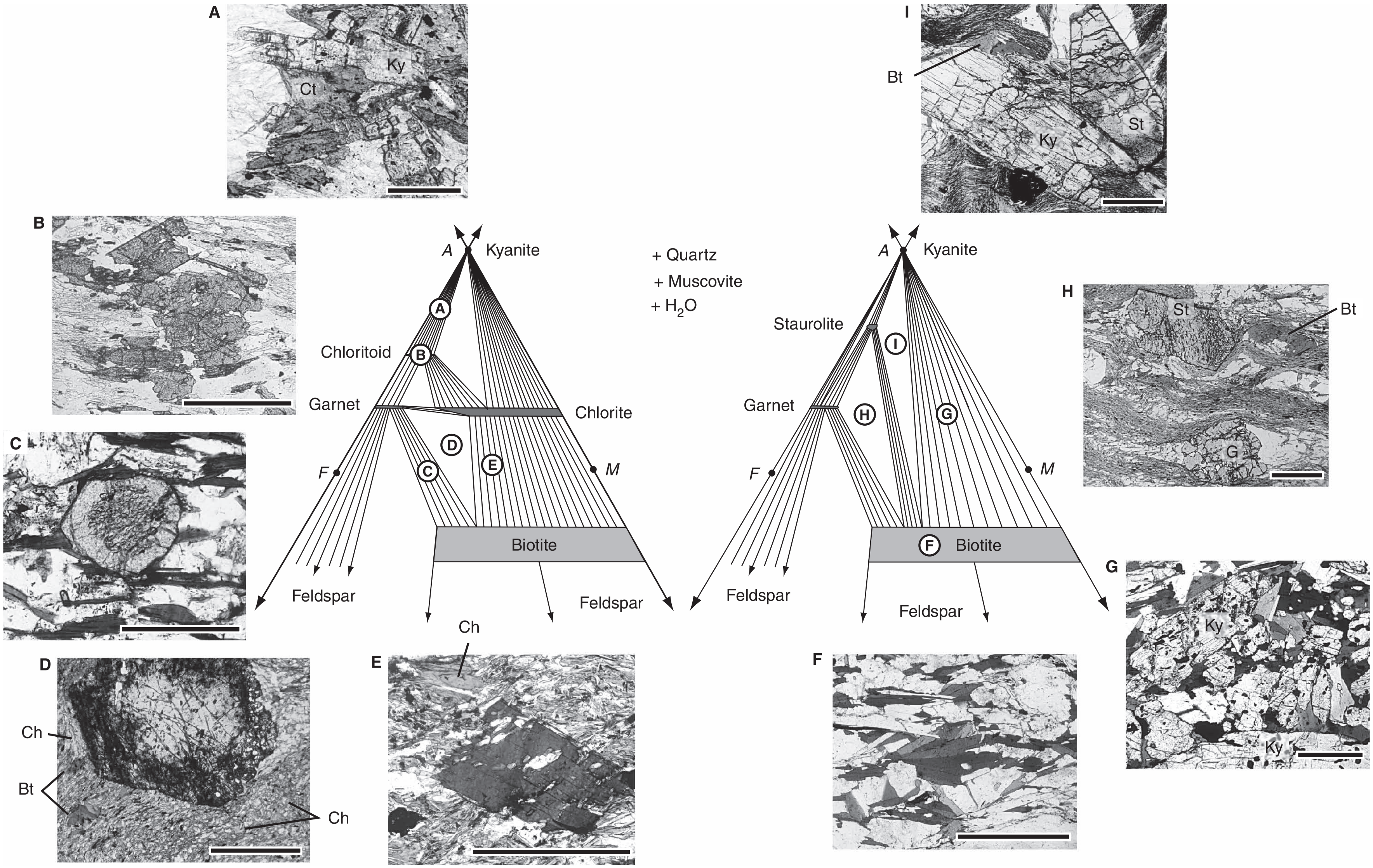


Fig. 18.8 Two representative *AFM* projections illustrated with photomicrographs of example mineral assemblages appropriate for moderate (Barrovian) pressures. Right diagram schematically shows some possible amphibolite facies mineral assemblages; left diagram is similar but is drawn for somewhat lower temperatures appropriate for the greenschist facies–amphibolite facies transition. Plane-polarized light. Scale bars = 0.5 mm for parts (B), (C), and (E); all others = 1 mm. (**A**) Chloritoid (Ct) + kyanite (Ky). (**B**) Chloritoid. (**C**) Garnet (center) + biotite (dark grains). (**D**) Garnet + biotite (Bt) + chlorite (Ch). (**E**) Biotite (dark porphyroblast) + chlorite (Ch). ( ) Biotite. (**G**) Kyanite (Ky) + biotite. Rock contains abundant Fe–Ti oxides (black). (**H**) Garnet (G) + staurolite (St) + biotite (Bt). (**I**) Staurolite (St) + kyanite (Ky) + biotite (Bt). Schematic tie-lines to K-feldspar shown only for reference; none of the rocks contains this mineral.

composition at the end of the tie line coming from the specified chlorite composition. All two-phase regions in Figure 18.8 are therefore ruled with tie lines indicating the compositions of coexisting phases. Note that if one of the two phases has a fixed composition, such as kyanite, the composition of the other phase cannot be varied independently. In this case, the bulk composition of the rock determines the composition of the other phase; that is, tie lines radiate from the phase of fixed composition through the bulk composition of the rock to the other phase.

Finally, if a rock contains only one of the minerals in an *AFM* diagram the variance is 5, that is, three environmental degrees of freedom and two compositional ones. Thus, any two compositional variables can be specified, and then the third is determined. When only one mineral is present, its composition is determined simply by the bulk composition of the rock, and it can indicate little about the metamorphic conditions, other than the rock must have crystallized in the stability field of that mineral.

The mineral assemblages in an *AFM* diagram consequently divide the diagram into three-phase triangles, two-phase regions ruled with tie lines, and one-phase regions. The three-phase assemblages have the smallest variance and therefore preserve the most information about metamorphic conditions.

18.10 ISOGRADS IN METAPELITIC ROCKS

An isograd was defined in Section 16.3 as a line on the surface of the Earth marking the first appearance of a metamorphic mineral or mineral assemblage. We will now see that such first appearances can result from either discontinuous or continuous reactions. These have important implications for field mapping, and will be illustrated with some of the reactions that produce staurolite in metapelites.

18.10.1 Discontinuous reactions

A discontinuous reaction changes the topology of a compatibility diagram. When such changes are due to the appearance or disappearance of a mineral, they are said to be *terminal* to that particular mineral. The topology can also be changed by a tie-line-switching reaction; these are *nonterminal* because the particular minerals, when separate, have stability ranges above and below that of the reaction.

Consider first the terminal discontinuous reaction

$$\text{almandine}_{\text{ss}} + \text{chlorite}_{\text{ss}} + \text{kyanite} = \text{staurolite}_{\text{ss}} + \text{quartz} + \mathrm{H_2O}$$

where the subscript ss indicates the mineral belongs to a solid solution series. Balancing such a reaction exactly requires knowledge of the Fe/Mg ratio in the various phases, which will be discussed in the next section. A *P–T* plot of this reaction is given in Figure 18.9. Because H_2O is evolved, the activity of H_2O plays an important role in determining the position of the reaction curve, which in this case is shown for $a_{\mathrm{H_2O}} = 1$. On the low-temperature side of the reaction, almandine + chlorite + kyanite form a three-phase triangle in the *AFM* diagram. Note that the composition of staurolite plots within this triangle, so that when the reaction temperature is reached, staurolite forms from these three minerals. This produces three new phase triangles depending on rock composition: almandine + staurolite + kyanite, kyanite + staurolite + chlorite, and almandine + staurolite + chlorite. For example, if the starting rock had only a small amount of chlorite relative to almandine and kyanite, the chlorite would be consumed during reaction and the product would contain almandine + staurolite + kyanite.

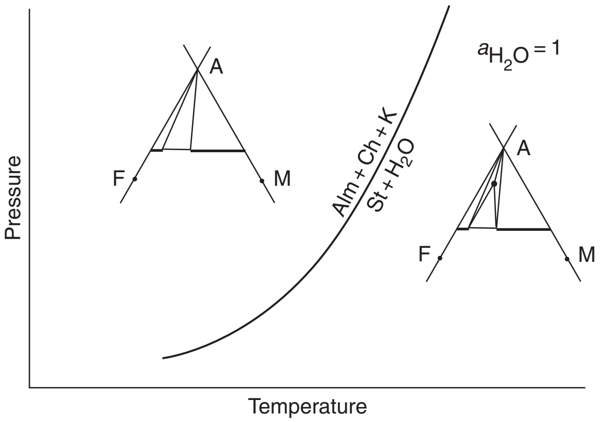


Fig. 18.9 *P–T* plot of the discontinuous reaction almandine + chlorite + kyanite = staurolite + H_2O for $a_{\mathrm{H_2O}} = 1$. This reaction is terminal to staurolite, because staurolite is not stable at temperatures below this reaction.

At a given pressure and activity of H_2O, this reaction takes place at only one temperature. The reaction marks the first true appearance of staurolite; that is, it is a terminal reaction on the low-temperature side of the staurolite stability field. Note that the reaction affects all rocks with bulk compositions lying in the triangle almandine + chlorite + kyanite. Staurolite therefore appears in a wide range of rock compositions as a result of this reaction. This is an important consideration when looking for reactions that can be mapped easily.

A second example of a terminal discontinuous reaction is the breakdown of chloritoid to produce staurolite, garnet, and chlorite during heating

$$\text{chloritoid}_{\text{ss}} + \text{quartz} = \text{garnet}_{\text{ss}} + \text{staurolite}_{\text{ss}} + \text{chlorite}_{\text{ss}} + \mathrm{H_2O} \quad (18.5)$$

As was the case for the first example, this reaction takes place at a specific temperature at a given pressure and $a_{\mathrm{H_2O}}$. When this reaction proceeds, it marks the upper temperature stability limit for chloritoid in quartz-bearing rocks.

An example of a nonterminal discontinuous reaction is

$$\text{almandine}_{\text{ss}} + \text{chlorite}_{\text{ss}} + \text{muscovite} = \text{staurolite}_{\text{ss}} + \text{biotite}_{\text{ss}} + \text{quartz} + \mathrm{H_2O} \quad (18.6)$$

which is shown in Figure 18.10. Again, because the reaction evolves H O, the must be specified. But unlike reaction

(18.5), there are two reactants that plot on the *AFM* plane (almandine and chorite), and two products that plot on the plane (staurolite and biotite). At temperatures below the reaction, staurolite is stable only in rocks having bulk compositions above the almandine–chlorite tie line. Less aluminous rocks contain almandine + chlorite + biotite. In either case, the three-phase *AFM* triangles have an almandine–chlorite tie line. Once reaction occurs, the almandine–chlorite tie line is replaced by a staurolite–biotite tie line. Thus, reaction involves a tie-line switch (hence, nonterminal), and produces two possible *AFM* assemblages depending on bulk composition: staurolite + almandine + biotite or staurolite + chlorite + biotite. Note that even rocks that originally had bulk compositions below the almandine–chlorite tie line contain staurolite once reaction occurs. The reaction therefore greatly expands the range of rock compositions that can contain staurolite. This reaction would therefore also be a convenient one to use for mapping purposes.

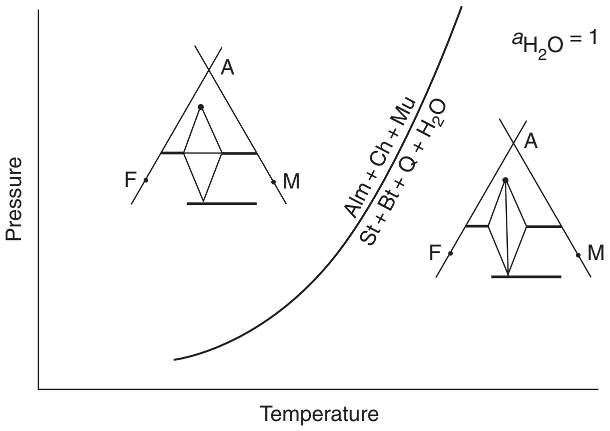


Fig. 18.10 *P–T* plot of the tie-line-switching nonterminal reaction of almandine + chlorite + muscovite = staurolite + biotite + quartz + H_2O for $a_{H_2O} = 1$.

18.10.2 Continuous reaction

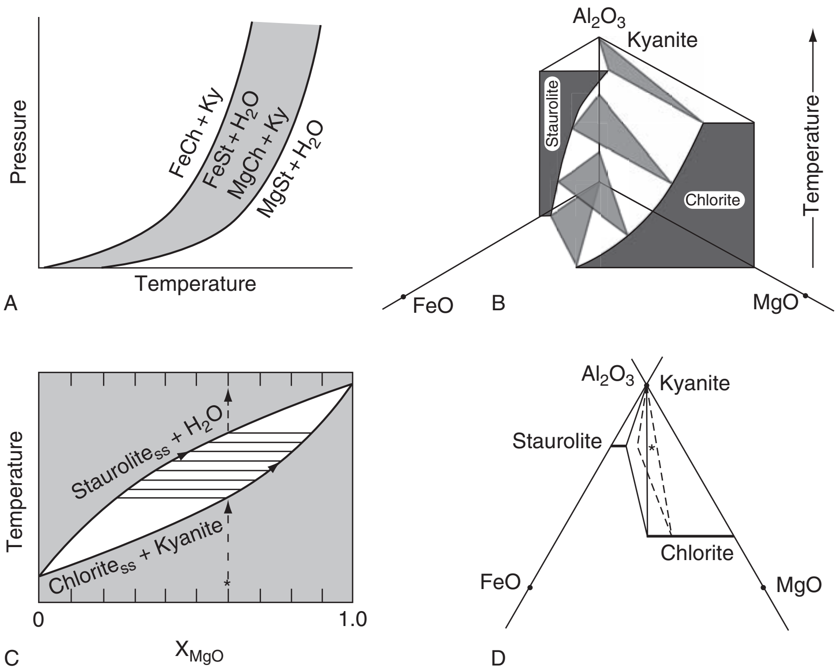
Consider the reaction

$$\text{kyanite} + \text{chlorite}_{ss} = \text{staurolite}_{ss} + \text{quartz} + H_2O \quad (18.7)$$

Both chlorite and staurolite form solid solutions, and without specifying the Mg/Fe ratio of one of the minerals, the reaction is not completely defined. As will be evident from the discussion in Sections 18.11 and 19.1, the Mg/Fe ratios of both minerals continuously change as the reaction proceeds with increasing temperature (hence, a continuous reaction). In a *P–T* diagram, the reaction is no longer a univariant line but is a broad zone (Fig. 18.11(A)). The reaction first starts with Fe-rich chlorite reacting with kyanite to form Fe-rich staurolite. As the temperature rises, the chlorite and staurolite become more magnesian. This causes the three-phase triangle kyanite + staurolite$_{ss}$ + chlorite$_{ss}$ to swing to more magnesian compositions by pivoting on its kyanite apex (Fig. 18.11(B)). These compositional changes are best seen in projections onto the T–X_{Mg} plane (Fig. 18.11(C)) or the *AFM* plane (Fig. 18.11(D)).

The temperature at which staurolite is produced by a continuous reaction depends on the bulk composition of the rock. Let us assume that the Mg/(Mg+Fe) ratio in the rock is 0.6 (* in Fig. 18.11(C) and (D)). At low temperature, this composition lies in the two-phase region kyanite + chlorite. As the temperature rises, the three-phase triangle kyanite + chlorite + staurolite swings to more magnesian compositions, and eventually the kyanite–chlorite side of the triangle

Fig. 18.11 (**A**) *P–T* plot of the continuous reaction kyanite + chlorite = staurolite + quartz + H_2O. The reaction begins at low temperatures involving the iron-rich components and progresses to more magnesian compositions with increasing temperature. (**B**) Perspective *T–AFM* plot showing how the composition of staurolite and chlorite coexisting with kyanite changes with increasing temperature. (**C**) T–X_{MgO} projection of compositions of staurolite and chlorite coexisting with kyanite, showing path followed by progressively metamorphosed rock having $X_{MgO} = 0.6$ (see text for discussion). (**D**) *AFM* plot showing swing of three-phase triangle kyanite – staurolite – chlorite to more magnesian compositions with increasing temperature.



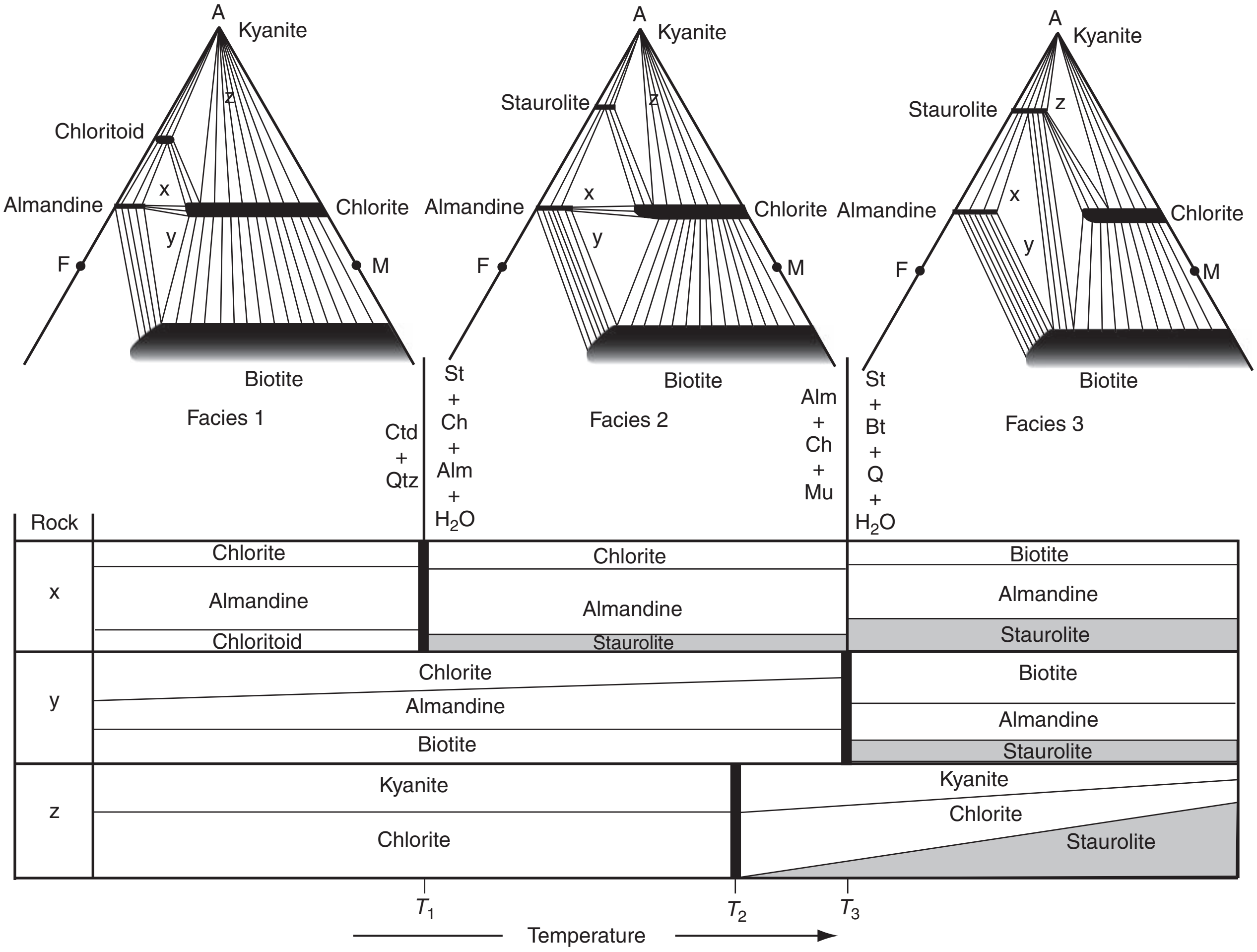


Fig. 18.12 *AFM* diagrams showing two discontinuous reactions and one continuous reaction for the formation of staurolite. The bar graphs represent schematically the modes of three different rocks, having compositions that are indicated in the *AFM* diagrams by points *x*, *y*, and *z*. The temperature of first appearance of staurolite (i.e. the staurolite isograd) is different in the three rock types, as indicated by the heavy vertical lines. See text for discussion.

passes through the bulk composition of the rock, and staurolite first appears. Only an infinitesimal amount of staurolite forms at first, which would likely go undetected in the rock, at least to the unaided eye. As the temperature rises and the three-phase triangle continues swinging to more magnesian compositions, the amount of staurolite increases and that of chlorite decreases as both minerals adjust their compositions. Eventually, the three-phase triangle swings far enough that its kyanite–staurolite side passes the bulk composition of the rock; all chlorite is consumed at this stage and the reaction is complete. The rock is then free to continue heating until the next reaction is encountered.

Let us now explore the significance of the staurolite isograd in terms of these three reactions. We will assume that in a sequence of pelitic rocks three different units have compositions indicated in the *AFM* diagrams of Figure 18.12 by points *x*, *y*, and *z*. Unit *x* plots in the three-phase triangle chloritoid + almandine + chlorite at low temperature. When the temperature reaches T_1, staurolite forms by the discontinuous chloritoid breakdown reaction (18.5), and the new assemblage is staurolite + almandine + chlorite. Staurolite makes an immediate entrance at this temperature, with its abundance determined by the position of the bulk composition of the rock within the staurolite–almandine–chlorite triangle. In those rocks sufficiently aluminous and Fe-rich to have originally contained chloritoid, the product staurolite would be an easily recognized phase. This reaction would make a very useful isograd, marking the trace of the T_1 isotherm at a given pressure and activity of water.

With rising temperature, rock *x* eventually reaches T_3 where the second discontinuous reaction takes place (Eq. (18.6)). The switch in tie lines changes the assemblage of minerals to staurolite + almandine + biotite above this temperature. Biotite makes an immediate entrance at this temperature, and chlorite abruptly disappears. The amount of staurolite also slightly increases. This reaction could be used to map a biotite isograd.

Rock *y*, being less aluminous than rock *x*, consists of the assemblage almandine + chlorite + biotite at low temperatures. Staurolite does not appear in this rock at T_1 because it lacks the appropriate reactant minerals. Not until the tie-line switch at T does staurolite appear, and then it does so abruptly, which would make this a mappable reaction. But note that this appearance of staurolite is different from that mapped by the reaction at T_1.

Rock *z* contains kyanite and chlorite at low temperature, and not until the three-phase triangle swings to more magnesian compositions does staurolite make its first appearance at T_2, but even then the amount is very small at first, increasing only with rising temperature. An isograd based on this continuous reaction would therefore be difficult to detect. The compositional range encompassed by the three-phase triangle is not large and therefore the range of rocks in which this reaction is likely to take place is small. Moreover, the temperature T_2 of first appearance of staurolite depends on the bulk composition of the rock. Thus, even if the first appearance of staurolite is detectable, its presence cannot be used as an indicator of temperature unless the bulk composition of the rock is known. This reaction is therefore a less practical one on which to base an isograd.

From the reactions above, it is clear that if an isograd is based simply on the first appearance of an index mineral, it will probably not be an accurate indicator of temperature, even if the pressure and a_{H_2O} are known, because the minerals can form by numerous reactions. The staurolite isograd defined in this way would occur at different temperatures in rocks *x*, *y*, and *z*, because of their different compositions (heavy vertical lines in lower part of Fig. 18.12). To avoid this compositional problem, the isograd should be based on a discontinuous reaction. These reactions generally affect a large range of rock composition. Also, an isograd should be based on the first appearance of a *mineral assemblage* rather than the first appearance of a *single mineral*. Thus in the examples above, two separate staurolite isograds could be mapped, one involving staurolite + almandine + chlorite, and the other staurolite + almandine + biotite. Clearly, if these two reactions had been lumped together and mapped as the first appearance of staurolite, the isograd would not necessarily indicate a line of equal grade of metamorphism.

18.11 EFFECT OF *P* AND *T* ON REACTIONS AMONG SOLID SOLUTION PHASES

Reactions involving solid solution phases can be of the continuous or discontinuous type. Because so many metamorphic minerals form solid solutions, in particular those in metapelitic rocks, it is important to have an understanding of the effects of temperature and pressure on reactions involving these minerals. For a complete discussion of this topic, see the papers by A. B. Thompson (1976a,b).

The ratio Mg/Mg + Fe (X_{Mg}) in coexisting minerals in metapelitic rocks decreases in the order cordierite > chlorite > biotite > chloritoid > staurolite > garnet. For example, in Figure 18.11(C) we saw that chlorite is more magnesian than coexisting staurolite. Knowing the relative distribution of Fe and Mg between these minerals allows us to deduce what reactions can take place among the minerals simply from their positions in the *AFM* plot. The direction in which these reactions proceed with increasing temperature can then be determined by noting on which side of the reaction H_2O is released (high-entropy side = high-temperature side).

In Section 10.9, the substitution of Mg for Fe in fayalite was seen to raise the melting point of olivine, and conversely, the substitution of Fe for Mg in forsterite lowered the melting point (Fig. 10.11). Although this is not a metamorphic reaction, the derivation of the cryoscopic equation, which quantified these effects, was perfectly general and is applicable to solid as well as liquid solutions. For example, in Section 10.10, polymorphic changes among the pyroxenes were seen to follow the same type of relation. We can generalize the observations by stating that the temperature of a reaction taking place between pure end-members is raised by a component that is more soluble in the low-temperature phase or phases and is lowered by a component that is more soluble in the high-temperature phase or phases. Magnesium, for example, is more soluble in olivine$_{ss}$ than in olivine liquid, so the melting point of fayalite rises with substitution of Mg. Similarly, the temperatures of metamorphic reactions are raised if Mg substitutes more readily into the phases on the low-temperature side of the reaction and lowered if Mg is more soluble in the high-temperature phases. Knowing the relative distribution of Fe and Mg between the minerals of pelitic rocks allows us, then, to predict the T–X_{Mg} relation of reactions involving these minerals. The effect of pressure on the reactions can be deduced if, in addition to knowing the Fe–Mg partitioning among phases, the ΔVs of the end-member reactions are known.

Let us consider the effects of temperature and pressure on a few specific metamorphic reactions. This will be done using pseudobinary T–X_{Mg} and P–X_{Mg} diagrams with compositions of minerals in the *AFM* plot being projected from the *A* apex of the diagram. We have already seen in Figure 18.11(C) the effect of Mg substitution for Fe on the continuous reaction of chlorite + kyanite to produce staurolite + quartz + H_2O (Eq. (18.7)). Staurolite must be on the high-temperature side of this reaction, and because Mg is more soluble in chlorite than in staurolite, the reaction temperature rises as X_{Mg} increases (shown also in Fig. 18.13(A)). Also, as X_{Mg} increases, so does the pressure at which the reaction takes place (Fig. 18.13(B)).

Consider next the reaction

$$\text{staurolite} + \text{muscovite} + \text{quartz} = \text{biotite} + \text{kyanite} + H_2O \tag{18.8}$$

In this reaction, staurolite appears on the low-temperature side, but because Mg is more soluble in biotite than in coexisting staurolite, the reaction temperature decreases with increasing X_{Mg} (Fig. 18.13(C)). Increasing X_{Mg}, however, raises the pressure at which the reaction takes place (Fig. 18.13(D)).

If reactions (18.7) and (18.8) are plotted in the same T–X_{Mg} diagram (Fig. 18.13(E)), they intersect to create an isobaric invariant assemblage of the four phases chlorite + kyanite + staurolite + biotite. From the modified phase rule ($\phi + f \quad + 3$), we have $4 + f = 3 + 3$, and thus there are two

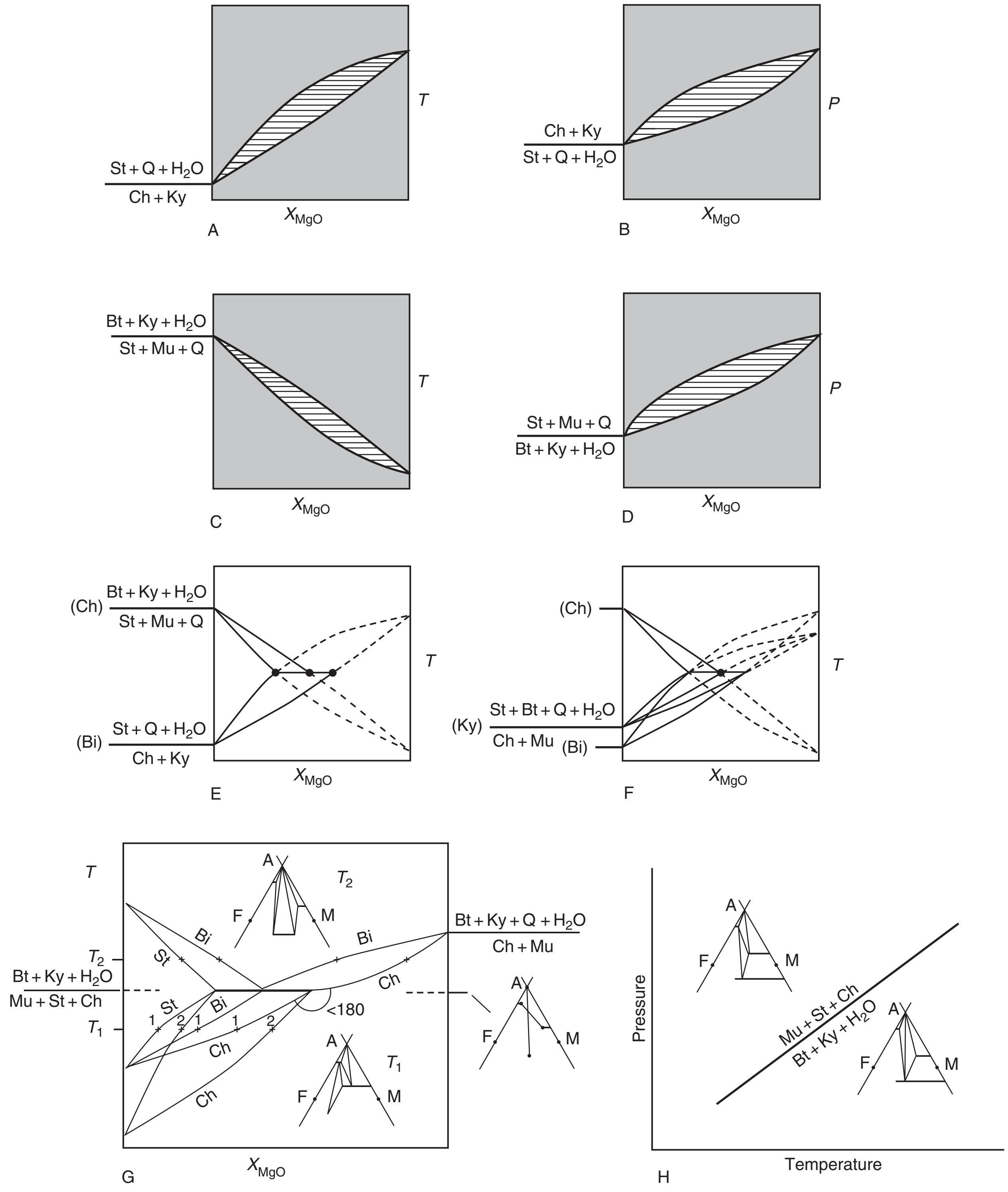


Fig. 18.13 (**A**) Isobaric T–X_{MgO} diagram for the continuous reaction chlorite + kyanite = staurolite + quartz + H_2O. (**B**) Same reaction as in part (A) plotted in P–X_{MgO} diagram. (**C**) Isobaric T–X_{MgO} diagram for the continuous reaction staurolite + muscovite + quartz = biotite + kyanite + H_2O. (**D**) Same reaction as in part (C) plotted in P–X_{MgO} diagram. (**E**) Reactions in parts (A) and (C) combine to give an isobaric invariant assemblage where they intersect involving muscovite + staurolite + chlorite = biotite + kyanite + H_2O. Each reaction is made metastable on the magnesian side of the invariant point by the other reaction. (**F**) Extending from the invariant point is a reaction involving chlorite + muscovite = staurolite + biotite + quartz + H_2O. (**G**) A fourth and final reaction extending from the invariant point involves chlorite + muscovite = biotite + kyanite + quartz + H_2O. At the isobaric invariant point, the compositions of staurolite, chlorite, and biotite are fixed, whereas above and below this temperature the compositions vary continuously along the isobaric univariant continuous reaction curves. (**H**) P–T diagram for the reaction muscovite + staurolite + chlorite = biotite + kyanite + H_2O. See text for discussion.

degrees of freedom; these are required to define the pressure and a_{H_2O} of the T–X_{Mg} diagram.

Four independent reactions must extend from this invariant point. Each one involves three of the four minerals present at the invariant point. Two of these are already known; the chlorite-absent reaction is Eq. (18.8) and the biotite-absent reaction is Eq. (18.7). The kyanite-absent and staurolite-absent reactions can be deduced from the positions of the phases in the *AFM* diagram. They are:

$$[\mathrm{Ky}]\quad \mathrm{chlorite_{ss} + muscovite = staurolite_{ss} + biotite_{ss} + quartz} + H_2O$$

$$[\mathrm{St}]\quad \mathrm{chlorite_{ss} + muscovite = biotite_{ss} + kyanite + quartz} + H_2O$$

The temperature of the kyanite-absent reaction must increase with increasing X_{Mg} because Mg is more soluble in chlorite than in coexisting staurolite, and chlorite appears on the low-temperature side of the reaction. Also, at the invariant point, the reaction involves the coexistence of three solid solution phases, chlorite, staurolite, and biotite, the compositions of which are defined by the points of intersection of the other two reaction lines. Three lines must extend from these points to indicate the projected compositions of the minerals forming the three-phase triangle in the *AFM* diagram. Because the kyanite-absent reaction involves the coexistence of staurolite and quartz, this reaction must occur in the stability field of staurolite + quartz; that is, the reaction must occur above the biotite-absent reaction because it is only there that staurolite + quartz is stable (Fig. 18.13(F)).

The staurolite-absent reaction involves the coexistence of the solid phases chlorite and biotite, the compositions of which are determined by the temperature at which the chlorite- and biotite-absent reactions intersect. Because Mg partitions into chlorite more than into biotite and chlorite is on the low-temperature side of the reaction, increasing X_{Mg} causes the temperature of the reaction to rise. However, because the assemblage chlorite + muscovite occurs on the low-temperature side of both the kyanite- and staurolite-absent reactions, the slope of the staurolite-absent reaction must be less than that of the kyanite-absent reaction (Fig. 18.13(G)) so as not to violate the Schreinemakers rule stating that *the angle between univariant lines bounding a divariant field at an invariant point must be less than 180°* (Section 8.4).

The point of intersection of these four isobaric, univariant, continuous reactions is an isobaric, invariant, discontinuous reaction involving

$$\mathrm{staurolite + chlorite + muscovite = biotite + kyanite + quartz} + H_2O \qquad (18.9)$$

At the isobaric invariant point, the compositions of staurolite, chlorite, and biotite are fixed; above or below this temperature, however, the compositions of these minerals vary along the lines representing the projections of the various three-phase triangles. At temperature T_1 (Fig. 18.13(G)), the compositions in the two three-phase triangles (St + Bi + Ch and St + Ch + K) can be read off the T–X_{Mg} diagram and plotted on an *AFM* diagram (inset in Fig. 18.13(G)). As the temperature rises, the compositions of staurolites 1 and 2 and chlorites 1 and 2 approach each other and eventually become the same at the invariant point where the tie-line-switching reaction of staurolite + chlorite = kyanite + biotite takes place. Above the invariant point, at temperature T_2 for example, we can again read off the T–X_{Mg} diagram the composition of the phases coexisting in the three-phase triangles and plot them in an *AFM* diagram (Fig. 18.13(G)).

Although the temperature and compositions of the phases are fixed at the isobaric invariant point, under other pressures or a_{H_2O} different values would be established. A diagram similar to Figure 18.13(G) can be constructed for the intersection of the four reactions in terms of P–X_{Mg} at constant T and a_{H_2O} (Thompson, 1976a). Finally, the reaction of Eq. (18.9) can be projected into P–T space to become the univariant reaction of Figure 18.13(H).

18.12 PETROGENETIC GRID FOR METAPELITIC ROCKS

Having discussed the types of reactions that can occur amongst minerals of solid solution series, we are in a position to make some comments on how a petrogenetic grid for metapelitic rocks could be constructed. A petrogenetic grid is based on discontinuous reactions because, unlike continuous reactions, these are functions only of pressure, temperature, and a_{H_2O}. Here, we focus on reactions involving the seven common minerals in the *AFM* diagram (Al_2SiO_5, staurolite, chloritoid, cordierite, garnet, chlorite, and biotite). A discontinuous reaction, such as that of Eq. (18.6), involves four minerals in the *AFM* diagram (+ quartz + muscovite + H_2O). We assume that $a_{H_2O} = 1$, which is probably reasonable for thick successions of metapelitic rocks that lack metacarbonate layers and graphite. Construction of a full grid is beyond the scope of this book, but we will examine some of the principles involved.

For purposes of illustration, consider an invariant point on the grid. If we start with the univariant discontinuous reaction of Eq. (18.6) and add to it the mineral chloritoid (Ct), we create an invariant assemblage for a given a_{H_2O}. Experiments and thermodynamic calculations indicate that this invariant point is at about 580 °C and 0.95 GPa (Fig. 18.14). Five univariant reactions radiate from the invariant point, one for each group of four minerals that can be formed from the five minerals at the invariant point. These reactions can be deduced from the position of minerals in the *AFM* diagram. We label them according to the absent phase.

[Ct] $\mathrm{G + Ch + Ms = Bt + St + Qtz} + H_2O$
[St] $\mathrm{Ct + Bt + Qtz} + H_2O = \mathrm{G + Ch + Ms}$
[Ch] $\mathrm{Ct + Ms + Qtz = Bt + G + St} + H_2O$
[Bt] $\mathrm{Ct + Qtz = G + St + Ch} + H_2O$
[G] $\mathrm{Ch + Ct + Ms = Bt + St + Qtz} + H_2O$

In each of these reactions quartz, muscovite, and water are available for balancing.

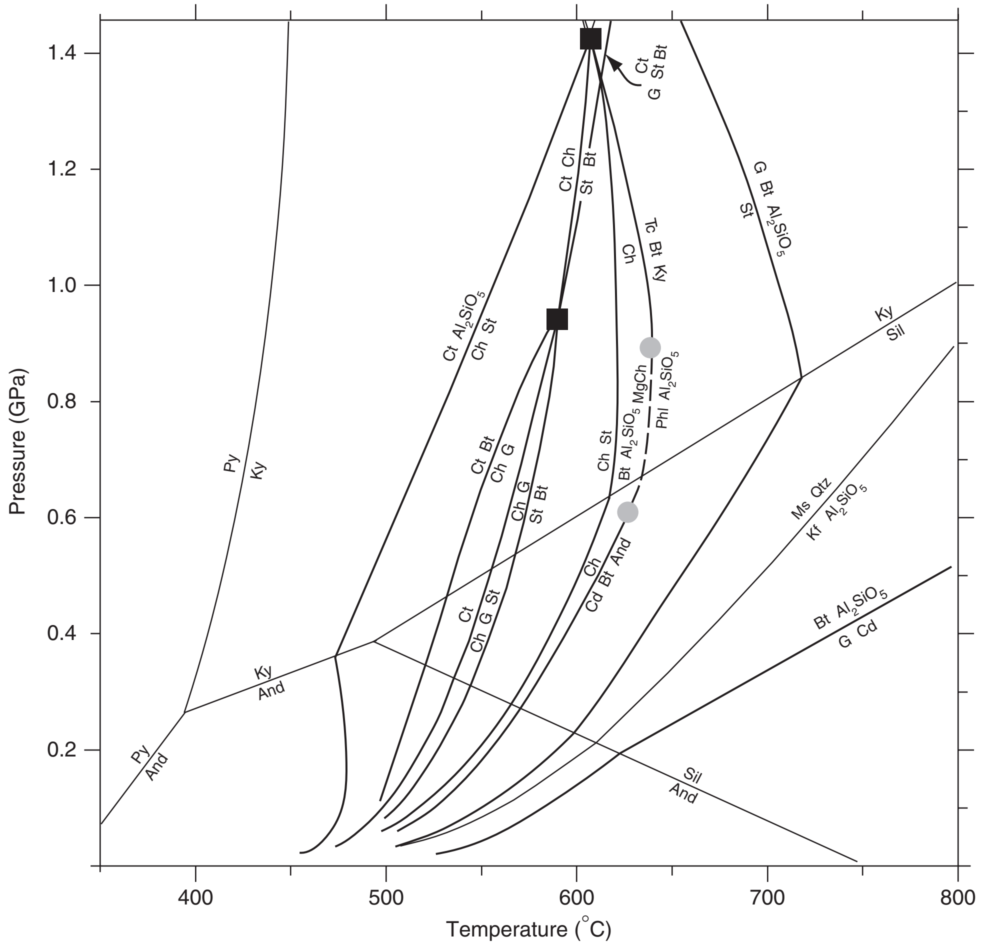


Fig. 18.14 Petrogenetic grid in the system SiO_2–Al_2O_3–FeO–MgO–K_2O–H_2O (KFMASH) for metapelitic rocks (after Spear and Cheney, 1989). The seven common metapelitic phases chlorite (Ch), chloritoid (Ct), staurolite (St), garnet (G), biotite (Bt), cordierite (Cd), and Al_2SiO_5 are considered; talc (Tc) is included as well. Furthermore, some reactions in the system SiO_2–Al_2O_3–K_2O–H_2O involving muscovite, pyrophyllite (Py), quartz (Qtz), K-feldspar (Kf), kyanite (Ky), andalusite (And), or sillimanite (Sil) are also shown. KFMASH invariant points denoted with solid squares (see Problem 18.10). The reactions involving chlorite breakdown to either Tc + Bt + Ky or Cd + Bt + And terminate at invariant points (gray circles) in the Fe-free system (KMASH). The reaction involving Mg-chlorite (MgCh), phlogopite (Phl), and Al_2SiO_5 in KMASH is shown (dotted line), but the other reactions radiating from the KMASH invariant points are omitted for clarity. Note that the two discontinuous staurolite-producing reactions of Figure 18.12 are present on this grid.

We already know the position of the chloritoid-absent reaction from a plot like Figure 18.10. The slopes of the reactions can be calculated from the Clapeyron equation and the ΔVs and ΔSs of the reactions. All the reactions involve water and have positive slopes (although the St-absent reaction is unusual in that water is on the low-temperature side). Using Schreinemakers rules, we then arrange the univariant lines around the invariant point (Fig. 18.14). The reader should check that the positions of these reactions do indeed agree with Schreinemakers rules.

On extending to higher pressures, the garnet-absent reaction intersects the reaction $Ct + Al_2SiO_5 = St + Ch + Qtz + H_2O$ to create a new invariant point involving chloritoid, chlorite, staurolite, biotite, and Al_2SiO_5 at about 605 °C and 1.43 GPa (Fig. 18.14). Again, thermodynamic data and Schreinemakers rules are used to arrange the reactions around the invariant point. Some of the univariant lines radiating from this second invariant point will, on being extended, intersect other univariant lines radiating from the first invariant point. For the system shown in Figure 18.14, this occurs at pressures in excess of 1.5 GPa and is not shown on the diagram. A new invariant point is created at each point of intersection, and five univariant lines will radiate around each of these, unless degeneracy occurs. These lead to new intersections, and P–T space becomes filled with a petrogenetic grid.

It turns out that construction of a full grid quickly becomes complicated because of the large number of reactions involved. To determine how many discontinuous reactions can be written between the seven *AFM* minerals, we need to know how many different combinations of minerals, taken four at a time, can be formed regardless of the order in the combination. According to the *rule of combinations*, the number of combinations of n different items taken r at a time is $n!/r!(n-r)!$. Thus, for the seven minerals, the number of combinations of four are $7!/4!(7-4)! = 35$. This, however, treats Al_2SiO_5 as a single mineral. The number of reactions is therefore much greater, for all those involving Al_2SiO_5 must be written as separate reactions for the different Al SiO polymorphs. At high temperatures, muscovite reacts with quartz, and K-feldspar becomes the common phase from which *AFM* projections are made. Therefore, reactions involving muscovite at low temperatures must be written as different reactions involving K-feldspar at high temperature. Clearly the number of reactions that must be considered in a petrogenetic grid for metapelitic rocks is large; a number of authors have constructed such grids, including Albee (1965), P. Hess (1969), Kepezhinskas and Khlestov (1977), and Harte and Hudson (1979). We illustrate the grid of Spear and Cheney (1989) in Figure 18.14, which was calculated using thermodynamic data for mineral end members as well as activity models that describe the energetics of the Fe–Mg solid solutions.

A petrogenetic grid can be a useful tool with which to analyze metamorphic mineral assemblages, because it should include all possible mineral assemblages if it has been constructed properly. It must, however, be used with care. The grid can be applicable to only one activity of H_2O, which in Figure 18.14 is $a_{H_2O} = 1$. If the activity of H_2O varies throughout a metamorphic region, as might happen if metacarbonate rocks were interlayered with metapelites, the simple petrogenetic grid could not be used. Other components, in particular Mn and Ca, can also cause changes in the grid (Spear and Cheney, 1989). Apart from these problems, the grid provides a useful means of explaining the sequence of successive mineral assemblages in progressively metamorphosed rocks, and it also provides a means of determining the approximate temperatures and pressures represented by particular mineral assemblages.

18.13 APPLICATION: REGIONAL PRESSURE ESTIMATION

To conclude, we will examine how some of the principles explored in this chapter can be used to estimate pressures in a real regional metamorphic mountain belt. To begin, consider the Al_2SiO_5 triple point. The location of this invariant point can be traced through a terrane by mapping the boundary between regions in which andalusite occurs and those in which kyanite transforms directly to sillimanite during progressive metamorphism. If the Al_2SiO_5 polymorphs are pure, this boundary corresponds to an isobar with a pressure of about 0.4 GPa. Carmichael (1978) defines such a line as a bathograd; that is, a *bathograd* is a mapped line that separates occurrences of a higher-pressure assemblage from occurrences of a lower-pressure assemblage.

Carmichael uses the intersection of two dehydration reactions with the univariant lines marking the polymorphic transformations of Al_2SiO_5 to define four additional invariant assemblages that can be used to map bathograds. The dehydration reactions involve the disappearance of staurolite and muscovite from quartz-bearing metapelites. The lower-temperature terminal reaction is

$$\text{quartz} + \text{muscovite} + \text{staurolite} = \text{biotite} + \text{garnet} + Al_2SiO_5 + H_2O$$

Where this reaction crosses the andalusite–sillimanite transformation line (Fig. 18.15(A)), both polymorphs are

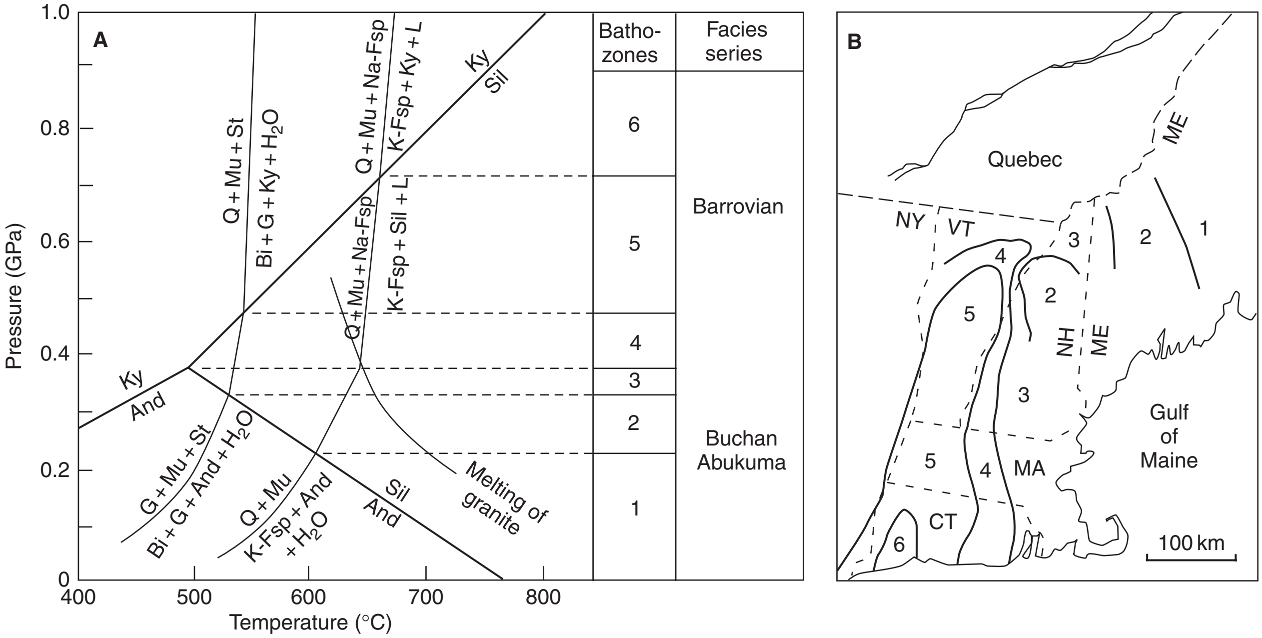


Fig. 18.15 Intersection of two dehydration reaction curves with the Al_2SiO_5 polymorphic transformation lines and the Al_2SiO_5 triple point itself provide a convenient means of distinguishing six pressure (depth) zones, which are referred to as bathozones. The boundaries between the bathozones, which are termed bathograds, are shown for New England. They show that the greatest uplift has occurred in western Connecticut and Massachusetts and in Vermont. (After Carmichael, 1978; published by permission of the *American Journal of Science*.)

stable, creating an invariant assemblage. Mapping the occurrence of this assemblage, or what is more likely in practice, mapping the boundary between regions that have the higher-pressure assemblage from those with the lower-pressure assemblage, delineates a bathograd whose pressure is approximately 0.33 GPa. Another invariant point exists at about 0.48 GPa, where the same dehydration reaction crosses the kyanite–sillimanite transition line. More recent petrogenetic grids place this staurolite breakdown reaction at somewhat higher temperatures, but the general bathograd concept remains valid (Pattison, 2001; Ghent and Simony, 2005).

The higher-temperature terminal reaction is

$$\text{quartz} + \text{muscovite} = \text{K-feldspar} + Al_2SiO_5 + H_2O$$

At low pressures, Al_2SiO_5 forms the polymorph andalusite, whereas at higher pressures it forms sillimanite. The bathograd marking the boundary between these two zones corresponds to a pressure of ~0.22 GPa. At higher pressures still, this reaction would intersect the kyanite–sillimanite transformation line, but the reaction is made metastable above a pressure of approximately 0.375 GPa by the beginning of melting of granite. Thus, above this pressure quartz + muscovite + Na-feldspar breaks down to form K-feldspar + Al_2SiO_5 + water-saturated granite melt. This reaction intersects the kyanite–sillimanite transition line at about 0.72 GPa (Fig. 18.15(A)).

Clearly, the position of all these points of intersection depends on the a_{H_2O}, which is assumed to be one in Figure 18.15 (A). The a_{H_2O} must be lowered considerably to cause a significant lowering of reaction temperatures. Because the activity of water in metapelitic rocks is believed to be moderately high in all but the highest-grade metamorphic terranes, the diagram for a_{H_2O} should provide reasonably accurate pressures.

The four points of intersection of the dehydration reactions with the Al_2SiO_5 transformation lines and the Al_2SiO_5 triple point itself create five boundaries that can be used to map bathograds. The five boundaries divide *P–T* space into six zones, which Carmichael refers to as *bathozones* (Fig. 18.15 (A)). Barrovian metamorphism falls in zones 5 and 6, whereas Buchan metamorphism corresponds to zone 2. The Abukuma type is a very low-pressure style of regional metamorphism that Miyashiro (1961) recognized in the central part of the Abukuma Plateau of Japan.

Figure 18.15(B) shows the distribution of the bathozones in the northern Appalachians of New England. Regions falling in the deepest bathozones must have undergone the greatest amount of erosion and exhumation since the time of metamorphism. The distribution of the bathograds in New England resembles a northerly plunging antiform extending through western Connecticut and Massachusetts and central Vermont, with the greatest unroofing occurring in Connecticut. This region corresponds closely to a large gravity high and is approximately along the suture between North America and the Avalon Terrane.

18.14 PROBLEMS

18.1 Metamorphic rocks at two localities have the following mineral assemblages:

Locality A:	quartz + enstatite + phlogopite
	quartz + phlogopite + orthoclase
Locality B:	quartz + enstatite + orthoclase
	enstatite + phlogopite + orthoclase

The metamorphic facies concept can be used to determine whether metamorphic conditions at these two localities were the same or different. Begin by constructing mineral facies diagrams for the two localities. Although five oxides (SiO_2, MgO, $KO_{1/2}$, $AlO_{3/2}$, $HO_{1/2}$) are needed to describe the compositions of all the phases, careful grouping of the oxides and treatment of water as an environmental variable allows the number of components to be reduced to three.

- **(a)** List three possible sets of three components that can be used to describe the minerals from the two localities on an anhydrous basis (water is treated as an environmental variable rather than as a component).
- **(b)** Construct a triangular mineral facies diagram for each locality, using one set of the components listed in part (a). Could these assemblages belong to the same facies? Explain your answer.
- **(c)** Write a balanced tie-line switching reaction that relates the mineral facies of the two localities.
- **(d)** Which locality would have experienced the highest temperature during metamorphism, and why?

18.2 In *ACF* diagrams only assemblages containing quartz are plotted. This allows SiO_2 to be omitted from the plot. Because quartz is present in all assemblages, it is available as a reactant or product to balance tie-line switching reactions relating *ACF* diagrams for adjacent mineral facies. To illustrate this, write balanced reactions relating the following low- and medium-pressure subfacies of the granulite facies. Start by drawing an *ACF* diagram for each subfacies.

Low-*P*:	anorthite + diopside + enstatite + quartz
	anorthite + enstatite + pyrope + quartz
Medium-*P*:	anorthite + diopside + pyrope + quartz
	diopside + enstatite + pyrope + quartz

18.3 How many reactions are required to convert the *ACF* diagram for the greenschist facies to that of the amphibolite facies in Figure 18.6?

18.4 At high metamorphic grades, muscovite is destroyed and K-feldspar is produced. *AFM* projections for these rocks must, therefore, use K-feldspar instead of muscovite.

- **(a)** Derive a formula for calculating the alumina index of compositions in the Al_2O_3–FeO–MgO–K_2O tetrahedron when projected onto the *AFM* plane from K-feldspar.

(b) What is the alumina index of biotite in the *AFM* plane when projected from K-feldspar? Use the biotite formula given in Table 18.2.

18.5 A suite of metagabbros and metaperidotites contain the following mineral assemblages:

Fo + En + An + Gt
Fo + En + An + Di
Fo + An + Gt + Sp
Fo + An + Di + Sp
Fo + En + Gt + Sp

where the abbreviations refer to

Forsterite	(Fo)	Mg_2SiO_4
Enstatite	(En)	$MgSiO_3$
Anorthite	(An)	$CaAl_2Si_2O_8$
Diopside	(Di)	$CaMgSi_2O_6$
Garnet	(Gt)	$CaMg_5Al_4Si_6O_{24}$
Spinel	(Sp)	$MgAl_2O_4$

The problem is to construct a triangular mineral facies diagram by projecting points in the CaO–MgO–$AlO_{3/2}$–SiO_2 tetrahedron onto the CaO–$AlO_{3/2}$–SiO_2 face from the phase common to all assemblages.

(a) First, plot all minerals in the tetrahedron, and join coexisting phases with tie lines. Because olivine is common to all these assemblages, the composition of the other minerals can be projected onto the CaO–$AlO_{3/2}$–SiO_2 face from the composition of olivine, thus reducing the graphical representation of the assemblages to a triangle. Draw, as best you can, the approximate positions of the projected points of the minerals. Note that the projection of spinel falls outside the triangle.

(b) Although the projected compositions of minerals can qualitatively be plotted graphically, it is necessary to have a means of doing this more accurately. We will now derive a general equation for the projection of any point within a tetrahedron onto a desired face from a given point. The composition of the point from which the projection is to be made can be expressed in terms of the four components *A*, *B*, *C*, and *D* as

$$\text{projection point}: a_pA + b_pB + c_pC + d_pD$$

where the lowercase letters are the mole fractions of each component. The composition of the point to be projected can similarly be defined as

$$\text{point } X: a_xA + b_xB + c_xC + d_xD$$

If point *X* is to be projected onto the *BCD* face, all *A* component must be removed from composition *X* by subtracting an appropriate amount, *n*, of the projection composition; that is, $a_xA - na_pA = 0$, in which case $n = a_x/a_p$. Now derive the general formula for the composition of the projected point in terms of *B C*, and *D*

(c) Using the formula from part (b), plot all five mineral assemblages on the triangular projection, joining coexisting phases with tie lines. Care must be taken with the projection of spinel. The general formula is still valid; the problem is to interpret correctly the significance of a negative coefficient.

18.6 This exercise examines the effect of temperature on reactions involving *AFM* phases.

(a) Draw a Thompson *AFM* diagram showing the three-phase tie triangles and two-phase tie lines for the following assemblages (+ muscovite + quartz)

staurolite + biotite + almandine
staurolite + biotite + kyanite
biotite + kyanite + cordierite

(b) Write a qualitative exchange reaction (no need to balance) relating staurolite–biotite–kyanite (+ muscovite + quartz). Determine which way this reaction would proceed with increasing temperature by noting on which side H_2O is liberated.

(c) Draw a schematic isobaric T–X_{Mg} diagram showing variation in the composition of staurolite and biotite Fe–Mg solid solutions in the assemblage staurolite + biotite + kyanite as a function of rising temperature. Discuss the variation in composition and modal proportions of staurolite and biotite as a function of *T* using this diagram.

(d) Show how the compositions of staurolite and biotite coexisting with kyanite change with rising temperature by showing the shift of the staurolite–biotite–kyanite tie-triangle with temperature.

18.7 The purpose of this problem is to see how several continuous reactions can form a discontinuous reaction that terminates the stability field of staurolite at high temperatures.

(a) Start by drawing two schematic T–X_{Mg} diagrams for the following continuous reactions:

$$\text{garnet} + \text{muscovite} = \text{biotite} + Al_2SiO_5 + H_2O$$
$$\text{staurolite} + \text{biotite} = \text{garnet} + \text{muscovite} + H_2O$$

Note that the second reaction requires three lines in the T–X_{Mg} diagram to show the compositions of the three coexisting solid solution phases.

(b) These two reactions intersect to create an isobaric invariant assemblage for a given a_{H_2O}. Draw the composition loops for the two reactions in the same T–X_{Mg} diagram, and then determine which parts of these loops must be metastable. Deduce what other two continuous reactions must intersect at this invariant point.

(c) Complete the T–X_{Mg} diagram by placing the other two reactions with appropriate relative slopes. Check that your construction does not violate Schreinemakers rules.

(d) Finally, select arbitrary temperatures above and below the invariant temperature, and show in *AFM* diagrams the compositions of the phases at these temperatures. What is the maximum temperature to which staurolite is stable in this isobaric diagram?

18.8 Metamorphism of siliceous dolomite can produce a variety of minerals, including tremolite, diopside, forsterite, calcite, and quartz. These minerals can be expressed in terms of the three components CaO–MgO–SiO_2 if the composition of the fluid is controlled by the environment. How many phases are needed to create an isobaric invariant assemblage in such a case? How many reactions can be written involving these minerals? Would you expect degenerate reactions to occur with these minerals?

18.9 From the ACF diagrams in Figure 18.6, list the major mineral assemblages formed in an impure limestone containing small amounts of SiO_2, MgO, and Al_2O_3.

18.10 **(a)** Two univariant reactions radiate from the high-pressure invariant point at about 1.43 GPa and 605 °C on Figure 18.14. They extend to pressures above the top of the graph. Both reactions result in chloritoid breakdown with increasing temperature. What are these two reactions?

(b) Metamorphic mineral assemblages that contain garnet + biotite are common in nature. Construct a simplified version of Spear and Cheney's (1989) petrogenetic grid (Fig. 18.14) appropriate for those bulk-rock compositions that produce *AFM* mineral assemblages containing garnet + biotite, and draw schematic *AFM* diagrams for each divariant field.

19 Geothermometry, geobarometry, and mineral reactions among solid solutions

19.1 INTRODUCTION

Petrogenetic grids provide a means of setting limits on the temperatures and pressures under which a given mineral assemblage can form, but these limits are commonly rather wide. A rock containing the divariant assemblage garnet + biotite + chlorite could be stable over a temperature range of nearly 100 °C and a pressure range in excess of 0.8 GPa (Fig. 18.14). Many of the minerals constituting these assemblages are solid solutions, and their compositions may vary considerably across a divariant field. We can use these variations to our advantage, however, because the chemical compositions of coexisting minerals are functions of temperature and pressure.

If we can determine how Fe, Mg, and other elements partition themselves between coexisting minerals as pressure and temperature vary, then precise metamorphic temperatures and pressures can be determined from analyses of the minerals. So to extract pressure–temperature information from rocks, the first step is to measure the chemical compositions of the minerals. This is usually done using the electron microprobe, which offers micrometer-scale spatial resolution and the ability to determine major, minor, and many trace elements. Then, the positions of reactions among the minerals in P–T space are calculated using thermodynamic data sets and activity models derived from laboratory experiments and, in some cases, natural mineral assemblages. As we saw earlier in Chapter 9, reactions that are useful for estimating temperatures are called *geothermometers*, whereas those useful for estimating pressures are called *geobarometers*. The science of estimating temperatures and pressures for rocks is called *thermobarometry.* In addition to chemical variations, the partitioning of oxygen and other stable isotopes between minerals may vary significantly with temperature and can, thus, be used for quantitative geothermometry.

Petrologists continue to devote considerable effort to the development of geothermometers and geobarometers, because estimates of pressure and temperature are critical for answering a broad range of geologic questions. Determining metamorphic fluid compositions, rates and processes of heat transfer through the crust, and rates of tectonic burial and exhumation are just a few examples of applications that require pressure–temperature estimates. In this chapter, we will review the general principles of geologic thermometry and barometry, and provide examples of specific reactions that are commonly used. We will also examine another type of phase diagram, the *pseudosection*, which displays quantitative pressure and/or temperature information for a particular bulk-rock composition.

19.2 SOLID SOLUTIONS AND THERMOBAROMETRY

The continuous reactions involving *AFM* phases that we examined in Chapter 18 can be thought of as resulting from two reactions proceeding simultaneously, one between the iron components and the other between the magnesium components. Reaction 18.7, for example, can be written as

$$41Al_2SiO_5 + 4Fe_5Al_2Si_3O_{10}(OH)_8$$
$$= 10Fe_2Al_9Si_4O_{23}(OH) + 13SiO_2 + 11H_2O$$

$$41Al_2SiO_5 + 4Mg_5Al_2Si_3O_{10}(OH)_8$$
$$= 10Mg_2Al_9Si_4O_{23}(OH) + 13SiO_2 + 11H_2O$$

If the second equation is subtracted from the first, we obtain

$$\begin{aligned}&10\text{Mg-staurolite} + 4\text{Fe-chlorite}\\ &\quad = 10\text{Fe-staurolite} + 4\text{Mg-chlorite}\end{aligned} \tag{19.1}$$

This is known as an Fe–Mg *exchange reaction*. The compositions of the coexisting staurolite and chlorite solid solutions depend on how far the reaction progresses to the right or left. At equilibrium, the compositions of the two minerals are related through the equilibrium constant for the reaction.

It will be recalled from Eq. (9.5) that the free energy of any assemblage is given by $G = \sum_i n_i \mu_i$. At equilibrium, the free energies of the left and right-hand sides of Eq. (19.1) must be equal; that is,

$$10\mu_{MgSt} + 4\mu_{FeCh} = 10\mu_{FeSt} + 4\mu_{MgCh}$$

On rearranging this gives

$$10(\mu_{FeSt} - \mu_{MgSt}) - 4(\mu_{FeCh} - \mu_{MgCh}) = 0$$

But the chemical potential of each component can be written in terms of the activity of that component according to

Eq. (19.34). For example, for the Fe-staurolite component we can write

$$\mu_{\text{FeSt}} = \mu^*_{\text{FeSt}} + RT\ln a_{\text{FeSt}}$$

where μ^*_{FeSt} is the chemical potential of pure Fe-staurolite at the given pressure and temperature. Substituting these expressions for each of the chemical potential terms and rearranging, we obtain

$$10\left(\mu^*_{\text{FeSt}} - \mu^*_{\text{MgSt}}\right) - 4\left(\mu^*_{\text{FeCh}} - \mu^*_{\text{MgCh}}\right) \\ = -10RT\ln\frac{a_{\text{FeSt}}}{a_{\text{MgSt}}} + 4RT\ln\frac{a_{\text{FeCh}}}{a_{\text{MgCh}}}$$

But $\mu^*_{\text{FeSt}} = \bar{G}_{\text{FeSt}}$. Thus the left-hand side of this equation is simply the free-energy change of the exchange reaction at the given pressure and temperature; that is,

$$\Delta G_{\text{exch}} = -RT\ln\frac{a^{10}_{\text{FeSt}}a^4_{\text{MgCh}}}{a^{10}_{\text{MgSt}}a^4_{\text{FeCh}}} \tag{19.2}$$

This free-energy change is known as the Fe–Mg *exchange potential*.

To relate the exchange potential to the composition of the minerals, we must relate the activities to mole fractions. From Eq. (10.14) it is known that if component *i* is mixed ideally on *n* equivalent sites in a mineral, the activity is given by $a_i = X_i^n$. Using this model, the activity of the Fe-staurolite component in staurolite becomes $a_{\text{FeSt}} = X^2_{\text{FeSt}}$ and for chlorite, $a_{\text{FeCh}} = X^5_{\text{FeCh}}$. Making these substitutions into Eq. (19.2) gives

$$\Delta G_{\text{exch}} = -RT\ln\frac{X^{20}_{\text{FeSt}}X^{20}_{\text{MgCh}}}{X^{20}_{\text{MgSt}}X^{20}_{\text{FeCh}}}$$

which on rearranging gives

$$-\Delta G_{\text{exch}} = 20RT\ln\frac{X_{\text{FeSt}}X_{\text{MgCh}}}{X_{\text{MgSt}}X_{\text{FeCh}}} \tag{19.3}$$

The ratio of the mole fractions of Fe and Mg in the coexisting phases is known as the *distribution coefficient*, which is designated K_{D}. In the example above, the distribution coefficient is for Fe and Mg between coexisting staurolite and chlorite and is therefore indicated as $K^{\text{StCh}}_{\text{DFeMg}}$. The distribution coefficient is a measurable property of a rock and requires careful analyses of coexisting minerals using the electron microprobe.

Equation (19.3) relates the distribution coefficient to the free-energy change of the exchange reaction between pure end members, the data for which are available from thermodynamic tables and software packages. Because ΔG_{exch} can be expressed as a function of pressure and temperature, the distribution coefficient can be used to interpret the pressures and temperatures under which coexisting minerals achieved their compositions. To do this we make use of Eq. (8.9), which for an exchange reaction that does not liberate a gas becomes

$$\Delta G_{\text{exch}} = \Delta H^0_{\text{exch}} - T\Delta S^0_{\text{exch}} + \left(P - 10^5\right)\Delta V^0_{\text{exch}}$$

where *P* is measured in pascal. For the general case, then, Eq. (19.3) can be written as

$$\Delta G_{\text{exch}} = -mRT\ln K_{\text{D}} \\ = \Delta H^0_{\text{exch}} - T\Delta S^0_{\text{exch}} + \left(P - 10^5\right)\Delta V^0_{\text{exch}}$$

and, with minor rearrangement

$$\ln K_{\text{D}} = -\frac{\Delta H^0_{exch}}{mRT} + \frac{\Delta S^0_{exch}}{mR} - \left(P - 10^5\right)\frac{\Delta V^0_{exch}}{mRT} \tag{19.4}$$

where *m* is a coefficient based on the activity model and the coefficients in the reaction ($m = 20$ for the example given in Eq. (19.3)). Whether a reaction can be used to indicate metamorphic temperatures (geothermometer) or pressures (geobarometer) depends on the magnitude of each term in this equation.

The volume change resulting from most Fe–Mg exchange reactions is very small ($\Delta V_{\text{exch}} \approx 0$), so Eq. (19.4) reduces to the equation of a straight line ($y = mx + b$), where y is $\ln K_{\text{D}}$; m, the slope of the line, is $\Delta H^0_{\text{exch}}/mR$; x is $1/T$ (K); and b, the intercept on the y axis, is $\Delta S^0_{\text{exch}}/mR$. Such an equation can be used as a geothermometer if the thermodynamic data are available for the end members, or the exchange reaction can be calibrated through experimental studies or against other geothermometers.

One exchange reaction that has received considerable attention involves coexisting garnet and biotite (see, for example, Thompson, 1976b; Ferry and Spear, 1978; Hodges and Spear, 1982). The exchange reaction, often called "GARB," can be written as follows

$$\underset{\text{(phlogopite)}}{\text{KMg}_3\text{AlSi}_3\text{O}_{10}(\text{OH})_2} + \underset{\text{(almandine)}}{\text{Fe}_3\text{Al}_2\text{Si}_3\text{O}_{12}} \\ = \underset{\text{(annite)}}{\text{KFe}_3\text{AlSi}_3\text{O}_{10}(\text{OH})_2} + \underset{\text{(pyrope)}}{\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}} \tag{19.5}$$

The expression for the distribution coefficient for the reaction is

$$K^{\text{Bi-G}}_{\text{DFe-Mg}} = \frac{X_{\text{FeBi}}/X_{\text{MgBi}}}{X_{\text{FeG}}/X_{\text{MgG}}} \tag{19.6}$$

which can be related to temperature and pressure (Eq. (19.4)) by

$$\ln K_{\text{D}} = -\left(\frac{\Delta H^0_{\text{exch}}}{3R}\right)\frac{1}{T} + \frac{\Delta S^0_{\text{exch}}}{3R} - \left[\left(P - 10^5\right)\frac{\Delta V^0_{\text{exch}}}{3RT}\right] \tag{19.7}$$

where the numeral 3 is for the three sites in the garnet and in the biotite structures involved in the exchange. The ΔV^0_{exch} is small, so that the terms in square brackets can be dropped without introducing serious error; the remaining expression for $\ln K_{\text{D}}$ can then be plotted as a linear function of $1/T$ (Fig. 19.1). This exchange reaction has been calibrated

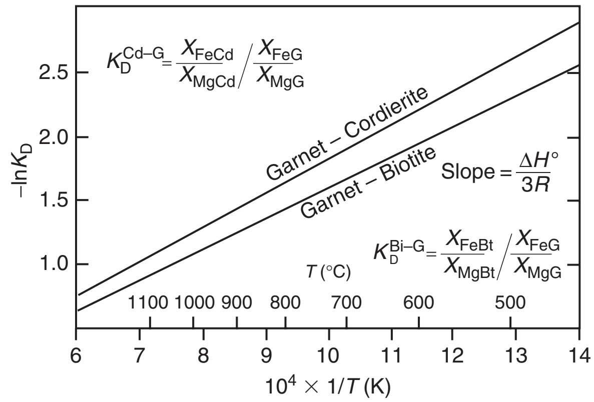


Fig. 19.1 Plots of $-\ln K_D$ versus $1/T$ (K) for the Fe–Mg exchange reactions between coexisting garnet–biotite and garnet–cordierite. See text for discussion (after Thompson, 1976b).

experimentally by Ferry and Spear (1978), who give the following equation for $\ln K_D$ incorporating ideal mixing of Fe and Mg in biotite and garnet

$$\ln K_D = -\left(\frac{52108}{3R}\right)\frac{1}{T} + \frac{19.51}{3R} - \left(P - 10^5\right) \times \frac{0.238 \times 10^{-5}}{3RT} \qquad (19.8)$$

where all values are expressed in SI units.

The biotite–garnet exchange reaction is a useful geothermometer, especially in view of the common occurrence of this assemblage in low to high grades of metamorphism (Fig. 18.8 (C), (D), (H)). As can be seen from Eq. (19.8), the ΔV^0_{exch} is small, so that even if the metamorphic pressure is not tightly constrained, a reasonably accurate temperature estimate (±50 ° C) can be made. More accurate estimates require consideration of nonideal mixing as well as pressure. For example, Ca and Mn substitution in garnet and Ti substitution in biotite affect the activities of the Fe and Mg end members (e.g. Berman, 1991). Furthermore, serious errors are introduced if large amounts of Fe^{3+}, F, or Cl are present. Upper amphibolite facies and granulite facies biotites require particular scrutiny, because they are commonly rich in Ti, F, and/or Cl (Essene, 1982).

For a reaction to be a good geobarometer, the $\Delta V$of reaction should be large. This eliminates simple ion-exchange reactions, which have very small ΔVs. Dehydration and decarbonation reactions have large ΔVs but these reactions are also functions of a_{H_2O} and a_{CO_2}, respectively, and these are commonly not known exactly. Reactions involving solids and having large ΔVs must therefore be used.

One of the most widely used geobarometers in metapelitic rocks involves the common assemblage plagioclase + garnet + Al_2SiO_5 + quartz (Ghent, 1976). It is based on the "GASP" reaction

$$\underset{\text{(anorthite)}}{3CaAl_2Si_2O_8} = \underset{\text{(grossular)}}{Ca_3Al_2Si_3O_{12}} + \underset{\text{(kyanite)}}{2Al_2SiO_5} + \underset{\text{(quartz)}}{SiO_2} \qquad (19.9)$$

The grossular is a component in garnet, which is predominantly almandine in pelitic rocks, and the anorthite forms a component in plagioclase, which is normally albite rich. An equilibrium expression for the reaction can be written in terms of the activities of these components in the two minerals using available thermodynamic data tables such as Holland and Powell (1998) (Chapter 7)

$$\ln\frac{a^G_{gross}}{\left(a^{Pl}_{an}\right)^3} = -\left(\frac{-40770}{R}\right)\frac{1}{T} + \frac{(-136.5)}{R} - \left(P - 10^5\right) \times \frac{\left(-6.605 \times 10^{-5}\right)}{RT} \qquad (19.10)$$

Note that Al_2SiO_5 and quartz are taken to be pure. Comparison of this equation with that for the biotite–garnet exchange (Eq. (19.8)) shows that the ΔV term in Eq. (19.10) is over a factor of 20 greater. For this reason, the reaction makes a good geobarometer. Note that the ratio of the activities also depends strongly on temperature. Thus, to use this geobarometer, the temperature must be known independently. This could be obtained, for example, from the biotite–garnet geothermometer. The difficulty in using this geobarometer, however, arises from having to know the activities of grossular in garnet and anorthite in plagioclase. These components are normally present in small amounts in their respective phases (particularly grossular in garnet), so serious errors can be introduced by assuming they behave ideally. Nonetheless, if we make this assumption, $a^G_{gross} = \left(X^G_{gross}\right)^3$, and $a^{Pl}_{an} = X^{Pl}_{an}$, and then

$$\ln\frac{a^G_{gross}}{\left(a^{Pl}_{an}\right)^3} = \ln\frac{\left(X^G_{gross}\right)^3}{\left(X^{Pl}_{an}\right)^3} = \ln K \qquad (19.11)$$

where K is the equilibrium constant. Various solution models for garnet and plagioclase have been used to refine this geobarometer (see Ghent *et al.*, 1979; Ghent and Stout, 1981; Hodges and Crowley, 1985; Berman, 1991). The amount of grossular component in the garnet of a typical pelitic rock is less than 10 mol %. Small errors in the analysis can therefore lead to large errors in the pressure determination. Also, errors in the temperature determination are translated into errors in pressure.

Another important geobarometer is based on the less common assemblage garnet + rutile + Al_2SiO_5 + ilmenite + quartz (Bohlen *et al.*, 1983)

$$\underset{\text{(almandine)}}{Fe_3Al_2Si_3O_{12}} + \underset{\text{(rutile)}}{3TiO_2} = \underset{\text{(ilmenite)}}{3FeTiO_3} + \underset{\text{(Al silicate)}}{Al_2SiO_5} + \underset{\text{(quartz)}}{2SiO_2} \qquad (19.12)$$

This reaction, known as "GRAIL" is almost independent of temperature, proceeding to the left at pressures greater than about 1 GPa for pure phases (Fig. 19.2). The equilibrium constant for the reaction can be expressed as

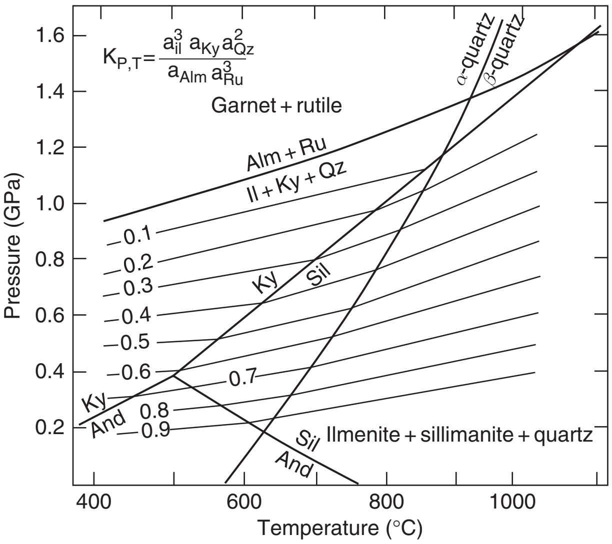


Fig. 19.2 *P–T–*$\log_{10}K$ plot for ilmenite–Al_2SiO_5–quartz–almandine–rutile geobarometer (Bohlen *et al.*, 1983; published by permission of *American Mineralogist*).

$$K = \frac{\left(a_{\text{Ilm}}^{\text{Ilm}_{\text{ss}}}\right)^3 \left(a_{\text{Al}_2\text{SiO}_5}^{\text{K,A,S}}\right) \left(a_{\text{SiO}_2}^{\text{Q}}\right)^2}{\left(a_{\text{Alm}}^{\text{G}_{\text{ss}}}\right) \left(a_{\text{TiO}_2}^{\text{Ru}}\right)^3} \tag{19.13}$$

In most metapelitic rocks, rutile, quartz, and Al_2SiO_5 polymorphs form essentially pure phases, and thus they have unit activities. However, ilmenite forms a solid solution with hematite, and almandine forms a solid solution in garnet. Ilmenite and hematite do not form an ideal solid solution series; indeed, they are separated by a large solvus at low temperatures. In metapelitic rocks, however, ilmenite usually contains less than 15 mol% hematite, so that ilmenite can be treated as an ideal solution without introducing serious error. Several solution models have been proposed for garnet. Bohlen *et al.* (1983) use a regular solution model (Section 9.6) with a temperature-dependent interchange energy. Using these models for the solid solution phases and the experimental data for the end-member reaction (Eq. (19.12)), Bohlen *et al.* have calculated the position of lines of constant $\log_{10}K$ for this reaction on a *P–T* diagram (Fig. 19.2). Note that these lines have a very shallow slope, so that metamorphic temperatures need not be known accurately for this geobarometer to be used. Temperature uncertainties of ±50 °C result in maximum errors in inferred pressure of only 0.05 GPa. Note also that the lines of constant *K* have inflections where they cross the univariant lines for the transformations between the various polymorphs of Al_2SiO_5 and of quartz. Many other reactions can be written that involve ilmenite and rutile besides GRAIL. The rutile tends to be on the high-pressure side of such reactions, reflecting its small molar volume (0.951×10^{-5} J Pa^{-1}; one oxygen basis).

Cation-based thermobarometry is not limited to reactions in metapelitic rocks. For example, Kohn and Spear (1990) provide two barometers suitable for garnet-bearing amphibolites. Plagioclase–garnet–pyroxene–quartz geobarometry has proven to be extremely useful for granulite facies rocks (e.g. Moecher *et al.*, 1988; Eckert *et al.*, 1991).

In subduction zone settings, the simple reaction

$$\underset{\text{(albite)}}{\text{NaAlSi}_3\text{O}_8} = \underset{\text{(jadeite)}}{\text{NaAlSi}_2\text{O}_6} + \underset{\text{(quartz)}}{\text{SiO}_2}$$

has a shallow d*P*/d*T* slope (Fig. 16.6) and is therefore a good geobarometer. It is very useful for estimating pressures for subducted metabasalts and metasediments, once the activities of albite in feldspar and jadeite in pyroxene are accounted for. The simple calcite = aragonite reaction is another barometer applicable to subducted rocks (Fig. 16.6). In addition to the polymorphic phase transitions quartz = coesite and graphite = diamond (Fig. 16.6), reactions among pyroxenes and garnet are important for unraveling the pressure–temperature conditions of eclogite facies and ultrahigh-pressure rocks (e.g. Carson *et al.*, 1999; Ravna and Terry, 2004). In particular, the Fe–Mg exchange between garnet and clinopyroxene has been calibrated in a number of studies (e.g. Ellis and Green, 1979; Pattison and Newton, 1989; Ravna, 2000) and is commonly used for eclogite (and granulite) geothermometry

$$\begin{aligned} &\underset{\text{(pyrope)}}{\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}} + \underset{\text{(hedenbergite)}}{\text{CaFeSi}_2\text{O}_6} \\ &\quad = \underset{\text{(almandine)}}{\text{Fe}_3\text{Al}_2\text{Si}_3\text{O}_{12}} + \underset{\text{(diopside)}}{\text{CaMgSi}_2\text{O}_6} \end{aligned}$$

One potential pitfall with application of this thermometer is that estimates of pyroxene Fe^{3+} content based on electron microprobe analyses can have large errors which then lead to errors in the temperature estimates. With respect to pressure, several options exist; a promising geobarometry reaction involving clinopyroxene, garnet, and muscovite components is (Ravna and Terry, 2004; Page *et al.*, 2007)

$$\begin{aligned} &\underset{\text{(diopside)}}{6\text{CaMgSi}_2\text{O}_6} + \underset{\text{(muscovite)}}{3\text{KAl}_2(\text{AlSi}_3)\text{O}_{10}(\text{OH})_2} = \underset{\text{(grossular)}}{2\text{Ca}_3\text{Al}_2\text{Si}_3\text{O}_{12}} \\ &\quad + \underset{\text{(pyrope)}}{\text{Mg}_3\text{Al}_2\text{Si}_3\text{O}_{12}} + \underset{\text{(celadonite)}}{3\text{KAlMg}(\text{Si}_4)\text{O}_{10}(\text{OH})_2} \end{aligned}$$

Here, celadonite represents the phengite substitution in the K-mica; note that substitution of Fe^{2+} for Mg^{2+} also occurs in natural phengites. Increasing pressures favor increasing amounts of phengitic substitution (Velde, 1965; Coggon and Holland, 2002). Much additional work remains to develop geothermometers and geobarometers applicable to the wide range of high-pressure and ultrahigh-pressure rocks encountered in the field.

19.3 THERMOBAROMETRY AND MULTIPLE REACTIONS

When using the general thermobarometry equation (19.4) with ideal mixing, we know K_D or K as well as the changes in enthalpy, entropy, and volume. So the only two unknowns are pressure and temperature. Thus, if one uses a thermometer

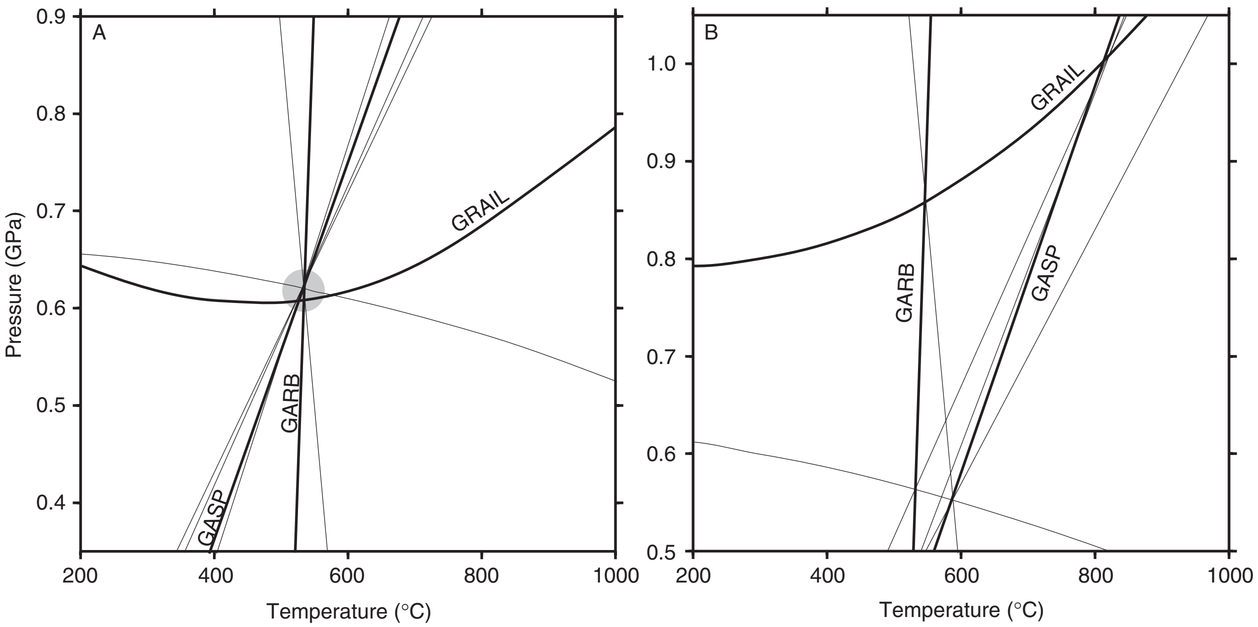


Fig. 19.3 *P–T* calculations for the GARB, GASP, and GRAIL reactions (thick lines), together with all other linearly dependent reactions that can be written by combining them (thin lines). (**A**) For this sample, the reactions intersect in a narrow region of *P–T* space (gray circle), providing a tightly constrained *P–T* estimate. (**B**) Here, the reaction intersections scatter widely, indicating disequilibrium or problems with the thermodynamic data, activity models, and/or chemical analyses of the minerals.

like GARB together with a barometer like GASP for the same rock, there are two equations that can be solved simultaneously for pressure and temperature.

This concept can be extended to include even more reactions like GARB + GASP + GRAIL, but here there are more equations than unknowns, so the solution will in general not be unique. If all the minerals reached equilibrium and the thermodynamic data, activity models, and chemical analyses of the minerals are sound, then the reactions should intersect in a small region on a pressure–temperature diagram. This region defines the conditions of equilibration (Fig. 19.3(A)). On the other hand, the curves may have intersections that scatter across a wide range of pressures and temperatures (Fig. 19.3(B)). This could mean that equilibrium was not attained, or that there were problems with the data used to solve Eq. (19.4). For high-quality microprobe analyses and well-calibrated, internally consistent thermodynamic data sets and activity models, the culprit is commonly disequilibrium.

Chemical equilibrium or the lack thereof can be investigated further by considering *all* reactions between end members. As discussed by Berman (1991), the total number of reactions that can be written for a system with c system components and n end members is

$$n!/[(c+1)!(n-(c+1))!] \tag{19.14}$$

an expression that follows from the rule of combinations noted in Chapter 18. Thus, for GARB, GASP, and GRAIL taken together, there are eight total reactions. Of these, only $n-c=3$ are linearly independent. The others are simply linear combinations of the three independent reactions. Simultaneous solution of all the reactions in a *multiequilibrium* calculation will yield a tightly constrained set of pressure–temperature intersections for a well-equilibrated sample, but many scattered intersections for samples characterized by disequilibrium (Berman, 1991) (Fig. 19.3). The degree of scatter gives an indication of how closely equilibrium was approached, and provides a measure of the uncertainty on pressure–temperature estimates.

Disequilibrium is powerful because it can provide important constraints on metamorphic histories. Different mineral assemblages may equilibrate under different pressure–temperature conditions at different times during metamorphism. For example, an early-formed mineral assemblage preserved as inclusions within garnet may yield very different pressure–temperature estimates than later-formed mineral assemblages in the matrix of a rock. Such differences allow parts of the pressure–temperature path of the rock to be reconstructed. To be most useful, the pressure–temperature estimates should be combined with textural and geochronologic information that constrains the timing of mineral growth.

To avoid simplifications, large multiequilibrium computations are usually best done using software packages such as THERMOCALC (Holland and Powell, 1998; www.earthsci.unimelb.edu.au/tpg/thermocalc) or winTWQ (Berman, 1991; http://gsc.nrcan.gc.ca/sw/twq_e.php). It turns out that the thermobarometry equation (19.4) is simplified because it assumes that the enthalpy, entropy, and volume changes of reaction remain constant. In reality, they are not constant as temperature and pressure change, leading to subtle but in some cases important curvature of reactions such as GRAIL

(Fig. 19.2). Ideal mixing is another simplification. Most solutions are nonideal, and nonideal activity coefficients are generally functions of both temperature and pressure. Computer software packages reduce considerably the potential for mistakes when calculating very complicated nonideal solution models, some of which involve dozens of terms, or when integrating values of thermodynamic quantities (cf. Eq. (7.45)). Nonetheless, the general principles of thermobarometry should be mastered first in order to understand what the software is doing and to recognize errors in input data or calculation parameters.

19.4 TRACE ELEMENT THERMOMETRY

The substitution of trace elements into mineral structures can also be strongly dependent on temperature or pressure. In Chapter 9, the Ti-in-quartz geothermometer (Wark and Watson, 2006) was shown to have important applications in igneous petrology. Two other recently calibrated geothermometers are Zr-in-rutile and Ti-in-zircon (Watson *et al.*, 2006). The amounts of Ti substitution into quartz or zircon are very small unless temperatures are extremely high. Thus, an ion probe is required to measure the Ti concentrations in all but the highest-grade granulite facies rocks. However, considerable substitution of Zr into rutile occurs at much lower temperatures, so the Zr-in-rutile thermometer based on substitution of Zr into rutile coexisting with zircon and quartz is widely applicable using concentration data obtained from the electron microprobe (Zack *et al.*, 2005; Watson *et al.*, 2006). The substitution can be written in terms of oxides as:

$$\mathrm{ZrO}_{2(\mathrm{zircon})} = \mathrm{ZrO}_{2(\mathrm{rutile})} \qquad (19.15)$$

where the subscripts (zircon) and (rutile) indicate the phases involved. If we take the zircon to be essentially pure so its activity is unity, the equilibrium constant expression (Eq. (9.56)) yields

$$a_{\mathrm{ZrO}_2}^{\mathrm{rutile}} = \gamma_{\mathrm{ZrO}_2}^{\mathrm{rutile}} \cdot X_{\mathrm{ZrO}_2}^{\mathrm{rutile}} = \exp\left(\frac{-\Delta G^{\mathrm{o}}}{RT}\right) \qquad (19.16)$$

in which $\gamma_{\mathrm{ZrO}_2}^{\mathrm{rutile}}$ and $X_{\mathrm{ZrO}_2}^{\mathrm{rutile}}$ are the activity coefficient and mole fraction for the trace amount of ZrO_2 in the rutile, respectively. If the activity coefficient remains constant or nearly so, then Eq. (19.16) predicts that the logarithim of the Zr concentration will be linear when plotted against $1/T$. Watson *et al*. (2006) give the following thermometer calibration based on laboratory experiments and natural samples

$$\log(Zr[ppm]) = 7.36 \pm 0.10 - \frac{4470 \pm 120}{T(K)} \qquad (19.17)$$

This thermometer has many attractive features. It is simple to apply and can yield a high-precision temperature estimate using the chemical composition of just one mineral (rutile). The Zr concentration of rutile can be measured relatively easily with the electron microprobe for crystals equilibrated at temperatures as low as ~500 ºC. The assemblage rutile + zircon is widely distributed and can be found in many greenschist, amphibolite, granulite, blueschist, and eclogite facies rocks.

Some caveats must also be kept in mind. Equation (19.17) is calibrated for a pressure of ~1 GPa. There is some indication that the thermometer is pressure dependent, and these effects could be important, particularly for high-pressure and ultrahigh-pressure rocks formed in excess of ~2 GPa. The rate of diffusion of Zr in rutile is not insignificant, so Zr contents are subject to retrograde resetting during cooling (Watson *et al.*, 2006). Thus, “peak” compositions are probably best preserved by rutile crystals encased in refractory porphyroblasts like garnet. If the rock lacks zircon, then the Zr content of rutile at a given temperature will be less than predicted by Eq. (19.17); the thermometer expression will therefore yield a minimum *T* estimate.

It is also important to point out that zircons are highly refractory and often react slowly, so there may be kinetic constraints on reaction (Eq. (19.15)), particularly at lower metamorphic grades. Other silicate minerals, however, can contain Zr, so when they break down Zr may be released. The solubility of Zr in typical fluids is very low, so it will tend to precipitate locally in new zircon rims on existing zircon grains, and become incorporated in rutile crystals growing at the same time. In this way, Eq. (19.16) is still valid, even though existing zircons need not have dissolved to provide Zr for rutile. Furthermore, the new zircon rims can be dated by measuring U/Pb isotope systematics with the ion probe, thereby determining the age of metamorphic reaction and zircon growth.

19.5 SOLVUS THERMOMETRY

In Section 10.20, we examined how temperatures can be estimated for igneous rocks using solvi in the pyroxene quadrilateral. The phase relations among coexisting pyroxenes can be used to estimate temperatures for high-grade metamorphic rocks too. There are a number of other solvi that are important in metamorphic rocks, including solvus relations among the ternary feldspars (Snoeyenbos *et al.*, 1995).

One solvus that is very useful for temperature estimation is the calcite–dolomite solvus. As can be seen in Figure 19.4, the Mg content of calcite coexisting with dolomite increases quite systematically with temperature. The Ca content of the dolomite is less temperature sensitive, so the thermometer is based on the calcite limb of the solvus. The variation in calcite Mg content with temperature is influenced by substitution of Fe and Mn into the calcite. Anovitz and Essene (1987) provide an empirical calibration of the solvus thermometer and common solid solution effects. Note that calcite deforms and recrystallizes easily, and Mg can diffuse through the structure relatively quickly at high temperatures. For these reasons, the thermometer is susceptible to resetting during retrogression. Calcite grains included in refractory porphyroblasts often provide the best chance of retaining

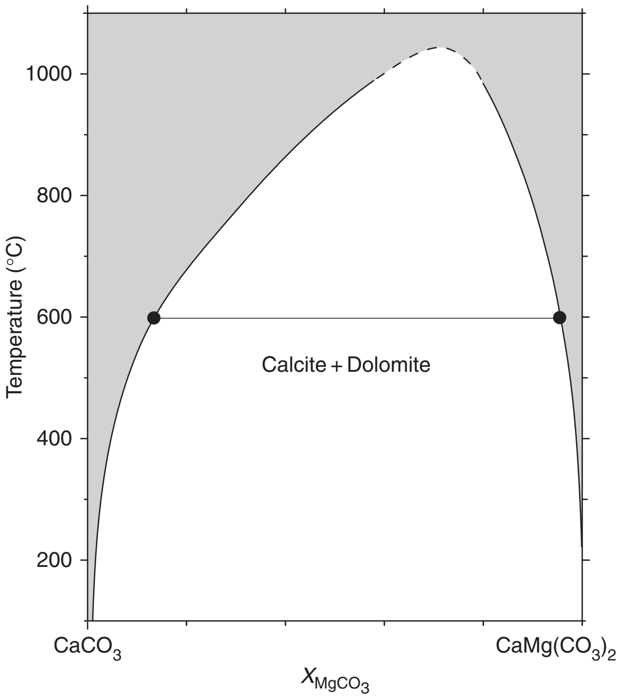


Fig. 19.4 The calcite–dolomite solvus (modified after Anovitz and Essene, 1987, and references cited therein.) Example tie-line shows compositions of coexisting calcite and dolomite at 600 °C.

the "peak" (maximum temperature) composition, particularly in high-temperature amphibolite and granulite facies rocks.

19.6 OXYGEN ISOTOPE THERMOMETRY

As we saw earlier in Eq. (13.31), the equilibrium partitioning of stable isotopes between minerals A and B is temperature sensitive, and can be described by

$$1000 \ln \alpha_{A-B} = a_0 + \frac{a_1 \times 10^6}{T^2} \tag{19.18}$$

where a_0 is a constant (usually zero if A and B are anhydrous), a_1 is the temperature coefficient of fractionation, T is in kelvin, and α_{A-B} is

$$\alpha_{A-B} = \frac{1000 + \delta^{18}O_A}{1000 + \delta^{18}O_B}$$

for oxygen isotopes.

In practice, one evaluates Eq. (19.18) for as many mineral pairs as possible to assess the internal consistency of the temperature estimates. This effort can be aided by a graphical analysis of the data (Javoy *et al.*, 1970). It is easily verified that $1000 \ln \alpha_{A-B}$ is approximated by $\delta^{18}O_A - \delta^{18}O_B = \Delta^{18}O_{A-B}$ for isotopic differences that are not too large. Quartz (or coesite) normally has the largest $\delta^{18}O$ value. Therefore, based on Eq. (19.18), we can write the following expression for $\Delta^{18}O_{\text{quartz}-\text{mineral}}$ that describes quartz coexisting in oxygen isotopic equilibrium with another mineral of interest

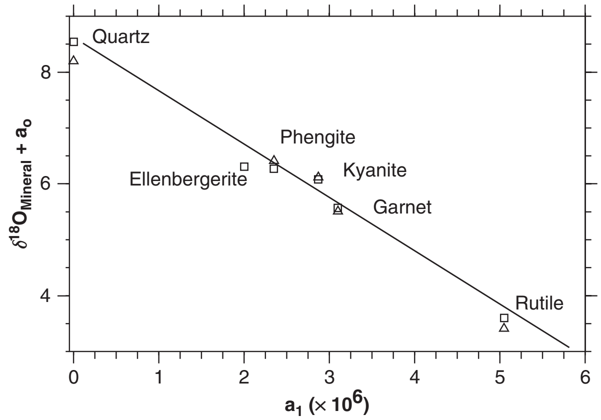


Fig. 19.5 Oxygen isotope systematics for minerals in UHP rock, Dora Maira Massif, western Alps. The minerals plot along a linear isotherm indicating an equilibration temperature of about 750 °C. Ellenbergerite forms under UHP conditions; it is a dense, hydrous, Al-bearing silicate mineral that is Mg-rich and that contains significant Ti, Zr, and P. (After Sharp *et al.*, 1993; published by permission of Springer-Verlag.)

$$\delta^{18}O_{\text{mineral}} + a_0 = \delta^{18}O_{\text{quartz}} - \frac{a_1 \times 10^6}{T^2} \tag{19.19}$$

On a graph with $\delta^{18}O_{\text{mineral}} + a_0$ on the y-axis and temperature coefficient on the x-axis, Eq. (19.19) defines a line with a negative slope ($-1/T^2$) and a y-intercept equal to the $\delta^{18}O$ value of quartz (Fig. 19.5). Plotted in this way, all minerals should fall on the same line (an isotherm) if isotopic equilibrium was reached and the temperature coefficients of fractionation are well known (Javoy *et al.*, 1970; Palin, 1992). Equilibrium $\Delta^{18}O_{A-B}$ values and, thus, isotherm slopes, decrease with increasing temperature.

An isotherm plot from Sharp *et al.* (1993) for ultrahigh-pressure (UHP) rocks from the Dora Maira Massif, western Alps, is shown in Figure 19.5. Chopin's (1984) discovery of coesite in the Massif was the first definitive identification of this mineral in regional metamorphic rocks anywhere in the world. The oxygen isotope systematics of the minerals record temperatures of about 750 ºC. Given that the pressures were 3 GPa or more, the temperatures indicate a low geothermal gradient consistent with subduction. Importantly, oxygen isotope thermometers are not sensitive to pressure, even at UHP conditions.

Oxygen isotope thermometry is critically dependent on the calibrations of the temperature-dependent coefficients of fractionation, and much work remains in this regard. The reader is referred to recent work by Valley (2001) and Sharp (2006) for discussions of these coefficients and isotope thermometry in general. Note that oxygen isotope thermometry is susceptible to retrogression like other thermometers we have discussed. However, this can give valuable perspectives on cooling histories (e.g. Eiler *et al.*, 1993; van Haren *et al.*, 1996). Finally, it should be pointed out that other stable isotope thermometers are available, including thermometry involving partitioning of C isotopes between calcite and graphite.

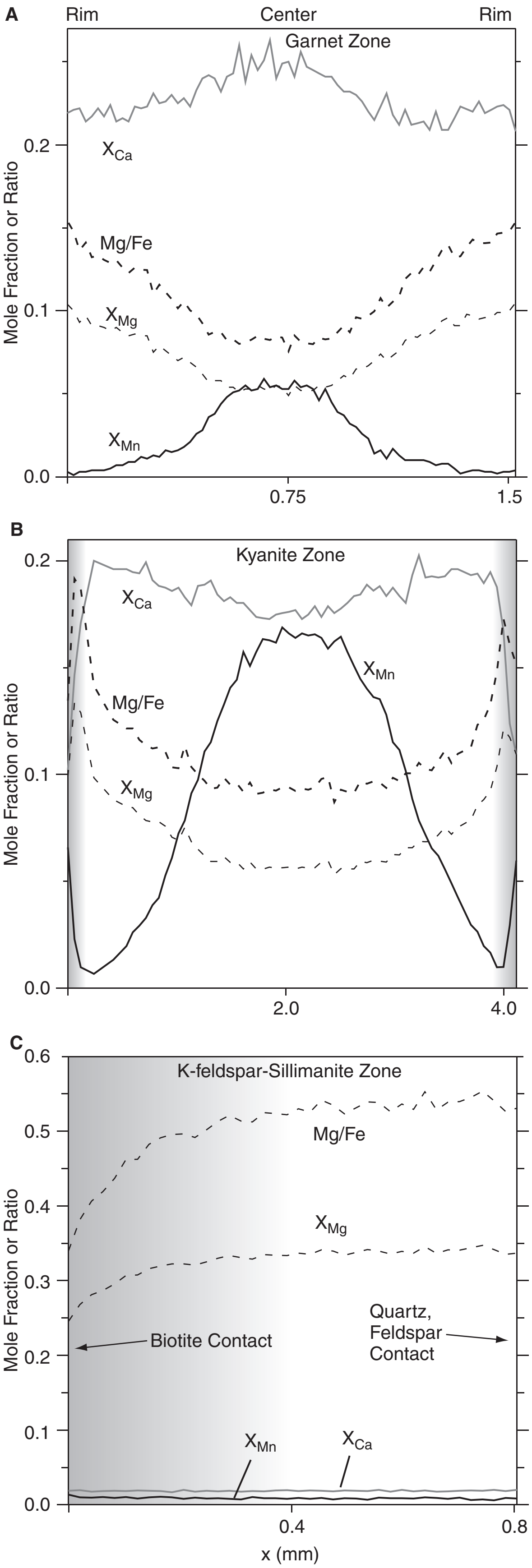


Fig. 19.6 Compositional profiles from rim to rim across garnet crystals obtained using the electron microprobe. (**A**) Garnet from Garnet zone, Scottish Highlands. Oscillations in composition along profiles are real variations in mineral chemistry, not analytical artifacts. Note typical bell-shaped Mn zoning profile, and increases in Mg/Fe toward the rims consistent with growth during prograde heating. (**B**) Kyanite zone, Wepawaug Schist, Connecticut. Margins of crystal affected by post-peak metamorphic retrograde reactions (gray shading). (**C**) K-feldspar–sillimanite zone, North Carolina. Left margin of crystal in contact with biotite, which exchanged Mg and Fe with the garnet by diffusion during cooling (gray shading). Right margin in contact with quartz and feldspar did not undergo retrograde exchange.

19.7 MINERAL ZONING AND THERMOBAROMETRY

Chemical or isotopic zoning in crystals (Chapters 10, 12) is essential to consider when using geothermometers and geobarometers (e.g. Kohn, 2003). Zoning can reflect growth under varying pressure–temperature–composition conditions; modification of original growth zoning by diffusion; and/or breakdown and resorption of grains. The process of intracrystalline (within grain) diffusion acts to smooth compositional heterogeneities within crystals. Consequently, minerals that have slow rates of intracrystalline diffusion commonly preserve zonation acquired as conditions changed during growth. Examples include garnets, pyroxenes, and amphiboles. Sheet silicates have faster rates of intracrystalline diffusion, so they are less commonly zoned. However, because diffusion rates are slow at low temperatures, zoning may be found in some sheet silicates formed at lower metamorphic grades.

The important question for thermobarometry is: what mineral compositions should be used to estimate temperature and pressure if the minerals are zoned? We will focus on garnet in metapelitic rocks because it is essential for many thermobarometry reactions, but the general concepts can be extended to other minerals. Two-dimensional chemical maps are best for illuminating complex zoning patterns (Fig. 12.27), but much can also be learned from simple chemical profiles extending from rim to rim across grains (Fig. 19.6).

Garnets have a strong preference for Mn. The rates of cation diffusion in garnet during growth under garnet zone conditions are very slow. Consequently, the Mn remains sequestered in the growing garnet and is depleted from the surrounding rock matrix. As a result, the Mn content of garnet typically decreases from center to edge, yielding the familiar "bell-shaped" Mn zoning profile (Figs. 12.17, 19.6(A), (B)). As we saw in Section 12.4, this growth zoning can be modeled using Rayleigh fractionation.

Compositional growth zoning is evident for other cations as well. Growth under conditions of increasing temperature during prograde metamorphism commonly results in an increase in the Mg/Fe ratio from center to edge, consistent with the GARB reaction (Fig. 19.6(A)). The Ca content of garnet can respond to changes in pressure or temperature via reactions like GASP.

Petrologists are often concerned with the "peak" metamorphic conditions; that is, the maximum temperature and pressure that a rock was subjected to. A complication is that the peak pressure need not have been attained at the same time as the peak temperature, a topic we will return to below and in Chapter 22. Even if we assume simultaneous attainment of peak temperature and pressure, determining the peak garnet composition can be challenging and must consider the metamorphic grade of the rock. For example, for garnet zone (greenschist facies) rocks, it is commonly assumed that garnet growth ceased once the rock began to cool and/or be exhumed. In this case, the peak composition is that of the last-formed garnet – the composition right at the rim of the crystal (Fig. 19.6(A)).

In the amphibolite facies, garnets can be consumed, grow, or simply remain in the matrix. Retrograde reactions are commonly significant, as the rocks cool over a long temperature interval. A very common feature is an upturn in the Mn content at garnet rims (Fig. 19.6(B)). This has been interpreted to reflect partial consumption of the garnet rim by retrograde reactions (Kohn and Spear, 2000). The phases produced by these reactions contain little Mn, so it is concentrated at the edge of the remaining garnet. The area of increased Mn is also commonly accompanied by decreased Mg/Fe and Ca. An important factor to consider in this regard is that amphibolite facies temperatures are high enough that diffusion can be significant in garnet. Consequently, decreases in rim Mg/Fe may reflect diffusive exchange with matrix biotite during cooling. On the other hand, changes in rim compositions may also be due to a late stage of growth during retrograde cooling and exhumation. For the example in Figure 19.6(B), the thickness of the retrograde rim is not dependent on the proximity to matrix minerals with which the garnet might have reequilibrated. Thus, retrograde reaction probably involved a grain boundary fluid.

Estimation of peak conditions is often done by measuring the garnet composition at the Mn minimum just inboard of any upturn in Mn content, because this garnet grew during prograde conditions and was presumably not severely modified by resorption, growth, and/or diffusion at the rim (Fig. 19.6(B)). It is also typical to measure biotite compositions at some distance (mm to cm scale) from such garnets, because these biotites will have undergone less retrograde Mg–Fe exchange during cooling than those in direct contact with garnet rims. While these rules of thumb are useful, great care must be exercised when inferring peak compositions (Kohn and Spear, 2000). Of course, measuring the compositions of minerals in contact with garnet rims can give a pressure–temperature estimate that records part of the retrograde history of the rock. This information can be critical for tectonic reconstructions of pressure–temperature–time (*P*–*T*–*t*) paths and exhumation (Fig. 19.7) (Hodges and Royden, 1984).

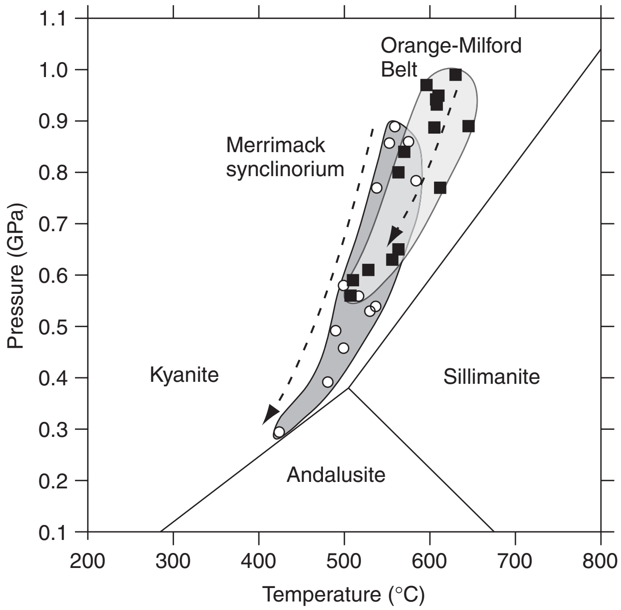


Fig. 19.7 *P*–*T* paths (arrows) for exhumation of kyanite-bearing outcrops, Connecticut, obtained using GARB, GASP, GRAIL, and other reactions. Filled squares denote *P*–*T* estimates obtained from metapelitic and metacarbonate rocks of the Wepawaug Schist, Orange-Milford belt, south-central Connecticut (Ague, 1994a, 2002). Peak metamorphic conditions at 0.9–1.0 GPa; minerals reequilibrated to varying degrees as the rocks were brought to the surface. Lowest *P*–*T* results from retrograded garnet rims (Fig. 19.6(B)) and minerals in contact with them. Open circles denote *P*–*T* estimates from kyanite-bearing veins in the Merrimack synclinorium, northeastern Connecticut. Lowest *P*–*T* estimates from mineral assemblages in contact with retrograded garnet rims (Ague, 1995).

In the upper amphibolite facies (Sillimanite zone) and the granulite facies, garnets commonly have "flat" compositional profiles, with some retrograde changes in composition at the rim like decreased Mg/Fe (Fig. 19.6(C)). Temperatures are high enough that diffusion can operate over hundreds of micrometers or even millimeter distances on geologic timescales. Thus, a common interpretation of "flat" compositional profiles is that zoning acquired during earlier, lower-grade stages of growth has been eliminated by diffusion at high temperatures. However, the possibility that "flat" profiles may in some cases result from growth at high metamorphic grades should also be considered in detailed petrological studies, as garnet can be produced at high temperatures by reactions like: quartz + muscovite + staurolite = biotite + garnet + Al_2SiO_5 + H_2O. In either case, the rims record retrograde cooling and/or decompression. So the best estimate of peak conditions is usually obtained from the garnet core composition (e.g. Tracy *et al*., 1976).

As discussed above, the compositions of minerals at some distance from the garnets should be measured to estimate peak conditions, whereas the garnet rim and minerals directly in contact with it can record part of the retrograde pressure–temperature history. For the example in Figure 19.6(C), retrograde exchange of Fe and Mg is only evident where the garnet is in contact with biotite. This type of phenomenon is common in the highest grade rocks, where it probably indicates that the rocks lacked a grain boundary fluid phase that could have helped bring about retrograde reactions all around the garnet rim.

Garnets in some rocks display abrupt, step-like changes in composition from center to edge (see Fig. 12.27). These are commonly the result of multiple stages of garnet growth that occurred under different pressure–temperature–composition conditions.

Growth zoning provides an invaluable record of how conditions changed during metamorphism of a rock package.

Consequently, it is a powerful tool to unravel metamorphic pressure–temperature histories. For example, a modification of the biotite–garnet geothermometer of Thompson (1976b) and Ferry and Spear (1978) was developed by Spear and Selverstone (1983) so that it can be used as a geobarometer as well. It takes advantage of the fact that the zoning so commonly exhibited by garnet crystals preserves a record of a range of metamorphic temperatures and pressures. In the Ferry–Spear geothermometer, the value of K_D is used in Eq. (19.4) to solve for T. In the Spear–Selverstone method, T is not solved for directly but only values of ΔT and ΔP are determined as functions of mineral compositions. If a specific temperature and pressure are known for one pair of compositions, the remainder can then be determined from the calculated values of ΔT and ΔP.

Spear and Selverstone's method involves expressing P and T as functions of n linearly independent compositional variables, where n is the variance of the system (i.e. P or $T = f(X_1, X_2 \dots X_n)$). As temperatures and pressures change, so do the compositions of coexisting biotite and garnet. The total differentials of T and P can be expressed as

$$\mathrm{d}T = \sum \left(\frac{\partial T}{\partial X_i}\right)_{X_{j\neq i}} \mathrm{d}X_i \quad \text{and} \quad \mathrm{d}P = \sum \left(\frac{\partial P}{\partial X_i}\right)_{X_{j\neq i}} \mathrm{d}X_i$$

Values of $(\partial T/\partial X_i)_P$, $(\partial P/\partial X_i)_T$, and $(\partial P/\partial T)_X$ can be solved for using molar entropies and volumes of all phases in the assemblage and a model for mixing in the solid solutions.

Consider, for example, the common assemblage in pelitic rocks of garnet + biotite + Al_2SiO_5 + muscovite + quartz + H_2O (Fig. 19.8). This forms a divariant assemblage that is bounded at low temperatures by the discontinuous reaction

$$\text{garnet} + \text{chlorite} = \text{biotite} + Al_2SiO_5$$

and at high temperatures by the discontinuous reaction

$$\text{muscovite} + \text{quartz} = \text{K} - \text{feldspar} + Al_2SiO_5$$

as indicated in Figure 19.8. Within the divariant field, the compositions of both garnet and biotite vary with temperature and pressure. This would show in an *AFM* plot as shifts in the tie triangle garnet–biotite–Al_2SiO_5. Spear and Selverstone (1983) have calculated the compositions of coexisting garnet and biotite, assuming both to be ideal binary solutions; the composition of the garnet is expressed as mole fraction of almandine, X_{Alm}, and that of biotite as mole fraction of annite, X_{Ann}.

The isopleths (lines of constant composition) for coexisting garnet and biotite form a series of intersecting lines that allow the pressure and temperature of equilibration to be read for any coexisting pair (Fig. 19.8). Note that with increasing pressure, both the garnet and biotite become progressively more magnesian. Because the isopleths for the two minerals intersect at high angles in the kyanite stability field, P–T conditions can be determined more accurately there than in the other fields where the angle of intersection is small. Within the sillimanite and andalusite fields, the garnet isopleths have little slope to them; that is, they are functions almost entirely of pressure.

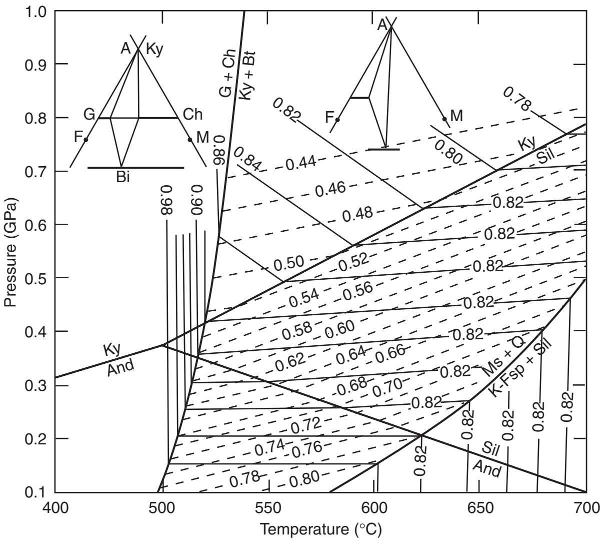


Fig. 19.8 *P–T* plot with isopleths giving the mole fractions of almandine in garnet (solid lines) and annite in coexisting biotite (dashed lines) in pelitic rocks (simplified from Spear and Selverstone, 1983; published by permission of Springer-Verlag).

Graphs such as the one shown in Figure 19.8, and others that take into account the effects of Mn, Ca, Na, and K on the garnet and biotite compositions (Spear and Selverstone, 1983), are extremely useful in unraveling the pressure–temperature–time (*P–T–t*) histories of rocks. Although assumptions about the solution models for the minerals may cause the absolute temperatures and pressures to be inaccurate, the range of temperatures and pressures are only slightly affected. Use of the method does require that biotite grains can be found included in zoned garnet, and that diffusion has not modified their composition since inclusion. Careful electron microprobe traverses across grain boundaries can usually resolve the extent to which compositions have been modified. Experience shows that zoned garnet crystals do indeed preserve long histories of changing *P–T* conditions (see Section 22.7).

19.8 INTRODUCTION TO PSEUDOSECTIONS

Pressure–temperature (*P–T*) grids and thermobarometry provide quantitative *P–T* information, whereas compatibility diagrams like *AFM* projections illustrate equilibrium mineral assemblages. Is there a way to capture both mineral assemblage and *P–T* relations graphically on a single phase diagram? The answer is yes, if the bulk composition of the rock is specified. Such phase diagrams are called *pseudosections*. What follows is based heavily on Powell *et al*. (1998); the interested reader is referred to this paper for further details.

We begin by considering an example of a pressure–temperature (*P–T*) pseudosection. Here, the basic idea is to take the bulk composition of interest and calculate the *P–T* stability fields of all the equilibrium mineral assemblages using known thermodynamic properties and activity models for the minerals. The example diagram shows univariant lines, divariant fields and so on quantitatively in terms of

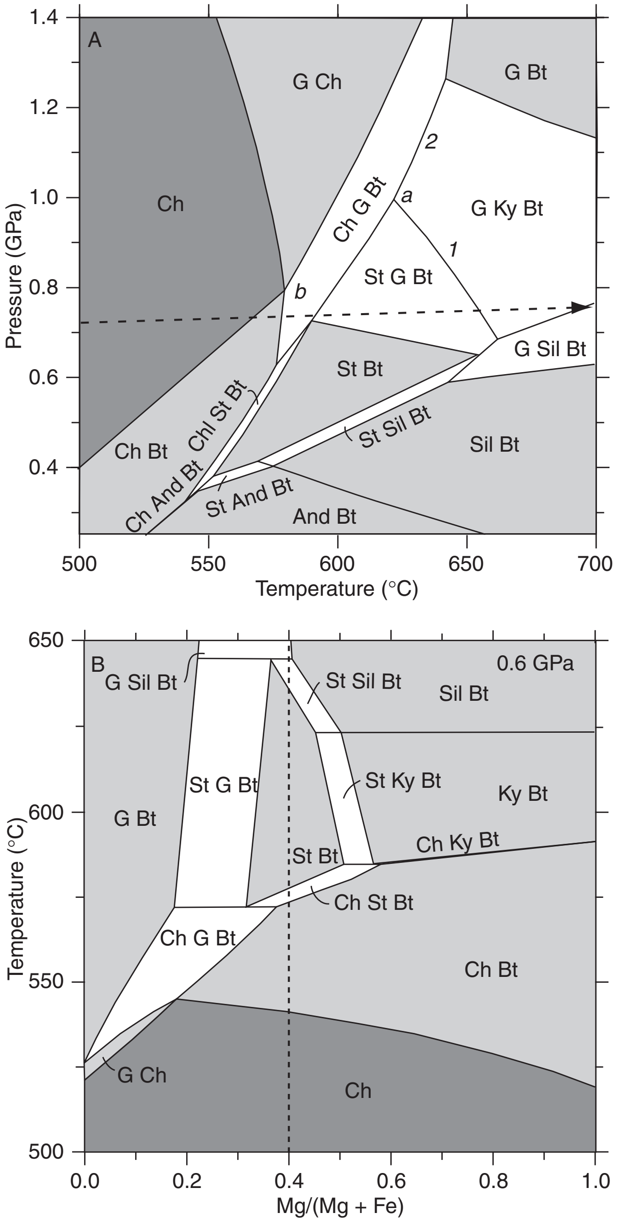


Fig. 19.9 (**A**) *P–T* pseudosection for a representative metapelitic bulk composition having Al_2O_3= 41.89, MgO = 18.19, FeO = 27.29, and K_2O = 12.63 (mol%). All mineral assemblages coexist with water (activity = 1) and muscovite. Quadrivariant field is dark gray, trivariant fields are medium gray, and divariant fields are white. Minerals are chlorite (Ch), garnet (G), biotite (Bt), staurolite (St), andalusite (And), kyanite (Ky), and sillimanite (Sil). Dashed arrow represents *P–T* path discussed in the text and in Problem 19.7; see text for discussion of other symbols.
(**B**) Temperature–composition pseudosection drawn for 0.6 GPa. Vertical dotted line represents the bulk composition depicted in (A). (Both (A) and (B) from Powell *et al*., 1998; published by permission of Blackwell.)

pressure and temperature for a model metapelitic bulk composition (Fig. 19.9(A)).

This bulk composition comprises K_2O-FeO-MgO-Al_2O_3-SiO_2-H_2O – the same oxides used for *AFM* projections (this chemical system is abbreviated as “KFMASH”). Additional similarities between the *AFM* projection and the *P–T* pseudosection are that quartz and muscovite are assumed to coexist with all mineral assemblages, and that the activity of water must be specified. A critical difference, however, is that the specific molar ratios of K_2O:FeO:MgO:Al_2O_3 in the bulk composition must be defined for the pseudosection. Although the example is for typical phases in metapelitic rocks, *P–T* pseudosections can be drawn for virtually any common rock type, including metamorphosed ultramafic rocks, carbonate rocks, or basalts (Problem 19.9).

The sizes of the stability fields, such as the divariant garnet–biotite–kyanite field, are rigorously constrained by the thermodynamic and solid solution properties of the minerals. The compositions of the biotite and garnet solid solutions in the garnet–biotite–kyanite field are not constant, but change such that the $(Fe^{2+}/Mg)^{biotite}/(Fe^{2+}/Mg)^{garnet}$ increases from low temperature to high temperature across the field (as predicted by the GARB reaction). The modal proportions of phases can also change continuously across stability fields. For a given temperature and pressure, the garnet–kyanite–biotite divariant assemblage corresponds to a three-phase garnet–kyanite–biotite triangle on an *AFM* diagram. If we changed the *P–T* conditions, the compositions of the biotite and garnet would change and we would have to draw a new set of *AFM* tie lines.

P–T pseudosections contain univariant curves and invariant points as do *P–T* grids. It is important to emphasize, however, that pseudosections only show those curves and points that are appropriate for the bulk rock composition of interest. A further distinction is that *P–T* pseudosections also explicitly show phase boundaries between all the other stability fields relevant for a bulk composition, such as boundaries between divariant and trivariant fields.

Discontinuous univariant reactions are interpreted in the usual way. For example, the garnet–kyanite–biotite field is bounded by several univariant curves (Fig. 19.9(A)). Curve *1* marks the disappearance of staurolite and appearance of kyanite with increasing temperature according to the general terminal discontinuous reaction: staurolite + muscovite + quartz = garnet + biotite + kyanite + H_2O. At higher pressures, curve *2* marks the disappearance of chlorite and appearance of kyanite. The intersection between three divariant fields is an invariant *P–T* point. So garnet, biotite, chlorite, kyanite, and staurolite all coexist at invariant point “a” in Figure 19.9(A).

What about the points and phase boundary lines involving fields of higher variance? Here, boundary lines mark the disappearance (or appearance) of one phase, whereas points involve the disappearance (or appearance) of two phases. So, for example, as temperature decreases at 0.9 GPa, there is a transition from the trivariant garnet + chlorite field to the quadrivariant chlorite field at about 575 ºC (garnet is eliminated). For another example, consider a temperature decrease from the divariant chlorite + garnet + biotite field, through point “b” at about 0.8 GPa and 580 ºC, into the quadrivariant chlorite field (Fig. 19.9(A)). Here two phases (biotite and garnet) are lost.

Interpretation of general mineral assemblage changes across the various fields is straightforward – a major advantage that pseudosections have over *P T* grids. For instance, a rock heated along the dashed *P–T* path in Figure 19.9(A) would be characterized by the following *AFM* assemblages with increasing grade: chlorite, chlorite + biotite, chlorite + garnet + biotite, staurolite + garnet + biotite, kyanite + garnet + biotite, and sillimanite + garnet + biotite (each of these coexists with muscovite, quartz, and H_2O). The reader will recall that this progression of index minerals corresponds to the Barrovian sequence. Of course, if the rocks also contained plagioclase feldspar, then partial melting reactions would also have to be considered at the highest temperatures (e.g. Fig. 18.15).

Construction of pseudosections is computationally very involved and is best done with the aid of appropriate computer programs. An outline of one approach is given here; the reader is referred to Powell *et al*. (1998) for details. A different approach that involves direct minimization of Gibbs free energy is given by Connolly and Petrini (2002).

As an example, we will examine qualitatively pseudosection calculations for point b on Figure 19.9(A). One way to proceed is to consider all the linearly independent chemical reactions between mineral end members in the system of interest. An equilibrium constant expression can then be written for each equation. So, for the Fe–Mg exchange reaction between garnet and chlorite end members

$$\begin{aligned}&\underset{\text{(pyrope)}}{5\mathrm{Mg_3Al_2Si_3O_{12}}} + \underset{\text{(daphnite)}}{3\mathrm{Fe_5Al_2Si_3O_{10}(OH)_8}}\\ &\quad = \underset{\text{(almandine)}}{5\mathrm{Fe_3Al_2Si_3O_{12}}} + \underset{\text{(clinochlore)}}{3\mathrm{Mg_5Al_2Si_3O_{10}(OH)_8}}\end{aligned}$$

we can write

$$0 = \Delta G^{\mathrm{o}} + RT\ln K = \Delta G^{\mathrm{o}} + RT\ln\left(\frac{a_{\mathrm{Py}}^5 a_{\mathrm{Daph}}^3}{a_{\mathrm{Alm}}^5 a_{\mathrm{Clin}}^3}\right) \tag{19.20}$$

where ΔG^{o} is the standard state Gibbs free energy change for the reaction at the pressure and temperature of interest, and K is the equilibrium constant. Ultimately, we need to solve for the equilibrium mineral compositions at the pressure and temperature of point b. The mineral compositions are incorporated into Eq. (19.20) by specifying the activity–composition relationships in K. For example, if the mixing in garnet is ideal, then the activity of almandine is simply: $(X_{\text{Fe-Garnet}})^3$ (Eq. (10.14)). Temperature and pressure enter the expression for ΔG^{o} (cf. Eq. (7.45)), and in the activity terms if the mixing is nonideal.

At point b (Fig. 19.9(A)), Powell *et al*. (1998) show that there are seven totally independent reactions relating the various chlorite, garnet, biotite, and muscovite end members they consider. This yields seven expressions of the form of Eq. (19.20), and a total of nine unknowns including pressure, temperature, and the mineral compositions. We are not done, therefore, because in order to solve the system of equations we need the same number of equations and unknowns. So in addition to these seven equations, a set of mass balance equations is written to describe how the modal proportions of the phases combine to produce the specified bulk composition. The mass balance equations relate the modes of chlorite, garnet, biotite, and muscovite to the amounts of K_2O, MgO, FeO, and Al_2O_3 in the bulk composition. Equations are not needed for SiO_2 and H_2O because they have the same compositions as the phases quartz and fluid. So there are four new mass balance equations (one each for K_2O, MgO, FeO, and Al_2O_3) and four new unknowns (the modal proportions of chlorite, garnet, biotite, and muscovite). But the number of unknowns can be reduced by two because the proportions of garnet and biotite go to zero at point b.

With this last simplification, the equilibrium constant and mass balance constraints form a nonlinear system of equations that can be solved directly to yield the pressure, temperature, mineral compositions, and mineral modes at point b. Calculation of *P–T* boundary lines is similar, but involves the elimination of one phase instead of two. In general, pseudosection calculations yield both stable and metastable equilibria; Schreinemakers rules are applied to determine the stable ones that are plotted.

P–T pseudosections have many uses. A common application is to predict the mineral assemblages formed in a rock along prograde and retrograde *P–T* paths to deduce reaction histories. This effort is aided by contouring mineral modes on the stability fields; these can then be compared with observed assemblages in the rocks. The stability fields can also be contoured for mineral compositions. Thus, if the mineral and bulk rock compositions are known, the general *P–T* conditions of equilibration can be estimated from the diagram (more precise estimates are usually best obtained using specific geothermometers and geobarometers). Open-system processes, like fluid-rock reaction or melt extraction accompanying high-grade metamorphism, can be portrayed using a series of *P–T* pseudosections which take account of the changing bulk composition.

Of course, one has to keep several potential pitfalls in mind. Phase relations will depend critically on water activity, and can also change if other constituents not considered in the model chemical system are significant. For example, MnO is not present in KFMASH, but it can be very important for stabilizing garnet in the greenschist facies. Furthermore, the example assumes that the minerals are chemically homogeneous and that equilibrium is attained everywhere on the diagram. The assumption of homogeneity can be problematic for minerals that are commonly zoned, like garnet (although the effects of zoning could be added to the calculations). *P–T* pseudosections have the potential to identify chemical disequilibrium, because if the observed phase relations differ markedly from the predicted equilibrium ones, it may be an indication that kinetic controls on reaction rates were important. However, caution must be exercised to make sure that the discrepancies are not due to other factors, like an over-simplified choice of chemical system, errors in the bulk composition, or large uncertainties on the thermodynamic data and/or activity models for mineral end members.

Pseudosections can also depict variations in bulk composition. For example, one could construct a temperature–composition pseudosection that specifies the total amount of MgO + FeO in the bulk composition. Phase relations could then be plotted in terms of temperature and the Mg/(Mg + Fe^{2+}) ratio of the bulk composition, holding pressure constant. Pressure–composition pseudosections that hold temperature constant could be constructed in an analogous way. The bulk composition in Figure 19.9(A) has Mg/(Mg + Fe^{2+}) = 0.4. Consequently, the phase relations of Figure 19.9(A) at 0.6 GPa plot along the vertical Mg/(Mg + Fe^{2+}) = 0.4 dotted line on Figure 19.9(B). A rock with this composition is predicted to contain staurolite and biotite at amphibolite facies conditions (~600 ºC, 0.6 GPa). On the other hand, a very high Mg/(Mg + Fe^{2+}) rock would contain biotite and kyanite. An example of this latter case can be found in Figure 18.8(G). Here, the rock crystallized at high oxygen fugacity and much of the iron occurs as Fe^{3+} in oxides (hematite–ilmenite solid solution). Consequently, the Mg/(Mg + Fe^{2+}) ratio of the rock is high, and the amount of Fe^{2+} available to make Fe-rich phases like garnet or staurolite is limited.

19.9 FIELD EXAMPLES

In closing, let us examine briefly two field areas in New England where geothermometers and geobarometers have been used extensively to give a regional picture of metamorphic conditions.

In south-central Maine, pelitic rocks of the Silurian Waterville Formation underwent Buchan-type (low-pressure) regional metamorphism during the Devonian Acadian Orogeny. Ferry (1980) has used eight different geothermometers/geobarometers to estimate the conditions during this metamorphic episode. To the west, in central Massachusetts, rocks ranging in age from late Precambrian to lower Devonian were also metamorphosed during the Acadian Orogeny but under Barrovian-like conditions. Geothermometers and geobarometers have been used by Robinson *et al*. (1982) and Tracy *et al*. (1976) to unravel the complex history of this region.

Rocks of the Waterville Formation in south-central Maine were deformed into a series of tight isoclinal folds. They appear to have undergone only one period of metamorphism. Many porphyroblasts overgrow and crosscut the schistosity (e.g. Figs. 16.4(D), 17.14(B)), indicating that much of the metamorphism followed the deformation that caused the schistosity. Mapping by Osberg (1971) has shown that the metamorphic grade increases from chlorite zone in the north to sillimanite zone in the south. The following succession of isograds has been mapped – biotite, garnet, staurolite + andalusite, staurolite + cordierite, and sillimanite (Fig. 19.10). Intrusive bodies of synmetamorphic quartz monzonite occur in the highest-grade zones.

Ferry (1980) analyzed minerals in approximately 200 samples of pelitic schist from the area shown in Figure 19.10 in order to determine the pressures and temperatures during metamorphism. Above the garnet isograd, the ubiquity of the assemblage garnet + biotite makes this mineral pair the most convenient to use for geothermometry. The isotherms determined using these minerals are shown in Figure 19.10. Temperatures rise from about 430 °C at the garnet isograd to 570 °C in the southern part of the region. Pressures throughout the region are all very near 0.35 GPa. The range of *P–T* conditions recorded in these rocks is shown in Figure 19.11 along with the Al_2SiO_5 phase diagram for reference.

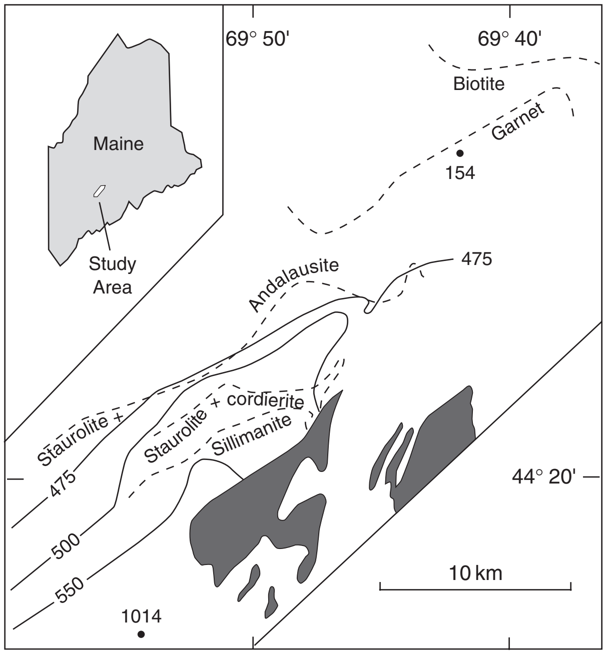


Fig. 19.10 Map of south-central Maine showing isograds (dashed lines) and isotherms (solid lines, °C) determined using the garnet–biotite geothermometer. Stippled areas are synmetamorphic quartz monzonite bodies. Two numbered dots refer to samples discussed in Problem 19.1 (simplified from Ferry, 1980; published by permission of *American Mineralogist*).

The second field area is in central Massachusetts and has a more complex geologic history than south-central Maine. At about 400 Ma the rocks were folded into nappes which were overturned to the west. These were then backfolded (overturned to the east) and at about 375 Ma the rocks were again strongly deformed as gneiss domes buoyantly rose in the Bronson Hill anticlinorium (Fig. 19.12). Before and during the nappe stage, sheetlike intrusions of gabbro, tonalite, and granite were emplaced. Sillimanite pseudomorphs after andalusite indicate an early low-pressure metamorphism in the eastern part of the area. Peak metamorphic conditions were attained during the backfolding stage with various retrograde reactions taking place during doming and unloading. The metamorphic isograds were deformed along with the primary rock units by the backfolding and doming.

The area is bounded to the west by the border fault of the Mesozoic Hartford and Deerfield basins. The lowest-grade rocks occur immediately east of this fault and directly north of the Deerfield basin. Their grade increases eastward toward the Bronson Hill anticlinorium through zones of chlorite, biotite, garnet, kyanite + staurolite, sillimanite + staurolite, sillimanite + muscovite, sillimanite + muscovite + K-feldspar, sillimanite + K-feldspar, and garnet + cordierite +

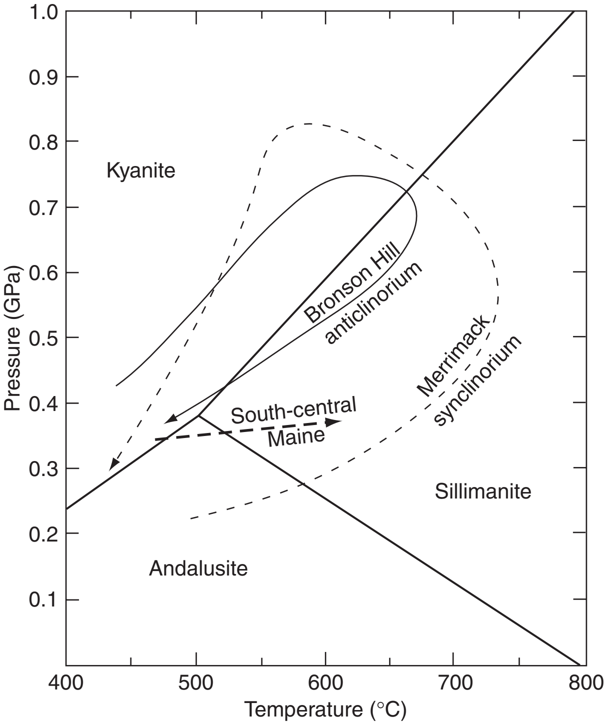


Fig. 19.11 Metamorphic *P–T* paths in south-central Maine and central Massachusetts. The Massachusetts rocks were subjected to higher pressures than those in Maine (see Fig. 18.15). The retrograde cooling and decompression path of the Merrimack synclinorium from ~0.85 GPa to ~0.3 GPa based on the *P–T* estimates shown in Figure 19.7.

sillimanite + K-feldspar. Farther to the east the grade decreases toward the Merrimack synclinorium through many of the same zones, but instead of a kyanite + staurolite zone, andalusite + staurolite is present.

By paying careful attention to the type of zoning in garnet crystals, their compositions have been used in conjunction with those of other minerals to determine *P–T* conditions at various stages in the history of these rocks. Figure 19.12 shows isotherms representing the peak temperatures reached by the rocks. Temperatures rise from about 600 °C in the kyanite + staurolite zone to 650 to 700 °C in the highest-grade zones. Pressures increase over this range from about 0.5 to 0.7 GPa. Rocks in different parts of the region reached their peak conditions via different paths. Rocks in the Merrimack synclinorium (eastern part) passed through the andalusite field along an *anticlockwise* or *counterclockwise P–T* path, whereas those in the Bronson Hill anticlinorium (central part) passed through the kyanite field along a *clockwise* path (Fig. 19.11) (see also Chapter 22). The *P–T* estimates of Figure 19.7 from an adjacent area in Connecticut further constrain the cooling and decompression path of the Merrimack synclinorium (Ague, 1995).

A striking feature of the isotherms in both the Maine and Massachusetts areas is that they do not deviate greatly from the pattern produced by the isograds. An isograd in these regions can therefore be thought of as a line of almost constant metamorphic temperature. This is found to be the case in many other metamorphic terranes. Thus, although rocks in different regions may be metamorphosed at different pressures, the different metamorphic zones reflect largely different temperatures. In the Waterville formation, for example, the pressure recorded in the rocks in passing from Chlorite to Sillimanite zones is a constant 0.35 GPa; yet the temperatures rise from 430 to 570 °C. Interestingly, the pseudosection for a typical metapelitic bulk composition suggests that some Barrovian metamorphic sequences may also develop by heating with little pressure change (dashed path in Fig. 19.9(A)). Regional metamorphism can therefore bear a striking resemblance to contact metamorphism (constant *P*, variable *T*), but on a larger scale.

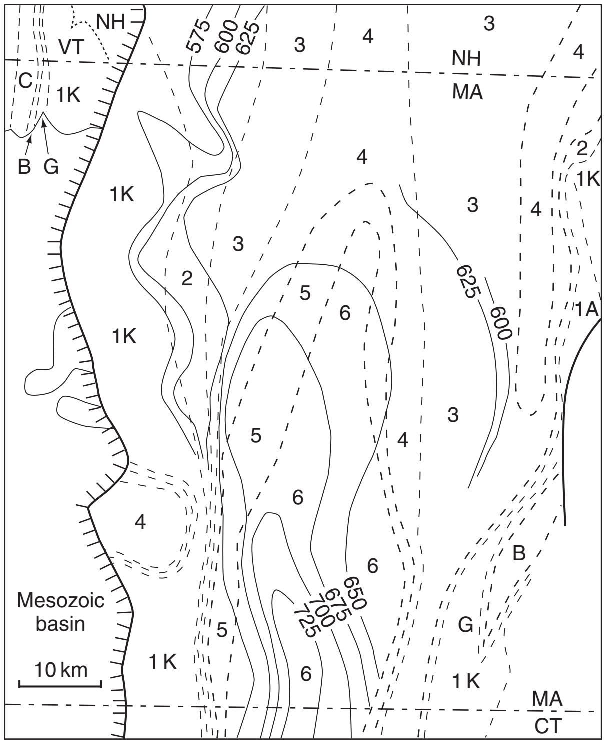


Fig. 19.12 Isograds (dashed lines) and isotherms (solid lines, °C) in the Bronson Hill anticlinorium and Merrimack synclinorium of central Massachusetts (MA) (after Robinson *et al.*, 1982). The metamorphic zones are chlorite (C), biotite (B), garnet (G), andalusite–staurolite (1A), kyanite–staurolite (1K), sillimanite–staurolite (2), sillimanite–muscovite (3), sillimanite–muscovite–K-feldspar (4), sillimanite–K-feldspar (5), and garnet–cordierite–sillimanite–K-feldspar (6).

Some words of caution are warranted regarding interpretation of the record preserved by geothermometers/geobarometers and mineral assemblages throughout a metamorphic terrane. As we have seen, rocks can follow complex *P–T–t* paths during burial and exhumation (Fig. 19.11). As a consequence, the "peak" temperature need not correspond to the "peak" pressure. Moreover, although points recording identical temperatures can be joined by lines that we label "isotherms," nothing proves that all rocks on an "isotherm" attained that temperature simultaneously. Isotherms and isograds need not be isochronous. Nor need two rocks that fall on different isotherms, for example one at 600 °C and the other at 700 °C, have attained those temperatures at the same time. In Chapter 22 we will see that time is an important variable, and that the records of temperature and pressure in rocks can vary systematically in age across a region.

19.10 PROBLEMS

19.1 The following table gives the mole fractions of Fe and Mg in coexisting garnet and biotite in pelitic rocks studied by Ferry (1980) in south-central Maine. Sample number 154 is from near the garnet isograd in one of the lowest-grade parts of the region, whereas number 1014 is from the highest-grade part of the region (Fig. 19.10). Geobarometers indicate that the pressure at the time of metamorphism was 0.35 GPa throughout the region. Using the Ferry and Spear (1978) calibration of the garnet–biotite geothermometer, determine the temperatures recorded by the minerals in these two rocks.

Sample number	Location isograd	Garnet		Biotite	
		X_{Fe}	X_{Mg}	X_{Fe}	X_{Mg}
154	Garnet	0.680	0.044	0.490	0.290
1014	Sill-Ksp	0.710	0.088	0.457	0.323

19.2 Recalculate the compositions of the garnet and biotite in sample 1014 from the previous problem in terms of mole fraction almandine and annite, assuming them to be simple binary Fe–Mg solutions. Using the isopleths in the P–T diagram of Spear and Selverstone (Fig. 19.8), determine the pressure and temperature of metamorphism of this rock. Is your answer consistent with the values obtained in the previous problem? Sample 1014 is from the sillimanite–K-feldspar isograd. In view of the position of the reaction defining this isograd in Figure 19.8, what can you conclude about the activity of H_2O during metamorphism in south-central Maine?

19.3 A metapelitic schist from the kyanite zone, south-central Connecticut, contains garnet with the following "peak metamorphic" rim composition (mole fractions): $X_{almandine} = 0.740$, $X_{pyrope} = 0.145$, $X_{grossular} = 0.098$, $X_{spessartine} = 0.017$. Coexisting biotite has the composition $(K_{0.855}Na_{0.018})(Mg_{1.283}Fe^{2+}_{1.084}Mn_{0.003}Ti_{0.086}Al^{VI}_{0.438})(Al^{IV}_{1.276}Si_{2.724})O_{10}(OH_{1.957}F_{0.041}Cl_{0.002})$, and coexisting plagioclase has the composition $An_{28}Ab_{71}Or_1$.

(a) Solve for the pressure and temperature of equilibration simultaneously using the GASP and GARB equilibria (Eqs. (19.8) and (19.10)). Assume ideal mixing. For the annite component in biotite, compute the mole fraction of Fe^{2+} in octahedral sites using $Fe^{2+}/(Fe^{2+} + Mg + Mn + Ti + Al^{VI})$. Use an analogous procedure for Mg.

(b) Repeat part (a), but take into account nonideal behavior in the solids. Use an activity coefficient of 1.630 for anorthite, 1.292 for grossular, 0.981 for almandine, 1.253 for pyrope, 1.163 for phlogopite, and 1.037 for annite (see activity models of Berman, 1990, Furhman and Lindsley, 1988, and McMullin *et al.*, 1991). So, for example, the activity of grossular in garnet is $(1.292\ X_{grossular})^3$. For biotite, assume that mixing on octahedral sites is nonideal but is ideal elsewhere. For example, the activity of annite is calculated as: $(X_K)(1.037X_{Fe})^3(X_{OH})^2$ (Berman, 1990; McMullin *et al.*, 1991). Here we have assumed that all Fe is ferrous; note that the calculations would need to be modified if the biotite contained significant Fe^{3+}. How do your results compare with those from part (a)?

(c) The rock contains abundant zircon and rutile crystals. Rutile inclusions in kyanite contain 171 ±21 ppm Zr (±2 sigma standard error). Use Eq. (19.17) to compute the temperature of equilibration. Is the result consistent with the GARB thermometry in parts (a) and (b)?

19.4 The major silicate minerals in a metapelitic rock from the same general area as the sample in Problem 19.3 are muscovite, biotite, quartz, plagioclase, garnet, kyanite, and staurolite. Quartz, muscovite, kyanite, and garnet rims preserve peak metamorphic $\delta^{18}O$ values, and the coefficients of fractionation are reasonably well known. The average $\delta^{18}O$ values are: quartz = 15.92 ±0.1‰, muscovite = 13.55 ±0.30‰, kyanite = 12.91 ±0.20‰, and garnet rims = 11.94 ±0.29‰ (±2 sigma standard errors; all values reported relative to the V-SMOW standard; data from van Haren *et al.*, 1996).

(a) Make an isotherm plot analogous to Figure 19.5 for these isotopic data. For quartz–muscovite, use $a_0 = -0.6$ and $a_1 = 2.2$ (Bottinga and Javoy, 1975). The a_1 values to use for quartz–garnet and quartz–kyanite are 3.1 and 2.25, respectively (Sharp, 1995).

(b) Using Eq. (19.19), compute the temperature of equilibration for the quartz–muscovite, quartz–kyanite, and quartz–garnet pairs.

(c) Using a calculator or a spreadsheet program, fit a line through the isotherm data plotted in part (a), and compute the temperature from the slope. How does this "average" temperature compare with the individual quartz–mineral pair values computed in part (b)? Are your results consistent with the thermobarometry in Problem 19.3?

19.5 A metapelitic rock contains almandine-rich garnet coexisting with quartz, kyanite, rutile and ilmenite. The mole fraction of ilmenite in the rhombohedral oxides is 0.95, and the mole fraction of almandine in the garnet is 0.75. The quartz, kyanite, and rutile are essentially pure. Independent geothermometry indicates a metamorphic temperature of 600 ºC. Using an activity coefficient of 0.98 for the almandine end member and ideal mixing for ilmenite, estimate the pressure of equilibration at this temperature using the GRAIL geobarometer (Fig. 9.2).

19.6 A high-grade metacarbonate rock contains forsterite olivine with inclusions of dolomite and calcite. The mole fraction of $MgCO_3$ (X_{MgCO_3}) in the calcite is on average 0.15. Calcite coexisting with dolomite in the matrix, on the other hand, has $X_{MgCO_3} = 0.025$. From Figure 19.4, what temperatures do the inclusion and matrix calcite record? What is the most likely reason for the discrepancy?

19.7 Draw schematic *AFM* projections for each mineral assemblage that would form during heating along the dashed *P–T* path in the pseudosection of Figure 19.9 (A). Assume that equilibrium is always maintained, that the rock does not change composition other than for water loss, and that the activity of water is one.

19.8 Simple pressure–temperature calculations with the THERMOCALC program. These instructions should work for version 3.30; later versions may use a different set of input files. Three files are needed. You can download two of them, tc-prefs.txt and tc-scr.txt, from the book's web site (www.cambridge.org/philpotts). The third file contains the mineral name and activity information for your calculation, and you can create it as described below. Also, make sure that all needed files, including the thermodynamic database (tc-ds55.txt for version 3.30), are in the same folder.

(1) Using a text editor, create an input file for the GARB and GASP reactions with the following format. The first column contains the mineral name abbreviation, the second holds the activities, and the third is unused. The abbreviations are q: quartz, ky: kyanite, alm: almandine, py: pyrope, gr: grossular, ann: annite, phl: phlogopite, an: anorthite. You will need to fill in the "activity value" entries with actual numerical values. The mole fractions of almandine, pyrope, and grossular in garnet are 0.75, 0.15, and 0.10, respectively. The biotite has X_{Fe} and X_{Mg} both equal to 0.4, and the plagioclase is An_{30}. Use ideal mixing (e.g. mole fraction cubed for garnet and biotite end members). You can also calculate nonideal activities using the AX program linked to the THERMOCALC site. The mineral end members can be listed in any order, and the number of spaces between the various entries on a line doesn't matter, but if the mineral abbreviations are wrong the example will not work. For example, if you enter Q instead of q for quartz, the program won't work properly (many word processors automatically capitalize the first letter, so this is something to watch out for). Make sure there are no blank lines at the top of the file or between lines in the file, and don't forget the asterisk at the bottom of the file. Name the file tc-tst.txt. If you are using Microsoft Word, save the file in the "plain text" format. Remember, the file needs to be in the same folder as the THERMOCALC program.

```
q        1.0               0.0
ky       1.0               0.0
alm      activity value    0.0
py       activity value    0.0
gr       activity value    0.0
ann      activity value    0.0
phl      activity value    0.0
an       activity value    0.0
*
```

(2) Start THERMOCALC, and select option 1 for phase diagram calculations.

(3) Read in your input file, following the prompts on your screen. If the program can't find your file at first, respond "yes" when the program asks to look for it. Then type in the full file name (e.g. tcdtst.txt). If it still can't find the file, try closing your text editor and rerunning the program. Once the file is read in, the program will inform you of the names of the files containing the output.

(4) Hit return on your keyboard for both the "which phases" and "in excess" prompts. This will select all the phases in the file.

(5) Answer "yes" for the "calculate T at P" option.

(6) Enter the pressure range of interest; 1 to 10 kbar works well for this problem. Just enter the numbers, e.g. 1 10; don't type "1 to 10" or "1,10" or "kbar". Note that pressure is in units of kilobars, not GPa or bars.

(7) Enter the temperature range of interest; 100 to 1000 ºC works well for this problem.

(8) Enter the desired pressure interval for output; 1 (kbar) is adequate.

(9) The program will type out the reactions, and the equilibrium *P–T* values which can then be plotted. The *P–T* estimate is obtained where the two reaction curves intersect.

(10) For another example, stop the program and go back and edit your input file, adding a line for muscovite (symbol mu) with an activity of 0.7. Now start THERMOCALC with this new file. There are more than two reactions, and the program can examine their intersections and estimate average *P–T* conditions and uncertainties. The documentation that accompanies the program provides extensive details.

(11) Select option 2, "average pressure-temperature calculations". After your new file is read in, hit return for the "which end members" prompt.

(12) From the next list of options, select option 3, "average PT".

(13) The program will ask if the data to be used look OK. If so, type "yes" or just hit return.

(14) Next enter the temperature (degrees C) and pressure (kbar) ranges for calculation, all on one line. 200 800 1 10 works for this example.

(15) Use 1 for the allowed angle change parameter.

(16) The program will then calculate the reactions, and provide estimates of temperature and pressure, along with their respective standard deviations.

(17) Try experimenting with THERMOCALC option 1 "phase diagram calculations" and option 3 "calculations on all reactions between end-members" with this same input data file. How does the output from these options relate to that of option 2 "average pressure-temperature calculations"?

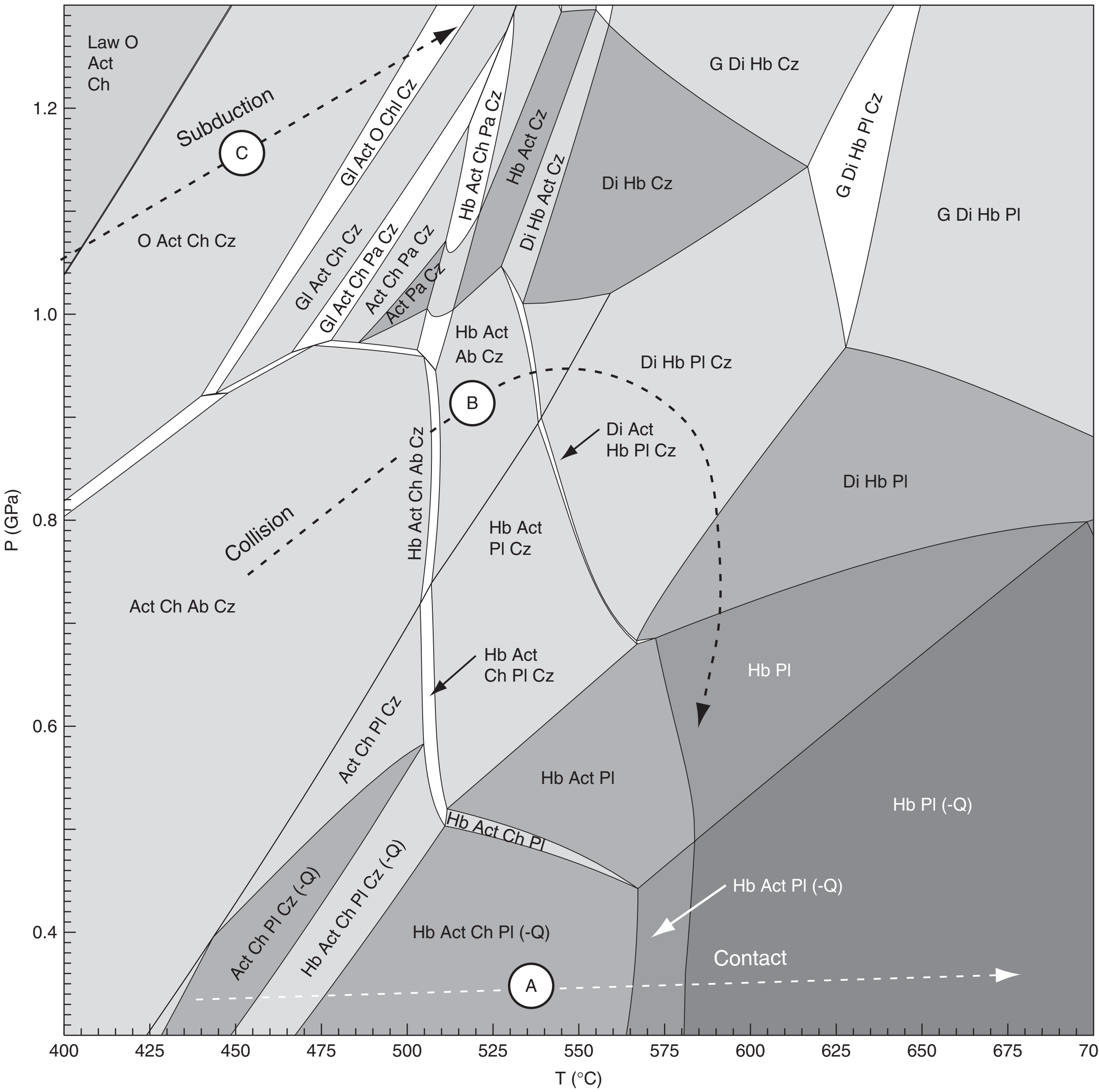


Fig. 19.13 Abbreviations: Ab, albite; Act, actinolite; Ch, chlorite; Di, diopside (clinopyroxene); garnet; Gl, glaucophane; Hb, hornblende; Law, lawsonite; O, omphacite; Pa, paragonite; Pl, plagioclase; Q, quartz. (Pseudosection from Diener *et al*., 2007; reproduced with permission of Blackwell.)

19.9 Figure 19.13 illustrates a pseudosection for a typical mid-ocean ridge basalt bulk composition in the system Na_2O-CaO-FeO-MgO-Al_2O_3-SiO_2-H_2O (NCFMASH). The mineral assemblages coexist with quartz and H_2O with unit activity, except in those stability fields marked (-Q) which lack quartz.

(a) The five different gray shadings represent mineral assemblages having different degrees of freedom. Use the phase rule to compute the degrees of freedom for each of the five types of stability field, making sure to give a specific example of each. For this case of unit water activity (water is present as a phase coexisting with the minerals), do the standard phase rule (Eq. (10.11)) and the modified phase (Eq. (18.2)) rule give the same result?

(b) The *P*–*T* trajectories labeled with arrows represent different possible metamorphic regimes. Trajectory A is appropriate for contact metamorphism, B for continental collision, and C for subduction. List the sequences of mineral assemblages developed during prograde heating along each trajectory. Do you think that the mineralogical progressions along each trajectory are distinct enough so that you could distinguish them in the field?

20 Mineral reactions involving H_2O and CO_2

20.1 INTRODUCTION

Most prograde metamorphic reactions involve dehydration or decarbonation. The large increase in entropy that accompanies the liberation of a volatile phase from a mineral ensures that rising metamorphic temperatures will favor reactions that produce a separate vapor phase. The properties of this phase are critical in determining which metamorphic reactions take place and under what conditions they occur. We have already seen in the simple example of the model metamorphic terrane treated in Chapter 18 that very different conclusions about the conditions of metamorphism can be reached depending on the assumptions made about the composition of the fluid phase (Fig. 18.4). The purpose of this chapter is to outline some of the important principles governing metamorphic reactions that involve a volatile phase.

The fluid phase in most metamorphic rocks is dominated by H_2O and CO_2. Both of these are initially derived almost entirely from the atmosphere (meteoric). Water is incorporated by minerals, such as the clays, during the weathering of rocks or the diagenesis of sediments. Carbon dioxide may also be similarly incorporated with the formation of calcite or dolomite. But the largest amount of CO_2 enters metamorphic rock as calcite of biological origin formed from the shells of organisms. Both H_2O and CO_2 are trapped directly as pore fluid, but during compaction and diagenesis of sediment most of this is expelled. The fluid phase in metamorphic rocks is therefore derived largely from the breakdown of minerals rather than from the initially trapped pore fluid. The composition of the fluid can thus vary considerably depending on the composition of the host rock.

Under typical metamorphic conditions, the fluid is above its critical point; that is, it is above the temperature and pressure at which liquid and gas can be distinguished. The critical point for water, for example, is at 374.1 °C and 22.12 MPa, and that of carbon dioxide is at 31 °C and 7.38 MPa. Thus, at the bottom of a 2.26-km-deep ocean, erupting lava would heat water, but it could not cause it to boil. In pore fluids, water is above its critical point at depths greater than about 2.5 km. These supercritical fluids are more compressible than minerals. Pressure therefore causes considerable changes in the ΔV of reactions involving fluids, which in turn affects the temperatures at which the reactions take place. To determine the volume of a fluid as a function of temperature and pressure it is necessary to have what is known as an equation of state. The ideal gas law ($PV = nRT$) is one such equation, but under the conditions prevailing during metamorphism, deviations from this law are large, and therefore other equations must be used.

The composition of the fluid is critical in determining metamorphic mineral equilibria. In Chapters 16 and 18 we saw how a change in the activity of water in the fluid dramatically affects the P–T conditions under which a dehydration reaction takes place. Similarly, the activity of CO_2 in the fluid affects decarbonation reactions. Many metamorphic reactions involve H_2O and CO_2 as reactants and/or as products. For these reactions, variations in the composition of the fluid can completely change the sequence of appearance of minerals with increasing temperature. It is therefore necessary to know how the activities of these components vary with composition in these fluids and how they affect mineral reactions.

20.2 *P–V–T* BEHAVIOR OF FLUIDS

Equation (8.10) describes how changes in the fugacity of a fluid affect the free-energy change of a reaction involving condensed phases and a gas. That equation was used in Problem 8.8 to calculate the dehydration reaction of quartz plus muscovite to produce sillimanite plus K-feldspar. For simplicity, water was assumed to behave ideally for this calculation. Because of the importance of this reaction, and others like it, we now need to make the calculation taking into account that water does not behave ideally under metamorphic conditions.

In deriving Eq. (8.10) it is necessary to evaluate the integral of $V\,dP$ at a fixed temperature, which requires a knowledge of the functional relation between V and P at a given temperature. An equation relating mass, volume, temperature, and pressure is known as an *equation of state*. The simplest such equation is the *ideal gas law*, which for one mole of gas is

$$\bar{V} = \frac{RT}{P} \tag{20.1}$$

which states that the product of the molar volume ($\bar{V}$) and pressure (P) is a constant for a given absolute temperature T. The constant of proportionality R, the *gas constant*, can be expressed in many different units (see Table 20.1). The great value of the ideal gas law lies in its independence of the composition of the gas.

Table 20.1 *Constants used in equations of state for gases*

Gas constant $R = 8.3144\ \mathrm{J\ mol^{-1}\ K^{-1}}$ (SI units)
$8.3144\ \mathrm{Pa\ m^3\ mol^{-1}\ K^{-1}}$ (SI units)
$83.144\ \mathrm{bar\ cm^3\ mol^{-1}\ K^{-1}}$
$8.3144 \times 10^7\ \mathrm{ergs\ mol^{-1}\ K^{-1}}$
$1.9872\ \mathrm{cal\ mol^{-1}\ K^{-1}}$
$0.082054\ \mathrm{liter\ atm\ mol^{-1}\ K^{-1}}$

Coefficients in van der Waals equation (Eq. (20.4))[a]

	a (Pa m^6 mol^{-2})	b (m^3 mol^{-1})
H_2O	0.5537	3.049×10^{-5}
CO_2	0.3640	4.267×10^{-5}

Coefficients in modified Redlich–Kwong equation of Kerrick and Jacobs (1981)[a]
For H_2O
a see text
$b = 2.9 \times 10^{-5}\ \mathrm{m^3\ mol^{-1}}$
$c = [290.78 - (0.30276 \times T) + (1.4774 \times 10^{-4} \times T^2)] \times 10^{-1}$
$d = [-8374 + (19.437 \times T) - (8.148 \times 10^{-3} \times T^2)] \times 10^{-7}$
$e = [76600 - (133.9 \times T) + (0.1071 \times T^2)] \times 10^{-13}$

For CO_2
a see text
$b = 5.8 \times 10^{-5}\ \mathrm{m^3\ mol^{-1}}$
$c = [28.31 + (0.10721 \times T) - (8.81 \times 10^{-6} \times T^2)] \times 10^{-1}$
$d = [9380 - (8.53 \times T) + (1.189 \times 10^{-3} \times T^2)] \times 10^{-7}$
$e = [-368654 + (715.9 \times T) + (0.1534 \times T^2)] \times 10^{-13}$

[a] All values in equations are for SI units.

Although many gases approach ideal behavior at low pressures, deviations from ideality become large at high pressures. At low pressures, nonideal gases can have larger or smaller molar volumes than would be predicted from the ideal gas law, but at high pressures the molar volumes are larger than predicted. Smaller volumes indicate that attractive forces between molecules must overcome the repulsive ones due to normal thermal vibrations. Larger volumes, on the other hand, indicate the existence of repulsive forces.

An obvious flaw in the ideal gas law is the prediction that the volume of a gas decreases to zero at the absolute zero of temperature. But gases must have a finite volume at absolute zero, and most, of course, liquefy well above this and then their volume changes very little with falling temperature. The ideal gas law can be modified by assigning a value of b to the finite volume that would exist at absolute zero. The equation then becomes

$$\bar{V} = b + \frac{RT}{P} \tag{20.2}$$

Although this equation can account for volumes greater than those predicted by the ideal gas law, it cannot account for those that are smaller because b is positive. In the latter case, the pressure that causes the gas to expand must be diminished by an attractive force between the molecules, which is known as van der Waals force. The pressure must therefore be reduced by this amount. To do this we first rearrange Eq. (20.2)

$$P = \frac{RT}{\bar{V} - b} \tag{20.3}$$

From this we subtract an amount $a/\bar{V}^2$, where a is a constant, and the $\bar{V}^2$ term is introduced to account for the fact that attractive forces balance out within the gas and produce an imbalance only on the outer walls of the volume. The resulting equation is

$$P = \frac{RT}{\bar{V} - b} - \frac{a}{\bar{V}^2} \tag{20.4}$$

which is known as *van der Waals equation*. Values of the constants and b for H_2O and CO_2 are given in Table 20.1.

Although van der Waals equation provides an adequate fit of calculated *P–V–T* relations to experimental data for many gases at low to moderate pressures and temperatures, under the more extreme conditions encountered during metamorphism, there are significant deviations. Consequently, van der Waals equation has been modified to improve its fit to the data. One such modified version is the Redlich–Kwong equation,

$$P = \frac{RT}{\bar{V} - b} - \frac{a}{\sqrt{T}\bar{V}(\bar{V} + b)} \tag{20.5}$$

which differs from van der Waals equation only in the formulation of the term for the attractive force, which is made a function of temperature. This equation has been further modified to give a still better fit for the high-pressure data for H_2O and CO_2; this is the modified Redlich–Kwong equation of Kerrick and Jacobs (1981):

$$P = \frac{RT(1 + y + y^2 - y^3)}{\bar{V}(1 - y)^3} - \frac{a}{\sqrt{T}\bar{V}(\bar{V} + b)} \tag{20.6}$$

where $y = b/4\bar{V}$, and a, which is both a function of pressure and temperature is given by

$$a = c + \frac{d}{\bar{V}} + \frac{e}{\bar{V}^2} \tag{20.7}$$

where the value of $\bar{V}$ is in cubic meters, and the values of *c, d*, and *e* are given in Table 20.1. Note that Eq. (20.6) reduces to the ideal gas law at high temperatures where the volumes are large, because the second term on the right-hand side approaches zero and y also approaches zero. Equation (20.6) provides good agreement between calculated and measured molar volumes up to 0.8 GPa and between 300 and 925 °C for water and 400 to 700 °C for CO_2, with the poorest fit being between 400 and 500 °C for water. The equation can be used to extrapolate to more extreme conditions with reasonable accuracy (approximately several percent). Holland and Powell (1998) discuss alternative expressions for water and CO_2 valid to well over 1 GPa.

Once equations of state are known, the integral of $\bar{V}\,dP$ in Eq. (8.4) can be evaluated and the free-energy change of a reaction involving a gas under any desired conditions can be determined. If the equation of state is the ideal gas law, RT/P is substituted for and the integral gives $RT \ln P$ (Eq. (8.6)).

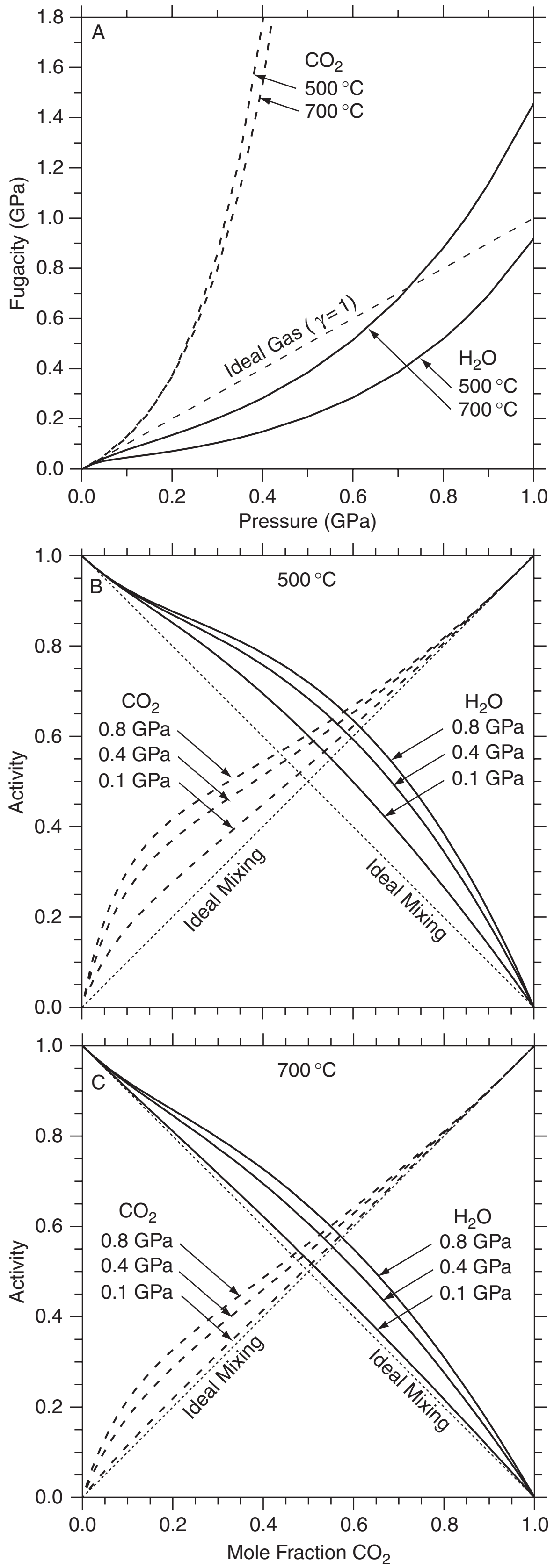


Fig. 20.1 (**A**) Plot of the fugacities of H_2O and CO_2 as a function of pressure at 500 and 700 °C based on the modified Redlich–Kwong equation of Kerrick and Jacobs (1981). Ideal gas behavior is shown by the line marked with a fugacity coefficient of $\gamma = 1$. (**B**, **C**) Activity–composition relations for H_2O–CO_2 fluids at 0.1, 0.4, and 0.8 GPa for 500 °C (part (B)) and 700 °C (part (C)). Ideal mixing shown with thin dotted lines; nonideal relations calculated following Kerrick and Jacobs (1981). Note that fluid behavior is closest to ideal at high temperatures and low pressures.

For real gases, however, it is necessary to integrate more complicated expressions, such as the modified Redlich–Kwong equation. This will not be done here, but for those interested, the integration is given by Kerrick and Jacobs (1981).

It will be recalled from Eq. (8.7) that a function known as the fugacity, *f*, was introduced so that the simple form of the integral of $\bar{V}\,dP$ for ideal gases could be retained for nonideal ones; that is, the integral is $RT \ln f$. By equating the actual integral of $\bar{V}\,dP$ obtained using the equation of state with $RT \ln f$, a fugacity coefficient, γ ($f = \gamma P$, Eq. (8.8)), is obtained. Kerrick and Jacobs (1981) give the following expression for the fugacity coefficients of H_2O and CO_2 using the values for the constants given in Table 20.1:

$$\begin{aligned}\ln\gamma = {} & \frac{8y-9y^2+3y^3}{(1-y)^3} - \ln Z - \frac{c}{RT^{3/2}(\bar{V}+b)} \\ & - \frac{d}{RT^{3/2}\bar{V}(\bar{V}+b)} - \frac{e}{RT^{3/2}\bar{V}^2(\bar{V}+b)} \\ & + \left(\frac{c}{RT^{3/2}b}\cdot\ln\frac{\bar{V}}{\bar{V}+b}\right) - \frac{d}{RT^{3/2}b\bar{V}} \\ & + \left(\frac{d}{RT^{3/2}b^2}\cdot\ln\frac{\bar{V}+b}{\bar{V}}\right) - \frac{e}{RT^{3/2}2b\bar{V}^2} \\ & + \frac{e}{RT^{3/2}b^2\bar{V}} - \left(\frac{e}{RT^{3/2}b^3}\cdot\ln\frac{\bar{V}+b}{\bar{V}}\right) \qquad (20.8)\end{aligned}$$

where *Z*, the *compressibility factor*, is given by $P\bar{V}/RT$. Despite the length of this equation, it provides a simple means of calculating the fugacity coefficient for any desired molar volume and temperature (Problem 20.2).

Fugacities can be calculated from the fugacity coefficients for any given pressure. Figure 20.1(A) shows the calculated fugacities of H_2O and CO_2 at 500 and 700 °C up to a pressure of 1 GPa (Problem 20.2). Water and CO_2 behave quite differently. Below a pressure of 0.73 GPa at 700 °C, the fugacity of water is less than the pressure of water, that is, $\gamma_{H_2O} < 1$. Above this pressure, however, the reverse is true, so $\gamma_{H_2O} > 1$. The pressure at which $\gamma_{H_2O} = 1$ decreases with rising temperature. At 1000 °C, for example, it occurs at 0.4 GPa.

The fugacity of CO_2, on the other hand, is invariably greater than the pressure of CO_2, and markedly so at high pressures. For example, at 1 GPa and 500 °C, the fugacity of CO_2 is 62 GPa; this would plot off-scale on Figure 20.1(A). Such large deviations from ideality significantly affect mineral equilibria (Problem 20.3). A decarbonation reaction proceeding at 500 °C and 1 GPa would “experience” a fugacity (effective pressure) of CO_2 of 62 GPa. This means that the equilibrium temperature is much higher than would be expected from ideal behavior.

Equations (20.6) and (20.8) allow us to calculate the fugacities of pure H_2O and CO_2 over a wide range of conditions. But because metamorphic fluids may contain both H_2O and CO_2, it is necessary to know the fugacities of these gases in mixtures. At certain pressure–temperature conditions, some gases form ideal solutions, even though separately they do not obey the ideal gas law. For such gases, the fugacity of a component is given by

$$f_i = X_i F_i = X_i \gamma_i P \tag{20.9}$$

where X_i is the mole fraction of component i and F_i is the fugacity that component i would have if pure and at the same temperature and total pressure as the mixture. Because the mixing is ideal, the activity of i is equal to its mole fraction. This simple relation, known as the Lewis and Randall rule, allows us to obtain approximate values for the fugacities in mixed gases. The mixing of water and CO_2, for example, approaches ideality at low pressures and high temperatures.

In general, however, the mixing is nonideal and we write

$$f_i = a_i F_i = \lambda_i X_i \gamma_i P$$

where a_i is the activity of species i and λ_i is the corresponding activity coefficient ($\lambda_i = 1$ for ideal mixing). As shown in Figure 20.1(B) and (C), the degree of nonideality in mixing tends to increase with increasing pressure and decrease with increasing temperature. Furthermore, the value of the activity coefficient depends critically on composition. For example, λ_{CO_2} is 3 in a water-rich fluid with $X_{CO_2} = 0.1$ at 500 ºC and 0.8 GPa, but it is near unity in a CO_2-rich fluid at the same conditions (Fig. 20.1(B)). The deviations from ideality are positive for both H_2O and CO_2 (see Fig. 9.3) so the mixing is endothermic. Qualitatively, this means that H_2O–H_2O and CO_2–CO_2 interactions are energetically more favorable than H_2O–CO_2 interactions. Consequently, there is a tendency for like species to cluster together in the fluid. Kerrick and Jacobs (1981) provide expressions for nonideal mixing behavior, as do others, including Shi and Saxena (1992) and Aranovich and Newton (1999).

20.3 CARBONACEOUS ORGANIC MATTER

Many sediments contain carbonaceous organic matter. As temperature rises, this material becomes increasingly crystalline, ultimately forming graphite at intermediate to high metamorphic grades (Wopenka and Pasteris, 1993). When water reacts with the reduced carbon, redox reactions can occur involving C–O–H species including CO_2, CH_4, and H_2 (Shi and Saxena, 1992).

Water, CO_2, and CH_4 are generally the most abundant fluid species, but their proportions depend critically on the redox state of the system. For example, consider fluid speciation for relatively low oxygen fugacities corresponding to an order of magnitude (one "log unit") smaller than those of the quartz fayalite magnetite buffer. This oxygen fugacity is commonly referred to as QFM-1. Calculations for 0.5 GPa indicate that fluids equilibrated with graphite are mostly H_2O and CH_4 at lower temperatures, and that the proportion of CO_2 increases considerably as temperature increases (Fig. 20.2). For higher oxygen fugacities of QFM-0.5, methane is less abundant, and the fluid consists almost entirely of H_2O and CO_2 above 500 °C (Fig. 20.2). At T greater than about 600 °C, the proportion of CO_2 increases dramatically. Note that the fluid contains CO_2, even though carbonate minerals need not be present. These results indicate that methane may be important for rocks rich in organic matter reacting at low oxygen fugacities, particularly at low metamorphic grades. The assumption of equilibrium with graphite should be viewed with caution, however, because the prograde conversion of amorphous organic matter to graphite is kinetically sluggish and incomplete at low to intermediate metamorphic grades.

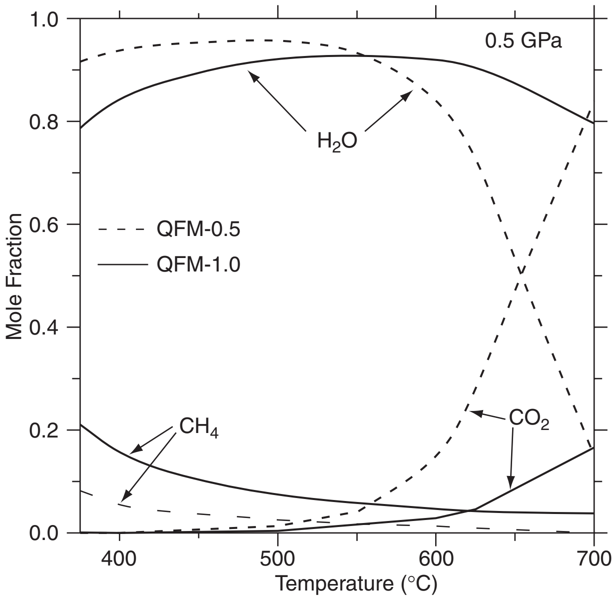


Fig. 20.2 Major species in C–O–H fluids equilibrated with graphite calculated for a total pressure of 0.5 GPa. QFM-1 denotes calculations for oxygen fugacities one order of magnitude (a "log unit") less than the quartz–fayalite–magnetite buffer. QFM-0.5 denotes more oxidizing conditions at oxygen fugacities 0.5 log units lower than QFM.

Although C species like methane are undoubtedly important under some metamorphic conditions, much can be learned from H_2O–CO_2 fluids, and these will be the focus of the remainder of this chapter.

20.4 METAMORPHOSED SILICEOUS CARBONATE ROCKS

Metamorphosed impure limestones and dolomites have long attracted the attention of petrologists because of the large number of mineral assemblages that can form in such a relatively simple system. These rocks can be represented essentially in terms of three components, CaO, MgO, and SiO_2, and a CO_2–H_2O fluid phase. Alumina may also be a component if the carbonate rocks initially contain clay minerals.

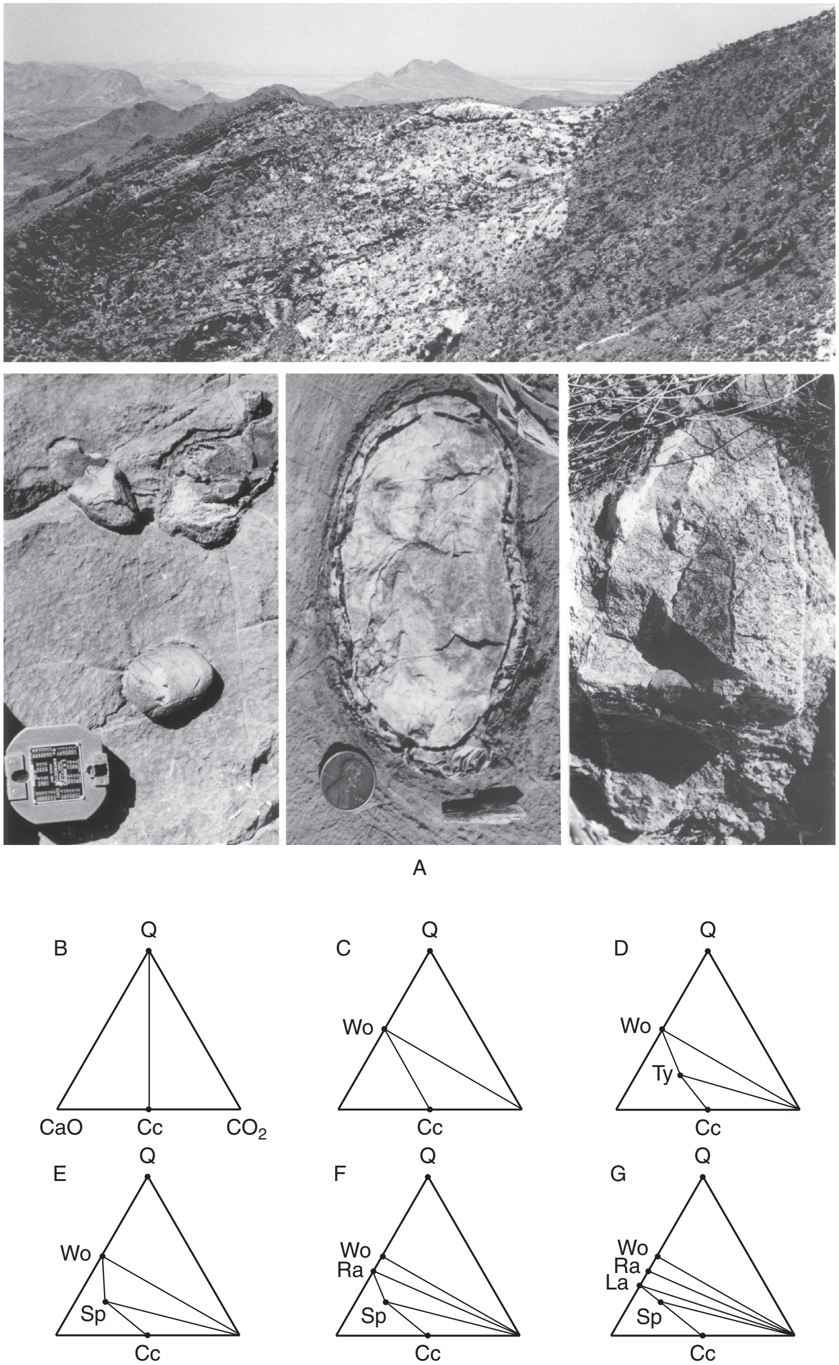


Fig. 20.3 (**A**) Contact metamorphic aureole of Christmas Mountains Gabbro, Big Bend Region, Texas (Joesten and Fisher, 1988). Heat from the gabbro (on right) recrystallized and whitened limestone out to 50 m from contact. Chert nodules in limestone reacted with enveloping calcite to produce concentric layers of calcsilicate minerals. (**B**) Unmetamorphosed chert nodules in limestone 132 m from contact. The mineral facies diagram below shows the stable assemblage of calcite (Cc) + quartz (Q) + fluid (CO_2) in this unmetamorphosed rock. (**C**) Chert nodule rimmed by wollastonite (Wo), which formed by reaction between quartz and calcite at 100 m from contact. (**D**) From 23 to 15 m from contact a reaction zone of tilleyite (Ty) formed between calcite and wollastonite. (**E**) Within 13 m of the contact, tilleyite became unstable and spurrite (Sp) formed instead. (**F, G**) Had the temperature risen higher and the rock been of appropriate composition, two more reactions would have occurred, producing rankinite (Ra) and larnite (La). (Photographs by R. L. Joesten; published by permission of the Geological Society of America.)

The simplest mineralogy is found in metalimestones (= marble) that contain silica. One of the most common modes of occurrence of silica in such rocks is as chert nodules. The nodules are particularly useful because they form recognizable objects that can be traced through various metamorphic grades, making it possible to study the mechanisms of successive reactions. Figure 20.3 shows the progression of reaction rims around chert nodules in the contact metamorphic aureole of the Christmas Mountains gabbro, Big Bend region, Texas (Joesten, 1974). Near the intrusion, the limestone changes from gray to white as a result of recrystallization (Fig. 20.3(A)). The first metamorphic mineral to form is wollastonite, which appears as a rim separating chert from calcite (Fig. 20.3(C)). Quartz and calcite in the unmetamorphosed limestone (Fig. 20.3(B)), on being heated, react to form wollastonite and liberate CO_2. Nearer the intrusion, wollastonite and calcite are no longer stable together and react to form tilleyite, with a further liberation of CO_2 (Fig. 20.3(D)). Still nearer the contact, tilleyite and wollastonite react to form spurrite, and more CO_2 is liberated (Fig. 20.3(E)). At still higher temperatures, the minerals rankinite and larnite might also have formed had the bulk composition of the rock been more siliceous (Fig. 20.3(F) and (G)).

The reactions taking place between chert and limestone all involve the generation of a CO_2 vapor. They are as follows:

$$\begin{array}{ccccccc} CaCO_3 & + & SiO_2 & = & CaSiO_3 & + & CO_2 \\ \text{(calcite)} & & \text{(quartz)} & & \text{(wollastonite)} & & \text{(vapor)} \end{array} \tag{20.10}$$

$$\begin{array}{ccccccc} 3CaCO_3 & + & 2CaSiO_3 & = & Ca_3Si_2O_7 \cdot 2CaCO_3 & + & CO_2 \\ \text{(calcite)} & & \text{(wollastonite)} & & \text{(tilleyite)} & & \text{(vapor)} \end{array} \tag{20.11}$$

$$\begin{array}{ccc} Ca_3Si_2O_7 \cdot 2CaCO_3 & + & CaSiO_3 \\ \text{(tilleyite)} & & \text{(wollastonite)} \\ = Ca_4Si_2O_8 \cdot CaCO_3 & + & CO_2 \\ \text{(spurrite)} & & \text{(vapor)} \end{array} \tag{20.12}$$

The effect of metamorphism is therefore to transfer CO_2 progressively from the minerals to the vapor phase. Apart from small amounts of fluid trapped in minerals as fluid inclusions or on grain boundaries, the large volumes of CO_2 produced by the reactions above are no longer present in the rocks. Therefore, another effect of metamorphism is to cause decarbonation.

If the carbonate rocks also include dolomite, the assortment of metamorphic minerals that can form is still larger. The composition of these minerals, projected into the ternary diagram $CaO–MgO–SiO_2$, is shown in Figure 20.4. Although all of these minerals may occur in metacarbonate rocks, the bulk composition of sedimentary carbonates is normally restricted to the left half of the diagram; that is, these rocks initially consist of a mixture of calcite, dolomite, and quartz. Metamorphosed ultramafic igneous rocks, however, have compositions in the right half of the diagram. With so many minerals, the number of three-phase assemblages that can be drawn is large, and working out which are possible and in what sequence they should appear with increasing metamorphic grade is no small task. In 1940, Bowen proposed that these mineral assemblages are formed by 13 decarbonation reactions, which lead to the appearance of 10 minerals in a definite order; from low to high temperature they are tremolite, forsterite, diopside, periclase, wollastonite, monticellite, akermanite, spurrite, merwinite, and larnite. He suggested the following mnemonic to help remember the sequence:

> Tremble, for dire peril walks,
> Monstrous acrimony's spurning mercy's laws.
> (Bowen, 1940)

Subsequently, additional reactions have been suggested. Tilley (1948), for example, found that talc appeared before tremolite. Bowen, on being asked by Tilley what this would do to the mnemonic, suggested stuttering at the beginning – *ta … tremble, for dire ….*

Not only have additional reactions been proposed, but Bowen's actual sequence of reactions has been shown to be only one of many. Indeed, diopside almost everywhere appears before forsterite. The cause for the different sequences is obvious when the reactions are studied in terms of a mixed $CO_2–H_2O$ fluid phase, but discussion of this aspect is postponed to the next section.

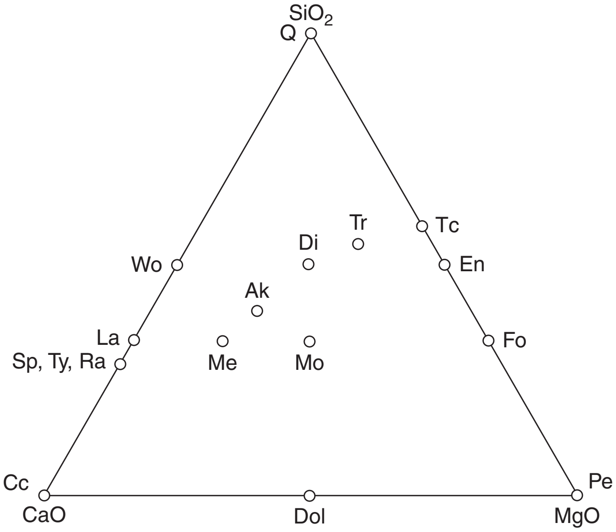


Fig. 20.4 Minerals formed by the metamorphism of siliceous limestone and dolomite plotted in terms of mole proportions of $CaO–MgO–SiO_2$. Talc (Tc) and tremolite (Tr) contain H_2O, and calcite (Cc), dolomite (Dol), spurrite (Sp), and tilleyite (Ty) contain CO_2. The other minerals are enstatite (En), forsterite (Fo), periclase (Pe), diopside (Di), monticellite (Mo), akermanite (Ak), merwinite (Me), wollastonite (Wo), larnite (La), rankinite (Ra), and quartz (Q).

Despite the limited applicability of Bowen's decarbonation series, his 1940 paper was an important landmark, because it contained the first proposal for the construction of a petrogenetic grid (Evans, 2007a). He visualized that a *P–T* grid would be formed by the intersection of the univariant decarbonation reactions with reactions involving only solid phases, which would likely have different slopes. Subsequent work has shown this to be the case, and today detailed grids have been worked out in terms, not only of pressure and temperature (Chapter 18), but also fluid-phase composition (e.g. Skippen, 1974).

For the moment, we will consider only one of Bowen's decarbonation reactions, leaving the others to the next section. The reaction of quartz and calcite to form wollastonite and CO_2 (Eq. 20.10) provides a good example of one of the many reactions in which only CO_2 is liberated. The free-energy change of the reaction at any given pressure and temperature can be expressed by Eq. (8.9), that is,

$$\Delta G^P_{r,T} = \Delta G^0_{r,T} + \Delta V_{r,T,c}(P - 10^5) + RT \ln \frac{f}{10^5}$$

where P and f are measured in pascal, making it necessary to subtract the pressure of the reference state (atmospheric pressure = 10^5 Pa). Note that $\Delta V_{r,T,c}$ is for the condensed phases only, which in this reaction is -1.9692×10^{-5} m^3 mol^{-1}. For fugacity, we can substitute γP. The equation then becomes

$$\Delta G^P_{r,T} = \Delta G^0_{r,T} + \Delta V_{r,T,c}(P - 10^5) + RT \ln \frac{\gamma P}{10^5} \tag{20.13}$$

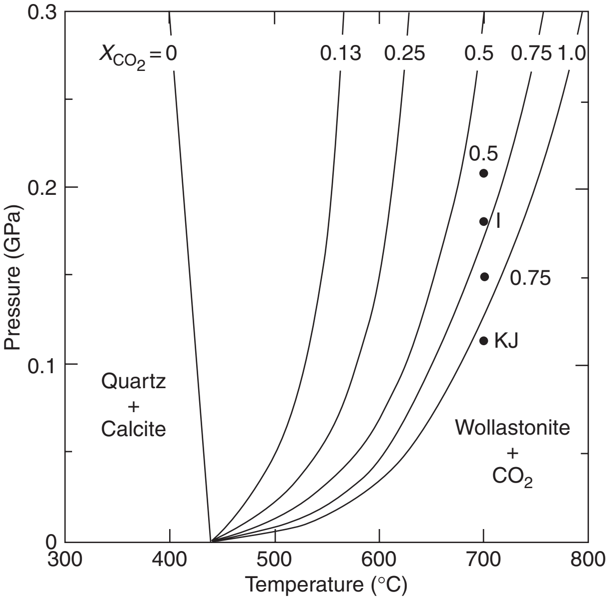


Fig. 20.5 Experimentally determined equilibria for the reaction quartz + calcite = wollastonite + CO_2 for fluid compositions ranging from pure CO_2 ($X_{CO_2} = 1$) to nearly pure water ($X_{CO_2} \approx 0$). (Data from Greenwood, 1967b, and Harker and Tuttle, 1956.) The four points refer to calculated equilibria for the reaction at 700 °C (see text for discussion). Point KJ is based on the fugacity data of Kerrick and Jacobs (1981); point I assumes ideal behavior of CO_2; points 0.75 and 0.5 are calculated using the Kerrick–Jacobs data for mole fractions of CO_2 in the fluid of 0.75 and 0.5.

At 700 °C and atmospheric pressure, for example, $\Delta G^P_{r,T} = \Delta G^0_{r,T} = -57\ 818\ \mathrm{J\,mol^{-1}}$ based on the data in Table 7.1. Because $\Delta G^P_{r,T}$ is negative, the reaction proceeds under these conditions and generates CO_2. To reestablish equilibrium at this temperature, $\Delta G^P_{r,T}$ must be reduced to zero by increasing the pressure ($(\partial \Delta G/\partial P)_T = \Delta V$, which is positive). When $\Delta G^P_{r,T} = 0$, then,

$$-\Delta G^0_{r,T} = \Delta V_{r,T,c}(P - 10^5) + RT \ln \frac{\gamma_{CO_2} P}{10^5} \qquad (20.14)$$

The value of γ_{CO_2} (abbreviated as γ for simplicity) depends on P, which is yet to be determined. We solve the equation by estimating a value for γ and then calculating P. If the chosen value of γ does not match with the determined pressure, a new value is selected. After several iterations, matching values of γ and P are found. For the conditions considered here (700 °C), a solution is obtained for a pressure of 0.117 GPa, a fugacity of 0.161 GPa, and a fugacity coefficient of 1.37. This point lies very near the experimentally determined equilibrium curve for this reaction (KJ in Fig. 20.5). Note that had CO_2 been assumed to behave ideally, the calculated equilibrium point would have plotted well above the experimentally determined curve at 0.183 GPa (I in Fig. 20.5).

This decarbonation reaction resembles a dehydration reaction, in that it has a positive slope that is shallow at low pressures and steepens with increasing pressure as the evolved vapor phase becomes less compressible. Because the fugacity coefficient of CO_2 is invariably greater than 1 and becomes progressively larger with increasing pressure (Fig. 20.1(A)), the "effective pressure" (fugacity) of the CO_2 "experienced" by the reaction is greater than the actual pressure. As a result, the increased temperatures required to drive the reaction at higher pressures are greater than those for dehydration reactions, for which the fugacity coefficient of H_2O is commonly less than 1. The slopes of decarbonation reactions on P–T diagrams are consequently shallower than those of dehydration reactions.

In calculating this decarbonation reaction, the vapor phase has been assumed to be pure CO_2. If it is not, then, according to Le Chatelier's principle, the reaction proceeds at lower temperatures. Equation (20.14) is still applicable, except that higher pressures are required at any given temperature to generate the equilibrium fugacities, because the vapor is not pure CO_2. If the fugacity of CO_2 in a mixed vapor phase obeys Eq. (20.9), the activity of CO_2 is given by its mole fraction, that is, $a_{CO_2} = X_{CO_2}$. Thus, according to Eq. (20.9), the fugacity of CO_2 in a mixed vapor is $X_{CO_2} F_{CO_2}$, where F_{CO_2} is the fugacity CO_2 would have if it were pure and under the same total pressure, P, as the system. Equation (20.14) can thus be written as

$$-\Delta G^0_{r.T} = \Delta V_{r,T,c}(P - 10^5) + RT \ln \frac{X_{CO_2} \gamma P}{10^5} \qquad (20.15)$$

where γ is the fugacity coefficient CO_2 would have if it were pure and at pressure P. This equation, then, can be used to calculate the equilibrium pressure of a decarbonation reaction for any given composition of the fluid phase, provided that the fugacity of CO_2 obeys Eq. (20.9).

Consider again, for example, the reaction of quartz and calcite to produce wollastonite. The fugacity of CO_2 at 700 °C was determined to be 0.161 GPa when the vapor is pure CO_2. This same fugacity is required for equilibrium when the activity of CO_2 is less than 1; the difference is that the fugacity must be generated as a fraction of the fugacity of pure CO_2 at a higher pressure and, therefore, with a different γ. It is necessary again to solve Eq. (20.15) iteratively for the new conditions, which gives a pressure of 0.15 GPa when $a_{CO_2} = 0.75$ and 2.1 GPa when $a_{CO_2} = 0.5$ (Fig. 20.5). Differences between the calculated and experimental data arise because of the assumption that H_2O and CO_2 mix ideally. To take account of nonideal mixing, we would use the following expression based on Eq. (20.15) and the discussion in Section 20.2

$$-\Delta G^0_{r.T} = \Delta V_{r,T,c}(P - 10^5) + RT \ln \frac{\lambda X_{CO_2} \gamma P}{10^5}$$

where λ is the activity coefficient of CO_2 in the mixture.

The experimentally determined equilibria for the reaction of quartz and calcite to form wollastonite and CO_2 are shown in Figure 20.5 for several mole fractions of CO_2 in the vapor. The maximum stability of quartz + calcite occurs when $X_{CO_2} = 1$. With decreasing CO_2 concentrations, the reaction curve steepens and has a negative slope at very low mole fractions. This means that when the fluid is almost pure H_2O, which might occur, for example, in a thin marble layer embedded in a thick pelitic sequence of rocks, increasing pressure causes wollastonite to form at progressively lower temperatures. This P–T diagram, which is typical of any decarbonation reaction, illustrates an important point; that is, the presence of a high-temperature mineral, such as wollastonite, provides little indication of actual metamorphic temperatures unless the composition of the fluid phase is known. At 0.3 GPa, for example, the temperature of appearance of wollastonite could be as low as 400 °C if $X_{CO_2} \approx 0$ or as high as 800 °C if $X_{CO_2} = 1$. In the next section we will see that certain mineral assemblages provide a record of the mole fraction of CO_2 during metamorphism.

20.5 THERMODYNAMICS OF MINERAL REACTIONS WITH $H_2O–CO_2$ FLUIDS

Dehydration and decarbonation reactions in the presence of a mixed fluid phase involve three intensive variables, *P, T*, and fluid composition, which can be conveniently represented in a three-dimensional diagram (Fig. 20.6(A)). A decarbonation reaction is plotted in this diagram for illustrative purposes. In the presence of pure CO_2, this reaction appears on the right-hand face of the diagram. This face is shown separately in Figure 20.6(B). The positions of the reaction in equilibrium with fluids containing mole fractions of CO_2 less than 1 are given by the dashed lines in the block diagram. These are lines of intersection of the reaction surface with planes parallel to the *P–T* face. The dashed lines can be projected onto the *P–T* face of the diagram to give the familiar curves for a decarbonation reaction at various mole fractions of CO_2 in the fluid phase (Fig. 20.6(B)). Isobaric (dotted lines in Fig. 20.6(A)) and isothermal sections can also be taken through the diagram (Fig. 20.6(C) and (D), respectively).

To determine the conditions under which a decarbonation (or dehydration) reaction will take place, the shape of the reaction surface in terms of *P, T*, X_{H_2O}, X_{CO_2} must be known. The surface, of course, represents the locus of points in *P–T*–fluid composition space where the free-energy change of the reaction is zero (by definition of equilibrium). By breaking the free-energy change of the reaction into its component parts and then expressing these as functions of *P, T*, and fluid composition, we generate the equation needed to describe the reaction surface. Although this derivation may appear complex because of the number of terms involved, it is, in fact, simple. The derivation is worth working through carefully because it employs many of the basic thermodynamic relations developed in Chapters 8 and 9. However, you may wish to go directly to the result, which is given in Eq. (20.16).

Consider a reaction of an assemblage of minerals, *A*, to produce a new assemblage *B* with the liberation of *m* moles of H_2O and *n* moles of CO_2.

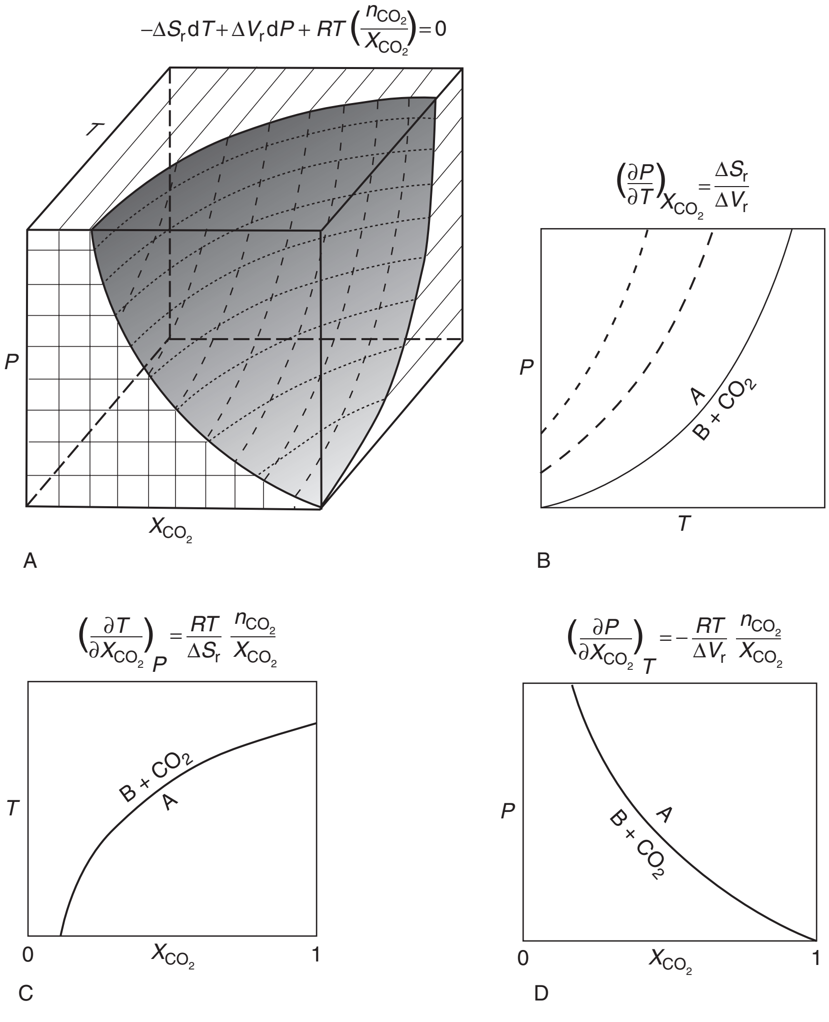


Fig. 20.6 (**A**) Plot of decarbonation reaction (A = B + CO_2) in terms of $P–T–X_{CO_2}$. The equation of the reaction surface is derived in the text (Eq. (20.16)). Sections through $P–T–X_{CO_2}$ space are shown for constant composition (**B**), constant pressure (**C**), and constant temperature (**D**).

$$A = B + mH_2O + nCO_2$$

To avoid excessive subscripting, H_2O in the fluid will be symbolized by H and CO_2 by C. At equilibrium

$$\Delta G_r = G_B + mG_H + nG_C + G_{mix} - G_A = 0$$

Note that in addition to the free energies of the various phases, a free-energy term resulting from the mixing of the fluid components is included. If the reactants and products are in equilibrium on the reaction surface at a point represented by (T^0, P^0, $X^0{}_H$, $X^0{}_C$), how will ΔG_r change with P, T, and X? In Section 1.6, the temperature at any depth in Earth was determined by expanding T in a Taylor series as a function of depth, starting from the known temperature on Earth's surface. Similarly, ΔG_r can be expanded in a Taylor series as a function of P, T, and X from the starting point $(T^0, P^0, X^0{}_H, X^0{}_C)$. Only the first two terms of the series need be used in the expansion, because eventually the expression is used to determine the conditions at a point that is separated from the first by only dP, dT, dX. In the following derivation the superscript ^{o} indicates a free energy value at the initial point. Note that $X_C = 1 - X_H$ in the binary fluid phase.

The free energy of the solids is a function only of T and P. For the solids constituting assemblage B, the Taylor series can be written as

$$G_B(T,P) = G_B^o + \left(\frac{\partial G_B}{\partial T}\right)_{P,X}(T - T^0) + \left(\frac{\partial G_B}{\partial P}\right)_{T,X}(P - P^0)$$

but $(\partial G_B/\partial T)_P = -S_B$ and $(\partial G_B/\partial P)_T = V_B$, so

$$G_B(T,P) = G_B^o - S_B(T - T^0) + V_B(P - P^0)$$

Similarly, for A,

$$G_A(T,P) = G_A^o - S_A(T - T^0) + V_A(P - P^0)$$

The free energy of the fluid is a function of T, P, and fluid composition. Also, at constant T and P, the free energy of the fluid is given by $G^f = m\mu_H + n\mu_C$. By expressing the free energy of the fluid in terms of chemical potentials, the free energy of mixing in the fluid is accounted for. We will assume ideal mixing in the fluid, so $\mu_i^f = \mu_i^* + RT\ln X_i^f$ (Eq. (9.28)). The free energy of the fluid is therefore

$$\begin{aligned} G^f(T,P,X_H,X_C) = G^{of} &+ m\left(\frac{\partial \mu_H}{\partial T}\right)_{P,X_H,X_C}(T - T^0) \\ &+ n\left(\frac{\partial \mu_C}{\partial T}\right)_{P,X_H,X_C}(T - T^0) \\ &+ m\left(\frac{\partial \mu_H}{\partial P}\right)_{T,X_H,X_C}(P - P^0) \\ &+ n\left(\frac{\partial \mu_C}{\partial P}\right)_{T,X_H,X_C}(P - P^0) \\ &+ m\left(\frac{\partial \mu_H}{\partial X_H}\right)_{T,P,X_C}(X_H - X_H^0) \\ &+ n\left(\frac{\partial \mu_C}{\partial X_C}\right)_{T,P,X_H}(X_C - X_C^0) \end{aligned}$$

Making use of Eq. (9.28), we expand the partial derivatives to give

$$\begin{aligned} m\left(\frac{\partial \mu_H}{\partial T}\right)_{P,X_H,X_C} &= m\left(\frac{\partial}{\partial T}(\mu_H^* + RT\ln X_H)\right)_{P,X_H,X_C} \\ &= m\left(\frac{\partial \mu_H^*}{\partial T}\right)_{P,X_H,X_C} \\ &\quad + m\left(\frac{\partial}{\partial T}(RT\ln X_H)\right)_{P,X_H,X_C} \end{aligned}$$

Recall from Eq. (9.2) that $\mu_H \equiv (\partial G/\partial n_H) = \bar{G}_H = \bar{H}_H - T\bar{S}_H$, which on being differentiated with respect to T gives $-\bar{S}_H$. Making this substitution yields

$$m\left(\frac{\partial \mu_H}{\partial T}\right)_{P,X_H,X_C} = -m\bar{S}_H + mR\ln X_H$$

The partial derivative with respect to P can be expanded as follows:

$$\begin{aligned} m\left(\frac{\partial \mu_H}{\partial P}\right)_{T,X_H,X_C} &= m\left(\frac{\partial}{\partial P}(\mu_H^* + RT\ln X_H)\right)_{T,X_H,X_C} \\ = m\left(\frac{\partial \bar{G}_H}{\partial P}\right)_{T,X_H,X_C} &= m\bar{V}_H \end{aligned}$$

The partial derivative with respect to X, on being expanded, gives

$$m\left(\frac{\partial \mu_H}{\partial X_H}\right)_{T,P,X_C} = m\left(\frac{\partial}{\partial X_H}(RT\ln X_H)\right)_{T,P,X_C} = \frac{mRT}{X_H}$$

For the carbon dioxide component of the fluid, we similarly obtain

$$n\left(\frac{\partial \mu_C}{\partial T}\right)_{P,X_H,X_C} = -n\bar{S}_C + nR\ln X_C$$

and

$$n\left(\frac{\partial \mu_C}{\partial P}\right)_{T,X_H,X_C} = -n\bar{V}_C$$

and

$$n\left(\frac{\partial \mu_C}{\partial X_C}\right)_{T,P,X_H} = \frac{nRT}{X_C}$$

We let $(T - T^0) = \delta T$, $(P - P^0) = \delta P$, $(X_H - X^0{}_H) = \delta X_H$, $(X_C - X^0{}_C) = \delta X_C$, and $\delta X_H = -\delta X_C$. We now combine all the terms to obtain the free energy of the fluid.

$$\begin{aligned} G^f &= G^{of} - m\bar{S}_H\delta T + (mR\ln X_H)\delta T - n\bar{S}_C\delta T \\ &\quad + (nR\ln X_C)\delta T + m\bar{V}_H\delta P + n\bar{V}_C\delta P \\ &\quad + \frac{mRT}{X_H}\delta X_H + \frac{nRT}{X_C}\delta X_C \\ &= G^{of} - (m\bar{S}_H + n\bar{S}_C)\delta T + R(m\ln X_H + n\ln X_C)\delta T \\ &\quad + (m\bar{V}_H + n\bar{V}_C)\delta P + RT\left(\frac{n}{X_C} - \frac{m}{X_H}\right)\delta X_C \end{aligned}$$

Let $(m\bar{S}_\mathrm{H} + n\bar{S}_\mathrm{C}) = S_f$ and $(m\bar{V}_\mathrm{H} + n\bar{V}_\mathrm{C}) = V_f$. Note from Eq. (9.20) that $R(m \ln X_\mathrm{H} + n \ln X_\mathrm{C}) = -S^f_\mathrm{mix}(m+n)$ because $m = X_\mathrm{H}(m+n)$ and $n = X_\mathrm{C}(m+n)$. Substituting these into the expression for G^f gives

$$G^f = G^{\circ f} - \left[S_f + S^f_\mathrm{mix}(m+n)\right]\delta T + V^f\delta P + RT\left(\frac{n}{X_\mathrm{C}} - \frac{m}{X_\mathrm{H}}\right)\delta X_\mathrm{C}$$

We now combine the expressions for the solids and the fluids to obtain ΔG_r. Note that $\Delta S_r = S_B + \left[S_f + S^f_\mathrm{mix}(m+n)\right] - S_A$, and $\Delta V_r = V_B + V_f - V_A$,

$$\Delta G_r = G_B - G_A + G^f$$

$$\begin{aligned}\Delta G_r = {} & G^\circ_B - S_B\delta T + V_B\delta P - G^\circ_A + S_A\delta T - V_A\delta P \\ & + G^{\circ f} - \left[S_f + S^f_\mathrm{mix}(m+n)\right]\delta T + V^f\delta P \\ & + RT\left(\frac{n}{X_\mathrm{C}} - \frac{m}{X_\mathrm{H}}\right)\delta X_\mathrm{C}\end{aligned}$$

$$\Delta G_r = \Delta G^\circ_r - \Delta S_r\delta T + \Delta V_r\delta P + RT\left(\frac{n}{X_\mathrm{C}} - \frac{m}{X_\mathrm{H}}\right)\delta X_\mathrm{C}$$

But for two points on the P, T, X_C surface that differ by only δP, δT, and δX_C, $\Delta G_r \longrightarrow \Delta G^\circ_r$. Therefore,

$$-\Delta S_r\delta T + \Delta V_r\delta P + RT\left(\frac{n}{X_{\mathrm{CO}_2}} - \frac{m}{X_{\mathrm{H}_2\mathrm{O}}}\right)\delta X_{\mathrm{CO}_2} = 0 \tag{20.16}$$

Equation (20.16) describes the divariant surface for any reaction evolving m moles of H_2O and n of CO_2 in terms of T, P, and X_{CO_2}, where mixing in the fluid is ideal. The slope of the reaction can be determined for any particular section through the P–T–X_fluid diagram of Figure 20.6. The three important sections are for constant composition, constant pressure, and constant temperature. At *constant composition*, $\delta X_{\mathrm{CO}_2} = 0$, and Eq. (20.16) reduces to

$$\left(\frac{\delta P}{\delta T}\right)_X = \frac{\Delta S_r}{\Delta V_r} \tag{20.17}$$

which is the Clapeyron equation (Eq. (8.3)). At *constant pressure*, $\delta P = 0$, and Eq. (20.16) reduces to

$$\left(\frac{\delta T}{\delta X_{\mathrm{CO}_2}}\right)_P = \frac{RT}{\Delta S_r}\left(\frac{n}{X_{\mathrm{CO}_2}} - \frac{m}{X_{\mathrm{H}_2\mathrm{O}}}\right) \tag{20.18}$$

Finally, at *constant temperature*, $\delta T = 0$, and Eq. (20.16) reduces to

$$\left(\frac{\delta P}{\delta X_{\mathrm{CO}_2}}\right)_T = \frac{RT}{\Delta V_r}\left(\frac{n}{X_{\mathrm{CO}_2}} - \frac{m}{X_{\mathrm{H}_2\mathrm{O}}}\right) \tag{20.19}$$

The univariant reaction in any one of these sections can be obtained by integrating the appropriate equation (Problem 20.7).

Five different types of reaction are possible involving H_2O and CO_2 (Greenwood, 1967a). These are shown in an isobaric, T–X_{CO_2} section in Figure 20.7. First is the simple decarbonation reaction ($m = 0$, $n > 0$), which has a positive slope and reaches its maximum temperature at $X_{\mathrm{CO}_2} = 1$. A dehydration reaction ($m > 0$, $n = 0$), on the other hand, has a negative slope and reaches a temperature maximum at $X_{\mathrm{H}_2\mathrm{O}} = 1$. Some reactions evolve both H_2O and CO_2 ($m > 0$, $n > 0$), in which case the reaction passes through a temperature maximum at a fluid composition equivalent to the composition of the mixture of gases evolved by the reaction. In other reactions, H_2O may be evolved while CO_2 is consumed ($m > 0$, $n < 0$). Such reactions occur at progressively higher temperatures as the mole fraction of water approaches one. Conversely, reactions that evolve CO_2 while consuming H_2O ($m < 0$, $n > 0$) occur at progressively higher temperatures as the mole fraction of CO_2 approaches one.

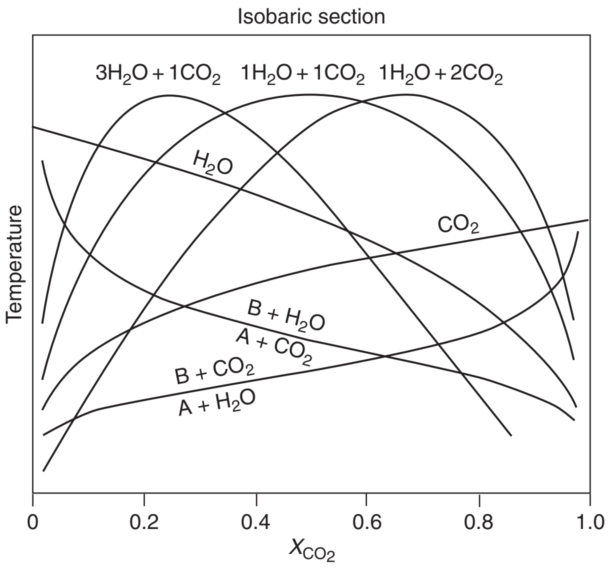


Fig. 20.7 Isobaric T–X_{CO_2} plot showing the general shape of univariant lines for reactions involving H_2O, CO_2, or both. The number of moles of H_2O and CO_2 involved in a reaction is shown on the side of the univariant line on which it is evolved. See text for discussion.

All five types of reaction have been studied experimentally (see Greenwood, 1967b, 1976; Skippen, 1974). Some examples are as follows:

$$\begin{array}{lllll} \mathrm{CaMg(CO_3)_2} + & \mathrm{SiO_2} = & \mathrm{CaMgSi_2O_6} + & \mathrm{2CO_2} & \\ \text{(dolomite)} & \text{(quartz)} & \text{(diopside)} & \text{(vapor)} & (20.20) \end{array}$$

$$\begin{array}{llll} \mathrm{Ca_2Mg_5Si_8O_{22}(OH)_2} + & \mathrm{3CaCO_3} & + \mathrm{2SiO_2} & \\ \text{(tremolite)} & \text{(calcite)} & \text{(quartz)} & \\ = \mathrm{5CaMgSi_2O_6} + & \mathrm{H_2O} + \mathrm{3CO_2} & & \\ \text{(diopside)} & \text{(vapor)} & & (20.21) \end{array}$$

$$\begin{array}{llll} \mathrm{5CaMg(CO_3)_2} + & \mathrm{8SiO_2} + & \mathrm{H_2O} & \\ \text{(dolomite)} & \text{(quartz)} & \text{(vapor)} & \\ = \mathrm{Ca_2Mg_5Si_8O_{22}(OH)_2} + & \mathrm{3CaCO_3} + & \mathrm{7CO_2} & \\ \text{(tremolite)} & \text{(calcite)} & \text{(vapor)} & (20.22) \end{array}$$

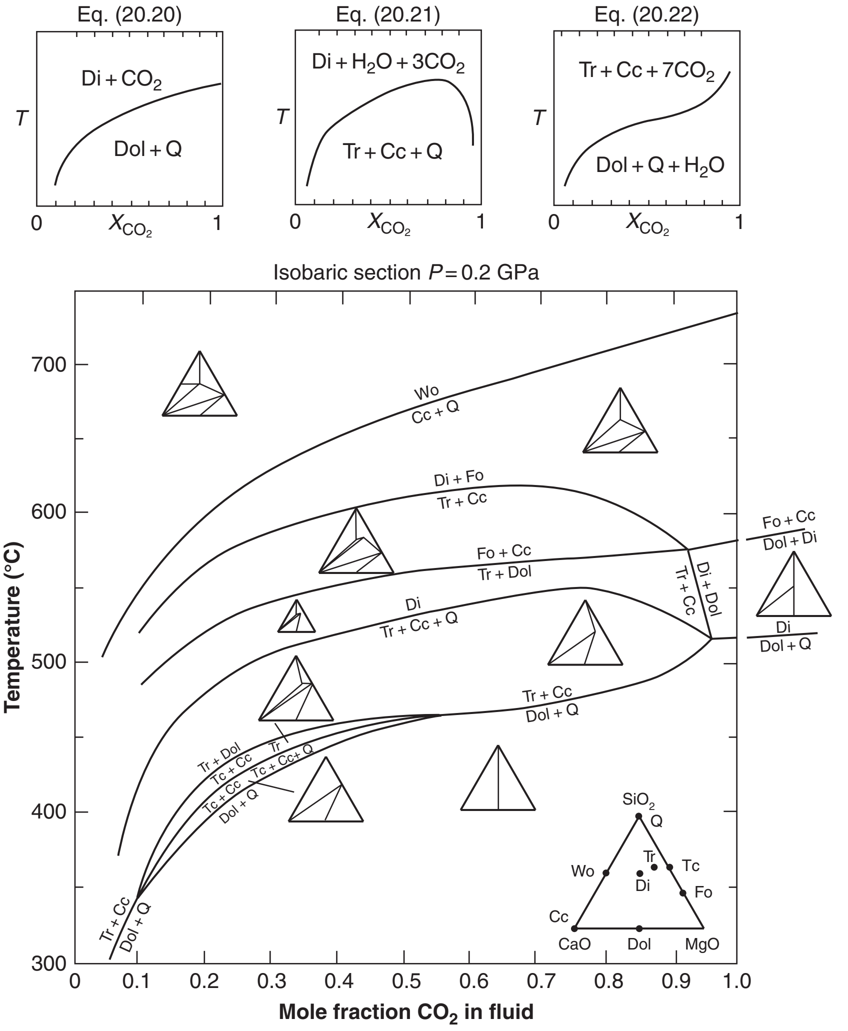


Fig. 20.8 Isobaric T–X_{CO_2} diagram for several reactions that occur in metamorphosed siliceous dolomitic limestone at a pressure of 0.2 GPa. Shown separately at the top of the diagram are three reactions, (20.20), (20.21), and (20.22). When combined, these reactions create the three lowest-temperature invariant points in Figure 20.9 (after Skippen, 1974; Greenwood, 1976) Abbrevations as in Fig. 20.4.

Each of these reactions is shown in a separate plot at the top of Figure 20.8. The reader should check that the slope of these reactions in each T–X_{CO_2} diagram is appropriate for the type of reaction.

If reactions (20.20) and (20.21) are plotted in the same diagram (Fig. 20.8), they intersect at a temperature slightly in excess of 500 °C at a fluid composition of $X_{CO_2} = 0.96$ for the pressure of this diagram (0.2 GPa). Along the isobaric univariant curve for reaction (20.20), three minerals coexist with the fluid: dolomite, quartz, and diopside. Along the curve for reaction (20.21), tremolite, calcite, quartz, and diopside coexist with the fluid. At the point of intersection, five minerals, tremolite, calcite, quartz, diopside, and dolomite, coexist. These phases can be described in terms of the three components CaO–MgO–SiO_2 (Fig. 20.4) plus a fluid phase, which is not included among the components if it is treated as an environmentally controlled intensive variable. The modified phase rule ($\phi + f = c + 3$) indicates that with five minerals, there is only one degree of freedom, which is used to define the pressure of the isobaric T–X_{CO_2} section. This assemblage of minerals therefore creates an isobaric invariant point. We can also consider the fluid explicitly and employ the "standard" phase rule ($\phi + f = c + 2$) with five components (CaO–MgO–SiO_2–H_2O–CO_2) and six phases (solids + fluid phase). Once again, we are left with one degree of freedom, which vanishes when the pressure is fixed, yielding an isobaric invariant point.

Schreinemakers rules (Section 8.5) can now be used to complete the array of univariant lines about this invariant point in terms of the two intensive variables, T and X_{CO_2}. Reaction (20.21) is the dolomite-absent reaction, which is designated [Dol]. Reaction (20.20) is the tremolite- and calcite-absent reaction ([Tr, Cc]). The other reactions that must pass through this point can be determined by noting where the minerals involved in the reactions plot in the CaO–MgO–SiO_2 diagram (Fig. 20.4). These reactions can then be tabulated according to the mnemonic scheme used in the Schreinemakers method for deriving the sequence of reactions around the invariant point.

(Tr)	(Cc)	(Q)	|Dol|	(Di)
	(Dol)	(Q)	|Tr, Cc|	(Di)
	(Cc)	(Tr)	|Di|	(Q)(Dol)
	(Tr)	(Cc)	|Q|	(Di)(Dol)

Note that only four reactions exist at this invariant point because of degeneracy caused by the collinearity of quartz, diopside, and dolomite in the CaO–MgO–SiO_2 plot.

The gas must appear on the high-temperature side of reactions (20.20) and (20.21). By labeling these two reactions in Figure 20.8, we see that only that part of reaction (20.20) ([Tr, Cc]) between the invariant point and the CO_2 side of the diagram is stable; on the water-rich side of the invariant point, the reaction is made metastable by reaction (20.21) ([Dol]). According to the mnemonic scheme, the diopside-absent reaction lies on one side of the tremolite–calcite-absent reaction, and the dolomite- and quartz-absent reactions on the other. The only arrangement of univariant lines around the invariant point that satisfies the mnemonic scheme is shown in Figure 20.8. The slopes of the lines can be determined from the balanced reactions and Eq. (20.18). The diopside-absent reaction is Eq. (20.22), which has a sigmoid shape in the T–X_{CO_2} diagram because H_2O is consumed and CO_2 is evolved. The quartz-absent reaction can be balanced as follows:

$$\underset{\text{(tremolite)}}{Ca_2Mg_5Si_8O_{22}(OH)_2} + \underset{\text{(calcite)}}{3CaCO_3} = \underset{\text{(diopside)}}{4CaMgSi_2O_6} + \underset{\text{(dolomite)}}{CaMg(CO_3)_2} + \underset{\text{(vapor)}}{H_2O + CO_2} \qquad (20.23)$$

Because this reaction produces 1 mole each of H_2O and CO_2, it rises from the invariant point to a temperature maximum at X_{CO_2} of 0.5. Once all univariant lines are correctly placed, facies diagrams are drawn in each isobaric divariant field. This serves to check that Schreinemakers rules are not violated and allows interpretation of metamorphic assemblages.

Before discussing the mineral assemblages found in the various divariant fields in Figure 20.8 we will examine the invariant point created by introducing a talc-producing reaction.

$$\underset{\text{(dolomite)}}{3CaMg(CO_3)_2} + \underset{\text{(quartz)}}{4SiO_2} + \underset{\text{(vapor)}}{H_2O} = \underset{\text{(talc)}}{Mg_3Si_4O_{10}(OH)_2} + \underset{\text{(calcite)}}{3CaCO_3} + \underset{\text{(vapor)}}{3CO_2} \qquad (20.24)$$

Intersection of this isobaric univariant reaction with reaction (20.22) produces two new invariant points involving the minerals talc, calcite, quartz, tremolite, and diopside. One invariant point occurs at 463 °C and $X_{CO_2} = 0.53$ and the other at 340 °C and $X_{CO_2} = 0.08$. The reason the univariant lines intersect at two points is that the curvature of reaction (20.22) is greater than that of reaction (20.24) (Problem 20.7). Three other reactions that must pass through these invariant points are

$$\underset{\text{(talc)}}{5Mg_3Si_4O_{10}(OH)_2} + \underset{\text{(calcite)}}{6CaCO_3} + \underset{\text{(quartz)}}{4SiO_2} = \underset{\text{(tremolite)}}{3Ca_2Mg_5Si_8O_{22}(OH)_2} + \underset{\text{(vapor)}}{2H_2O + 6CO_2} \qquad (20.25)$$

$$\underset{\text{(talc)}}{5Mg_3Si_4O_{10}(OH)_2} + \underset{\text{(calcite)}}{6CaCO_3} = \underset{\text{(tremolite)}}{Ca_2Mg_5Si_8O_{22}(OH)_2} + \underset{\text{(dolomite)}}{CaMg(CO_3)_2} + \underset{\text{(vapor)}}{H_2O + CO_2} \qquad (20.26)$$

$$\underset{\text{(talc)}}{5Mg_3Si_4O_{10}(OH)_2} + \underset{\text{(dolomite)}}{2CaMg(CO_3)_2} + \underset{\text{(quartz)}}{SiO_2} = \underset{\text{(tremolite)}}{Ca_2Mg_5Si_8O_{22}(OH)_2} + \underset{\text{(vapor)}}{4CO_2} \qquad (20.27)$$

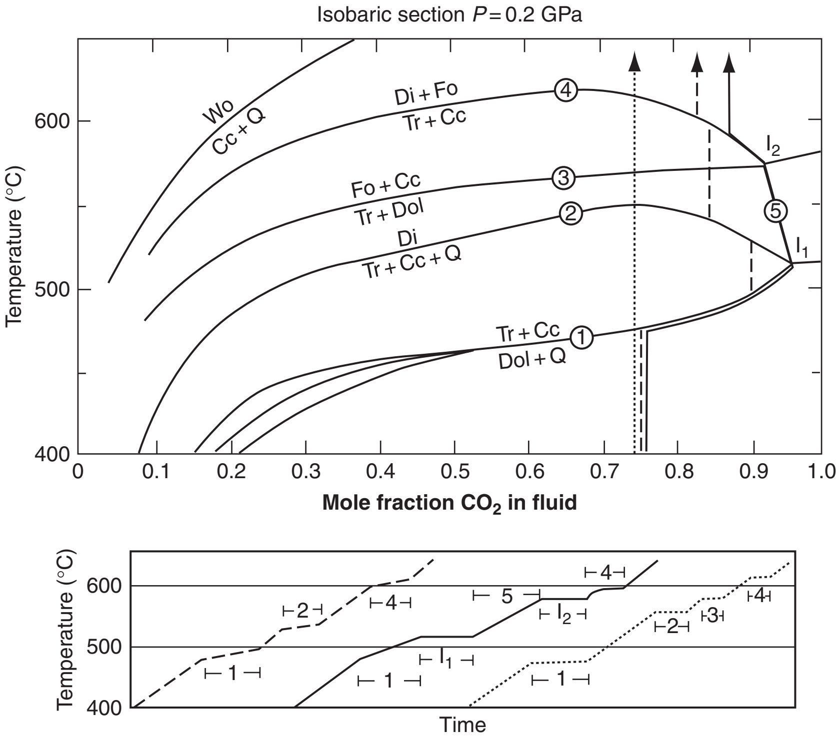
These reactions, with the exception of (20.27), are shown in Figure 20.8. For simplicity, reaction (20.27) is omitted because it involves only compositions outside the range of siliceous carbonate rocks. It can, however, occur in metamorphosed ultramafic igneous rocks (Problem 20.8).

This T–X_{CO_2} diagram explains why different sequences of minerals can form as a result of progressive metamorphism under different fluid compositions. The lowest-temperature field in Figure 20.8 contains the mineral assemblage typical of unmetamorphosed siliceous carbonate rocks; that is, calcite + dolomite + quartz. With rising temperature, the first metamorphic mineral to form depends on the fluid composition. At 0.2 GPa, if $X_{CO_2} > 0.53$, tremolite forms, whereas if $X_{CO_2} < 0.53$, talc forms, followed by tremolite. Bowen's decarbonation series, which starts with the appearance of tremolite, must therefore have been based on occurrences where the fluid was CO_2-rich, whereas in the area where Tilley found talc appearing before tremolite, the fluid must have been H_2O-rich. Note that tremolite again forms first if the fluid is extremely water-rich. In addition to accounting for differences in the first-appearing minerals, the diagram provides an explanation for the common occurrence of the sequence tremolite, diopside, forsterite, wollastonite with progressive metamorphism.

Twenty-five stable reactions involving a fluid relate the minerals quartz, calcite, dolomite, talc, tremolite, diopside, forsterite, and enstatite (Skippen, 1974). In addition, there are numerous reactions which do not involve a fluid and hence are unaffected by the composition of the fluid. Such reactions plot as horizontal lines in isobaric T–X_{CO_2} diagrams. If pelitic rocks are also present, a large number of dehydration reactions can occur. These have negative slopes in the T–X_{CO_2} plot. The resulting petrogenetic grid provides many mineral assemblages that are indicative of fluid compositions.

When the composition of the fluid is taken to be an environmentally controlled intensive variable, we are implicitly invoking communication of the rock with a large reservoir of fluid with fixed or buffered composition. Large quantities of this external fluid must flow or diffuse into the metacarbonate rock in order to keep the fluid composition constant during heating. Under such conditions we read the effects of increasing metamorphic temperature in the T–X_{CO_2} diagram by following a vertical line at the fixed composition of the fluid. For example, in Figure 20.9, which is for a total pressure of 0.2 GPa, we can follow the path taken by a rock

Fig. 20.9 Isobaric T–X_{CO_2} diagram (same as Fig. 20.8) showing three possible paths followed during progressive metamorphism of a rock initially in equilibrium with a fluid having a mole fraction of CO_2 of 0.75. If the rock is externally buffered with a reservoir having $X_{CO_2} = 0.75$, the path followed is indicated by the dotted line. If the rock is internally buffered, the composition of the fluid is modified by the reactions taking place; the solid line shows such a path for a fluid that is driven all the way to the invariant point, whereas the dashed line is for a rock that exhausts one of the reactants before reaching the invariant point (in this case dolomite). A schematic representation of the rise of temperature along these three paths during metamorphism is given below. See text for discussion.



that is heated in equilibrium with an externally buffered fluid phase having a composition of $X_{CO_2} = 0.75$ (dotted line in Fig. 20.9). Tremolite first appears at 475 °C, and because the fluid composition must remain constant, the temperature also remains fixed, while the tremolite-producing reaction goes to completion. Once one of the reactants is exhausted, the temperature of the rock is free to rise until it reaches the diopside-producing reaction at 550 °C. Again, the temperature remains constant while this reaction goes to completion. This step-like increase in temperature continues as the rock reaches successive reactions involving the phases remaining from previous reactions. Note that if the fluid has a different composition, the reactions take place at different temperatures, and therefore the steps on the heating curve are also at different temperatures. Mapping the first appearance of a mineral formed by such a reaction would therefore not provide a reliable indication of metamorphic temperatures unless independent evidence indicated the composition of the fluid was everywhere the same.

This type of metamorphism requires that large volumes of fluid be able to migrate freely and infiltrate into metacarbonate rocks. If the fluid flux is restricted in any way, reactions evolving H_2O and CO_2 in proportions that differ from that in the infiltrating fluid can develop local compositional inhomogeneities. In the extreme case of little or no infiltration, the fluid composition can be controlled entirely by local reactions in the rock – a scenario we will now turn our attention to.

The temperatures at which metamorphic reactions occur, and even the types of reactions that take place, can be quite different if the composition of the fluid is controlled by the rock rather than by an external buffer. Consider again a rock under a total pressure of 0.2 GPa being heated in equilibrium with a fluid that initially has $X_{CO_2} = 0.75$, but this time the fluid is internally buffered. Tremolite again first forms at 475 °C, but now, with the reaction consuming H_2O and liberating CO_2 (Eq. (20.22)), the composition of the fluid becomes richer in CO_2 (dashed line in Fig. 20.9). As it does so, the temperature rises in order to keep the rock on the isobaric univariant reaction line. Reactants are gradually consumed as the rock follows the reaction curve, and once one of them is exhausted, the rock is free to leave the reaction curve, with the temperature rising along a vertical line marking the new fluid composition. This path is followed until the next reaction involving the remaining phases is reached, whereupon the rock follows that univariant line with the composition of the fluid changing as it does so. Depending on the slope of the next reaction curve, the fluid can become enriched in either CO_2 or H_2O as the reaction proceeds.

When reactions taking place in the rock internally buffer the composition of the fluid, the temperature of the rock does not remain constant but rises gradually as the composition of the fluid changes until an invariant point is reached. In Figure 20.9 two internally buffered paths are shown. The one indicated with the dashed line runs out of reactants before reaching the invariant point I_1. Its time–temperature path consequently has a gradual slope with no plateaus; even inflections caused by reactions occur at different temperatures, depending on local abundances of minerals.

The other path, shown by a solid line, reaches the invariant point (I_1), which then produces a distinct plateau on the time–temperature plot that is independent of the bulk composition of the rock. Temperature and fluid composition remain fixed at the invariant point until a mineral reactant is exhausted. The fluid composition then rises along a univariant line to the

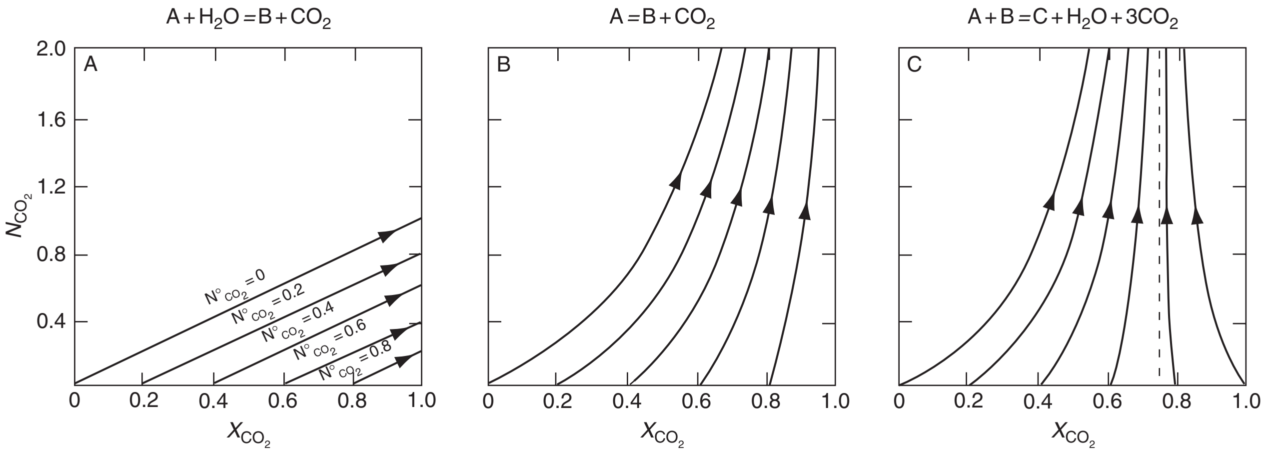


Fig. 20.10 Change in the mole fraction of CO_2 in a fluid caused by metamorphic reactions that involve H_2O and CO_2, where N_{CO_2} is the number of moles of CO_2 entering the fluid from the reaction. In (**A**), the reaction consumes $1H_2O$ and evolves $1CO_2$. A fluid initially having $X_{CO_2} = 0.2$ follows the line marked $N^0_{CO_2} = 0.2$ as the reaction proceeds. For the reaction in (**B**), only CO_2 is evolved, and the reaction in (**C**) evolves $1H_2O$ and $3CO_2$ (after Greenwood, 1975).

next invariant point (I_2) where again the fluid composition and temperature remain constant until one phase is exhausted. Greenwood (1975) has shown that for internal buffering, only small amounts of reactants need be consumed to cause major changes in the composition of the fluid. Indeed, he concludes that in most internally buffered rocks the changes are great enough to move fluid compositions along univariant lines all the way to isobaric invariant points. Therefore, internally buffered rocks spend a considerable amount of time at invariant points during metamorphism, and it is here that the most conspicuous changes in mineral assemblages occur.

The degree to which a reaction can buffer and change the composition of the fluid depends on the proportions in which H_2O and CO_2 are generated by the reaction relative to their mole fractions in the fluid. If we let the number of moles of H_2O and CO_2 in the fluid be N_{H_2O} and N_{CO_2}, respectively, then, as shown by Greenwood (1975), the mole fraction of CO_2 in the fluid is

$$X_{CO_2} = \frac{N_{CO_2}}{\left(N^0_{H_2O} + N^0_{CO_2}\right) + A\left(N_{CO_2} - N^0_{CO_2}\right)} \qquad (n \neq 0) \tag{20.28}$$

where $A = (m + n)/n$, m and n are the numbers of moles of H_2O and CO_2 produced by the reaction, respectively, and $N^0_{H_2O}$ and $N^0_{CO_2}$ are the initial numbers of moles of H_2O and CO_2, respectively, in the fluid. This equation relates the mole fraction of CO_2 in the fluid to the proportion of H_2O and CO_2 produced by the reaction (the A term), the initial composition of the fluid ($N^0_{H_2O} + N^0_{CO_2}$), and the extent to which the reaction has progressed ($N^0_{CO_2} - N^0_{CO_2}$).

For a reaction such as $MgCO_3 + H_2O = Mg(OH)_2 + CO_2$, the value of A is 0, so Eq. (20.28) becomes $X_{CO_2} = N_{CO_2}$; that is, there is a one-to-one relation between the number of moles put into the fluid by the reaction and the mole fraction of CO in the fluid. Figure 20.10(A) shows how the X_{CO_2} in the fluid changes during reaction. For example, if the fluid was initially pure H_2O, the fluid would follow the path labeled ($N^0_{CO_2} = 0$); if the fluid initially had $X_{CO_2} = 0.2$, it would follow the line labeled $N^0_{CO_2} = 0.2$. The buffering capacity of a reaction can be given quantitatively by differentiating Eq. (20.28) with respect to N_{CO_2}. For the magnesite-to-brucite reaction, dX_{CO_2}/dN_{CO_2} is a constant and equal to 1; that is, the buffering capacity of the reaction remains constant regardless of how long it takes place.

For a reaction such as calcite + quartz = wollastonite + CO_2, the value of A is 1, so Eq. (20.28) becomes $X_{CO_2} = N_{CO_2}/(1 + \Delta N_{CO_2})$ where $\Delta N_{CO_2} = N_{CO_2} - N^0_{CO_2}$, which is a measure of the progress of the reaction. In this case, dX_{CO_2}/dN_{CO_2} is a function of the progress of the reaction, with the value decreasing with time. Thus, in Figure 20.10(B) we see that for an initially pure H_2O fluid, $dX_{CO_2}/dN_{CO_2} = 1$, but as the reaction progresses, the derivative decreases (slope steepens in plot of N_{CO_2} versus X_{CO_2}) and the reaction causes progressively less change in the composition of the fluid. Notice also that the rate of change of composition of the fluid is less for those fluids that initially have a higher mole fraction of CO_2.

For a reaction such as Eq. (20.21), which produces $1H_2O$ and $3CO_2$, the change in composition of the fluid is similar to that for the wollastonite reaction, except that the fluid composition to which the reaction converges is $X_{CO_2} = 0.75$ rather than pure CO_2 (Fig. 20.10(C)). Here again, the change in composition of the fluid is greatest for early stages of reaction and for fluids with compositions that deviate most from the composition being evolved by the reaction.

Let us determine the amount of reaction that would have to take place along the reaction curve for 5Dol + 8Q + H_2O = Tr + 3Cc +$7CO_2$ (Eq. 20.22) in Figure 20.9 to change the mole fraction of CO_2 in the fluid from its initial value of 0.75 to a value of 0.9, where one of the reactants (we will assume dolomite) is exhausted and the fluid leaves the univariant line. Because the reaction consumes 1 mole of H O (1) and evolves 7 moles of CO_2 ($n = 7$), $A = 6/7$. If we start with 1 mole of fluid that has $X_{CO_2} = 0.75$, it initially contains 0.75 mole of CO_2 ($N^0_{CO_2}$). From Eq. (20.28) we find that the number of moles in the fluid in its final state is 1.406, which is an increase of 0.66 mole of CO_2; at the same time 0.094 mole of H_2O and 0.47 mole of dolomite are consumed. Because dolomite is exhausted by the reaction, the volume of rock affected must have initially contained this amount of dolomite, which is equivalent to 0.087 kg. This means that for 1 mole of initial fluid, 0.087 kg of dolomite is consumed in changing the composition of the fluid from a mole fraction of 0.75 to 0.90. From this we see that if rocks buffer the fluid, variations in modal abundances brought about by progressive metamorphism could be used to estimate the amount of fluid that passed through the rock during metamorphism (Section 20.6 and Chapter 21).

Isograds based on the first appearance of minerals in rocks that were externally buffered have little significance if the composition of the fluid is unknown. Similarly, the first appearance of minerals formed by isobaric univariant reactions in rocks that were internally buffered cannot provide unique indications of temperature. But reactions taking place at isobaric invariant points, to which the composition of the fluid is driven by the univariant reactions in the internally buffered rock, do record unique temperatures and fluid compositions for a given pressure. Isograds based on the appearance of minerals formed by these reactions are therefore the most reliable. In nature, conditions probably fall between the internal and external buffering extremes, a point we will explore further in the following section and in Chapter 21.

Regardless of whether the system was externally or internally buffered, reactions involving CO_2 and/or water are extremely useful for estimating the composition of the fluid phase if temperature and pressure are known independently from thermobarometry or geologic relations (Problem 20.12). Consider a metacarbonate rock that contains calcite, quartz, and wollastonite that equilibrated with a H_2O–CO_2 fluid (reaction 20.10). Assuming, for simplicity, ideal H_2O–CO_2 mixing and pure solids, the equilibrium constant expression for the reaction is

$$K = \exp\left(-\frac{\Delta G_r^*}{RT}\right) = \frac{f_{CO_2}}{10^5} = \frac{X_{CO_2}\gamma_{CO_2}P}{10^5}$$

Here ΔG_r^* is the standard state Gibbs free energy change of the reaction; it is evaluated using the pressure and temperature of interest for solids, and the temperature of interest and reference pressure (10^5 Pa) for CO_2. One can then solve for the mole fractions of H_2O and CO_2 in the fluid by noting that $X_{CO_2} = 1 - X_{H_2O}$. If the calculations are being done in bars, the reference P is one bar and K is simply equal to f_{CO_2} (10^5 Pa = 1 bar). For nonideal mixing, activity coefficients would also need to be taken into account (Problem 20.12).

20.6 REACTION PROGRESS AND FLUID INFILTRATION

Natural reactions in metacarbonate rocks can fall anywhere between the two extremes of internal and external buffering. For internal buffering, the fluid is generated by reactions taking place in the rock in question; no fluid is introduced, but some may be lost. For external buffering, fluid is introduced from a source outside the rock in question; such sources include dehydrating metapelitic rocks and degassing magmas. The important question is whether we can determine the extent of buffering due to infiltration. If this can be done, the amount of reactive infiltration can be estimated (Rice and Ferry, 1982; Ferry, 1983).

In metamorphic terranes where rocks have been internally buffered, mineral assemblages form divariant zones (univariant isobaric zones) where modal abundances vary continuously as a result of continuous reactions (dashed and solid paths in Fig. 20.9). These zones are separated by abrupt changes in modal abundance resulting from discontinuous reactions at isobaric invariant points. Also, because the fluid composition is buffered by local mineral assemblages, initial differences in rock composition may result in different fluid compositions. These differences produce chemical potential gradients in the fluid components over relatively short distances, such as those between layers of different composition in an outcrop. On the other hand, the chances of finding isobaric univariant reaction assemblages in rocks that are externally buffered are very small. The vertical paths followed by such buffered assemblages through T–X_{CO_2} diagrams (dotted line in Fig. 20.9) indicate that isobaric divariant assemblages are far more likely to be found.

It will be recalled from Section 20.5 that the degree to which a reaction can buffer and change the composition of a fluid is given by Eq. (20.28), which for a binary H_2O–CO_2 fluid can be written as

$$X_{CO_2} = \frac{N_{CO_2}}{N^0_{H_2O} + N^0_{CO_2} + A\left(N_{CO_2} - N^0_{CO_2}\right)}$$

where X_{CO_2} is the mole fraction of CO_2 in the fluid at a given time, N_{CO_2} and N_{H_2O} are the number of moles of CO_2 and H_2O in the fluid at that time, $N^0_{CO_2}$ and $N^0_{H_2O}$ are the initial numbers of moles of CO_2 and H_2O in the fluid, and $A = (m + n)/n$ where m and n are, respectively, the numbers of moles of H_2O and CO_2 produced by the reaction. A measure of the progress of the reaction is given by the last term in the equation, $N_{CO_2} - N^0_{CO_2}$, which is simply the number of moles of CO_2 produced or consumed by the reaction. To account for differences in the amount of CO_2 resulting from the stoichiometry of the reaction, the $N_{CO_2} - N^0_{CO_2}$ term is divided by the stoichiometric coefficient for CO_2 in the reaction (n) to give what is referred to as a *reaction progress variable*, ξ (Brimhall, 1979; Rice and Ferry, 1982), which we can define for this case as

$$\xi \equiv \frac{N_{CO_2} - N^0_{CO_2}}{n} \qquad (20.29)$$

We can then substitute ξn for $N_{CO_2} - N^0_{CO_2}$ and $\xi n + N^0_{CO_2}$ for N_{CO_2} in Eq. (20.28), and recalling that $A = (m + n)/n$, we obtain

$$X_{CO_2} = \frac{X^0_{CO_2}\left(N^0_{H_2O} + N^0_{CO_2}\right) + \xi n}{\left(N^0_{H_2O} + N^0_{CO_2}\right) + \xi(m+n)} \tag{20.30}$$

which can be rearranged to give ξ

$$\xi = \frac{\left(N^0_{CO_2} + N^0_{H_2O}\right)\left(X^0_{CO_2} - X_{CO_2}\right)}{X_{CO_2}(m+n) - n} \tag{20.31}$$

This equation then relates the change in the composition of the fluid to the reaction progress variable, from which we can determine the change in the modal abundances of minerals.

Reactions in many metamorphic rocks are likely to occur under conditions intermediate to those of complete internal or external buffering. Mineral assemblages in these reactions still buffer the fluid composition, but some fraction of the fluid is derived from an external source. In such cases the extent of infiltration may be estimated from changes in the modal abundances of minerals. If no fluid were introduced, the amount of reaction necessary to establish an equilibrium fluid composition for any given change in temperature along an isobaric univariant reaction can be calculated from the stoichiometry of the reaction (Problem 20.14). But this amount of reaction produces a predictable change in the modal abundances of the reactant and product minerals if we have an estimate of the porosity of the rock. Differences between the predicted and actual change in the modal abundances indicate the degree of infiltration. For example, if the externally derived fluid had, coincidently, the equilibrium composition, no reaction would be required whatsoever, and the modal abundances of the minerals would not change. On the other hand, if the infiltrating fluid was poorer in the component to which the buffered fluid must progress, an additional amount of reaction would be required, and the change in the modal abundances of the minerals would be larger than expected. To estimate the quantity of infiltrating fluid, the reaction progress must be related to the composition of the fluid and the abundances of minerals involved in the reaction.

Consider, for example, a rock containing quartz and calcite that has a porosity of 1% and is in equilibrium with a CO_2–H_2O fluid in which $X_{CO_2} = 0.01$ (Rice and Ferry, 1982). On being heated to 470 °C at 0.35 GPa ($P_{total} = P_{CO_2} + P_{H_2O}$), the following reaction starts:

$$\text{quartz} + \text{calcite} = \text{wollastonite} + CO_2 \tag{20.32}$$

As reaction progresses with rising temperature, the evolved gas enriches the fluid in CO_2, causing the assemblage to follow the isobaric univariant reaction curve in the T–X_{CO_2} diagram (Fig. 20.11). Regardless of whether or not externally derived fluid is added, the fluid must have the composition given by the isobaric univariant curve if quartz + calcite + wollastonite are to be in equilibrium at any given temperature. Thus, at 600 °C, the fluid must have a composition of $X_{CO_2} = 0.086$.

Using Eq. (20.31), we now calculate how much quartz and calcite would have to react to change the fluid from $X^0_{CO_2} = 0.01$ to $X_{CO_2} = 0.086$ in rising from 470 °C to 600 °C. If we treat a rock volume of 1000 cm^3 (a large hand specimen), the pore fluid occupies 10 cm^3 and with a composition of $X_{CO_2} = 0.01$, would contain 0.447 mol of H_2O ($N^0_{H_2O}$) and 0.005 mol of CO_2 ($N^0_{CO_2}$). Substituting these values into Eq. (20.31) gives

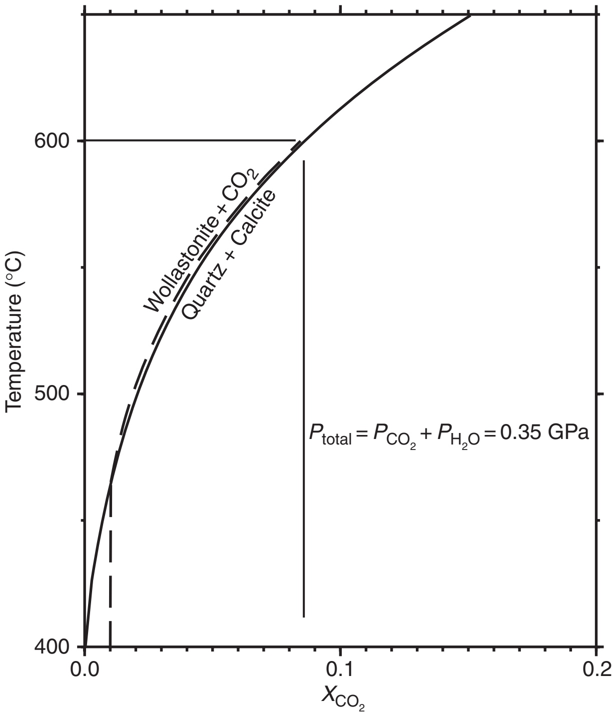


Fig. 20.11 Part of the isobaric T–X_{CO_2} diagram for the reaction quartz + calcite = wollastonite + CO_2. The dashed line indicates the path followed by a fluid that initially has $X_{CO_2} = 0.01$ and is then buffered by the reaction as the temperature rises to 600 °C. See text for discussion.

$$\xi = \frac{(0.005 + 0.447)(0.01 - 0.086)}{0.086 \times (1) - 1} = 0.0376$$

From Eq. (20.29) we therefore obtain

$$N_{CO_2} - N^0_{CO_2} = \xi n = 0.0376$$

Thus, only 0.0376 mol of CO_2 need be produced to effect the change in fluid composition, and because of the stoichiometry of the reaction, 0.0376 mol each of quartz and calcite would be consumed and 0.0376 mol of wollastonite would be formed. The molar volume of wollastonite is 39.93 cm^3 (Table 7.1). Thus the volume of wollastonite produced in the reaction is 1.5 cm^3, which is a 0.15% change in modal abundance. Clearly, very little wollastonite need be formed if the rock is internally buffered.

Rumble *et al.* (1982) have described the contact metamorphism of a siliceous limestone in west-central New Hampshire in which wollastonite formed under a pressure of 0.35 GPa and a maximum temperature of 600 °C. Near the igneous contact, the metamorphosed limestone contains 70% wollastonite, an amount far greater than would be expected if the rock were internally buffered, even if the porosity had been considerably higher than 1%. This indicates infiltration by an H O-rich fluid, the excess wollastonite being formed when the reaction was establishing the equilibrium fluid composition ($X_{CO_2} = 0.086$).

If again we deal with 1000 cm^3 of this New Hampshire rock, 70 modal % of wollastonite corresponds to 700 cm^3, or 17.53 mol of wollastonite. From the stoichiometry of the reaction (Eq. (20.32)), 17.53 mol of CO_2 must also have been produced. Despite this large production of CO_2, the composition of the fluid at 600 °C was only $X_{CO_2} = 0.086$. The CO_2 must therefore have been considerably diluted with a CO_2-poor fluid. If this fluid is assumed to have been pure H_2O, we can write

$$X_{CO_2} = 0.086 = \frac{N_{CO_2}}{N_{CO_2} + N_{H_2O}^{inf}} = \frac{17.53}{17.53 + N_{H_2O}^{inf}}$$

where $N_{H_2O}^{inf}$ is the number of moles of H_2O that infiltrated the rock during metamorphism. Thus 186.3 mol of H_2O must have passed through the rock during metamorphism. The molar volume of H_2O at 600 °C and 0.35 GPa is 19.6 cm^3 (Eq. (20.6)). Consequently, 3652 cm^3 of H_2O must have infiltrated the rock during the reaction. If we add to this the 871 cm^3 of CO_2 released by the reaction (17.53 × 49.7 cm^3), the volume of the fluid phase was 4523 cm^3. Thus the ratio of the fluid to rock volumes (known as the *fluid:rock ratio*), was 4.5:1. Importantly, this is a *minimum* estimate; if the fluid had not been pure H_2O and had contained some CO_2, still larger volumes of fluid would have been required to infiltrate the rock.

Reactions that take place at isobaric invariant points provide another measure of the extent of infiltration. Again we consider reactions treated by Rice and Ferry (1982). A rock containing anorthite, calcite, quartz, zoisite, tremolite, and diopside in equilibrium with a CO_2–H_2O fluid is isobarically invariant (Fig. 20.12(A)). At 0.35 GPa the fluid has a composition of $X_{CO_2} = 0.094$ and the temperature is 481 °C. The invariant point is produced by the intersection of four isobaric univariant reactions. The reaction at the invariant point can be described in terms of any two of these reactions, as long as they are linearly independent. For example,

$$\begin{aligned}&\text{tremolite} + 3\ \text{calcite} + 2\ \text{quartz}\\&\quad = 5\ \text{diopside} + 3CO_2 + H_2O\end{aligned} \qquad (20.33)$$

and

$$2\ \text{zoisite} + CO_2 = 3\ \text{anorthite} + \text{calcite} + H_2O \qquad (20.34)$$

Reaction progress variables for these two reactions (ξ_{33} and ξ_{34}) are defined as (Eq. (20.29))

$$\xi_{33} \equiv \frac{N_{CO_2} - N_{CO_2}^0}{3}$$

and

$$\xi_{34} \equiv \frac{N_{CO_2} - N_{CO_2}^0}{1}$$

If the rock is internally buffered, the volatiles produced by the reactions must be in the proportions defined by the isobaric invariant point; that is, $X_{CO_2} = 0.094$. The number of moles of CO and H O produced by each reaction can be expressed in terms of the reaction progress variables. Thus reaction (20.33) produces $3\xi_{33}$ moles of CO_2 and ξ_{33} moles of H_2O, and reaction (20.34) produces $-\xi_{34}$ moles of CO_2 (negative sign indicates CO_2 is consumed) and ξ_{34} moles of H_2O. The mole fraction of CO_2 in the fluid at the invariant point can therefore be written as

$$\begin{aligned}X_{CO_2} = 0.094 &= \frac{3\xi_{33} - \xi_{34}}{(3\xi_{33} - \xi_{34}) + (\xi_{33} + \xi_{34})}\\ &= \frac{3\xi_{33} - \xi_{34}}{4\xi_{33}}\end{aligned} \qquad (20.35)$$

from which it follows that $\xi_{33}/\xi_{34} = 0.381$. Thus for the mole fraction of CO_2 to remain fixed at 0.094, reactions (20.33) and (20.34) must progress in the proportions 0.381:1. Graphically this can be represented by a straight line in a plot of ξ_{33} versus ξ_{34} (Fig. 20.12(B)).

Because of the stoichiometry of reactions (20.33) and (20.34), the number of moles of diopside (N_{Di}) produced is $5\xi_{33}$, and the number of moles of anorthite (N_{An}) is $3\xi_{34}$. Thus $\xi_{33} = N_{Di}/5$, and $\xi_{34} = N_{An}/3$, which can be substituted into Eq. (20.35) to give the ratio of diopside to anorthite that must be produced for the fluid to retain its composition of $X_{CO_2} = 0.094$; that is,

$$\frac{N_{Di}}{N_{An}} = 0.635 \qquad (20.36)$$

If the rock is infiltrated by fluid with a composition different from that of the isobaric invariant point, the production of diopside and anorthite will differ from the ratio given by Eq. (20.36).

If a rock in which tremolite + calcite + quartz + zoisite is reacting to form diopside and anorthite is infiltrated by a fluid composed of pure H_2O, the mole fraction of CO_2 in the fluid at the isobaric invariant point is

$$X_{CO_2} = 0.094 = \frac{3\xi_{33} - \xi_{34}}{4\xi_{33} + N_{H_2O}^{inf}} \qquad (20.37)$$

where $N_{H_2O}^{inf}$ is the number of moles of H_2O infiltrating during the reaction. On rearranging, Eq. (20.37) gives

$$N_{H_2O}^{inf} = 27.915\xi_{33} - 10.638\xi_{34} \qquad (20.38)$$

Equation (20.38) can be expressed in terms of volume of H_2O by multiplying both sides of the equation by the molar volume of H_2O, which is 22.11 cm^3 at 481 °C and 0.35 GPa. This gives

$$V_{H_2O}^{inf} = 617.20\xi_{33} - 235.21\xi_{34}\ \text{cm}^3 \qquad (20.39)$$

Equation (20.39) defines straight lines in a plot of ξ_{33} versus ξ_{34} (Fig. 20.12(C)). In the limiting case where $V_{H_2O}^{inf} = 0$, Eq. (20.39) becomes Eq. 20.35. The values of $V_{H_2O}^{inf}$ plotted in Figure 20.12(C) have been multiplied by 10^{-3}. Because we are dealing with a rock volume of 1000 cm^3, the values on the lines give the ratio of the volume of infiltrating fluid to volume of rock. Thus a point on the line labeled 1.0 indicates that 1000 cm^3 of H_2O would have infiltrated the rock during metamorphism. Conversely, if the modal abundance of diopside and anorthite indicate that ξ ξ = 1, then 400 cm^3 of

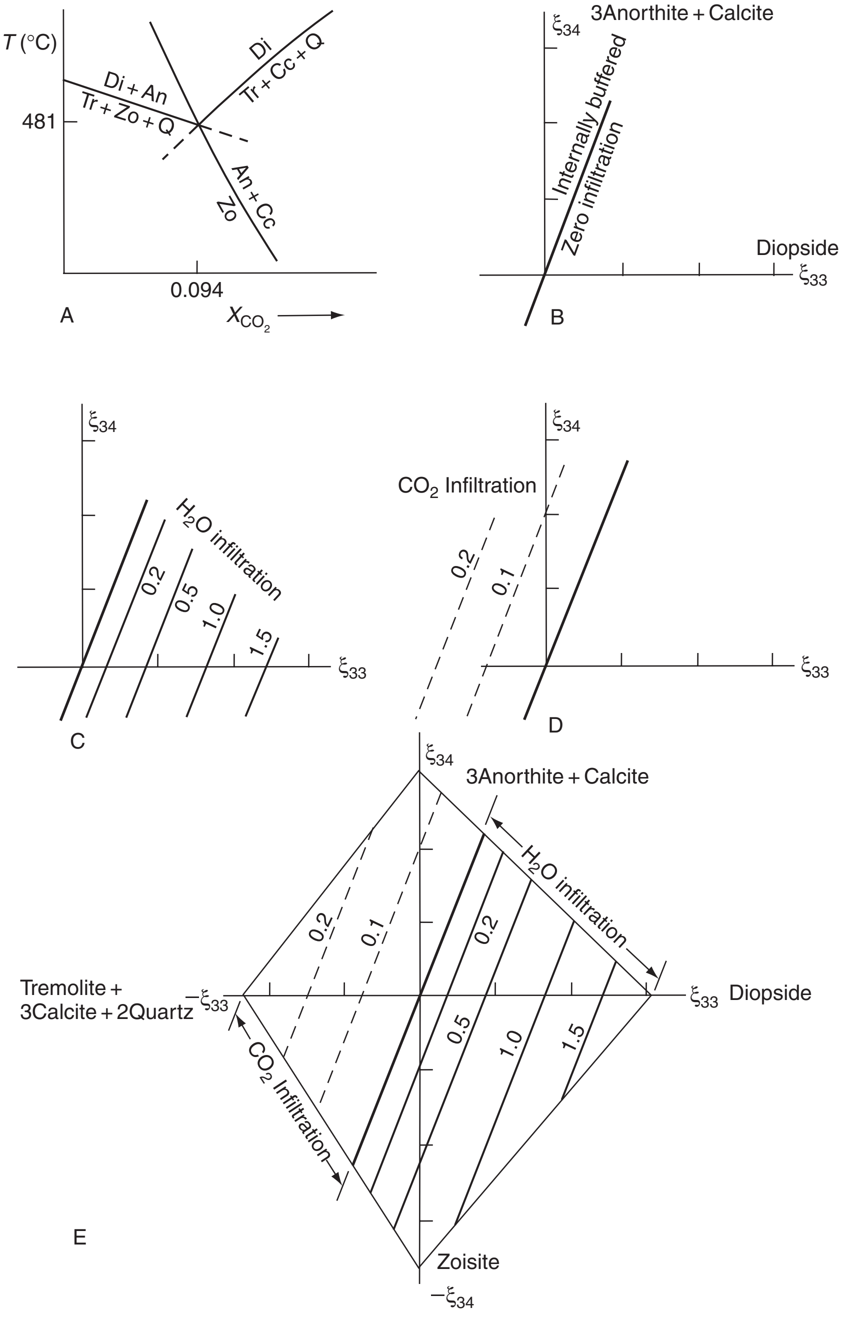


Fig. 20.12 Reaction taking place at the isobaric invariant point in the T–X_{CO_2} diagram (**A**) can be expressed as the sum of two reactions, tremolite + calcite + quartz = diopside + 3CO$_2$ + H$_2$O and zoisite + CO$_2$ = anorthite + calcite + H$_2$O (Eqs. (20.33) and (20.34)). Because fluid composition is fixed at the invariant point, reactions (20.33) and (20.34) must progress at rates that produce CO$_2$ and H$_2$O in the proportions required by the invariant point. Relative rates of progress of these two reactions are given by the solid line in (**B**), which plots the reaction progress variable (see text for definition) of the one reaction against that of the other. If pure H$_2$O infiltrates the rock, the reaction producing the larger amount of CO$_2$ (20.33) must progress more rapidly if the fluid composition is to remain at the invariant composition. Volumes of H$_2$O infiltrating 1000 cm^3 of rock are shown multiplied by 10^{-3} in (**C**). If pure CO$_2$ fluid infiltrates the rock, the reaction producing the most water must progress more rapidly to maintain the invariant fluid composition. Volumes of infiltrating CO$_2$ (× 10^{-3}) are shown by dashed lines in (**D**). These figures are combined into a single diagram in (**E**). Negative values of reaction progress variables indicate the reactions above occurring in reverse.

pure H$_2$O could have infiltrated the rock during metamorphism. As we saw above, had the infiltrating fluid not been pure H$_2$O, a larger volume of fluid would have been necessary to effect the same change.

Identical relations can be developed for infiltration by a fluid composed of pure CO$_2$. The mole fraction of CO$_2$ in the fluid at the isobaric invariant point is

$$X_{CO_2} = 0.094 = \frac{3\xi_{33} - \xi_{34} + N^{\text{inf}}_{CO_2}}{4\xi_{33} + N^{\text{inf}}_{CO_2}} \qquad (20.40)$$

On rearranging the equation and converting moles of CO$_2$ ($N^{\text{inf}}_{CO_2}$) to volume by multiplying by the molar volume of CO , which is 48.98 cm^3 at 481 °C and 0.35 GPa, we obtain

$$V^{\text{inf}}_{CO_2} = 51.85\xi_{34} - 136.07\xi_{33} \text{ cm}^3 \qquad (20.41)$$

Equation (20.41) defines straight lines in a plot of ξ_{33} versus ξ_{34} (Fig. 20.12(D)). These are parallel to the lines for infiltration by H$_2$O fluids but plot on the opposite side of the line marking the internally buffered reaction. In the limiting case where $V^{\text{inf}}_{CO_2} = 0$, Eq. (20.41) becomes Eq. 20.35. The values of $V^{\text{inf}}_{CO_2}$ plotted on the lines in Figure 20.12(D) have also been multiplied by 10^{-3}; these values express the ratio of the volume of the infiltrating fluid to rock volume (1000 cm^3). Again, the volumes of infiltrating fluid are only minimum values. If the fluid were not pure CO$_2$, larger volumes would be required to effect the same change.

In describing the reaction progress variables for reactions (20.33) and (20.34), the reactions were assumed to progress to the right; that is, diopside and anorthite were reaction products. But the reactions could progress to the left, and then diopside and anorthite would be reactants. Rather than define new reaction progress variables for the reverse reactions, we can simply use negative values of the progress variables for the forward reactions. Thus, in Figure 20.12, each ξ_{33} versus ξ_{34} plot extends to negative values of the two variables, dividing each plot into four quadrants. In the first quadrant, diopside and anorthite are both products of the isobaric invariant reaction. In the second quadrant, anorthite becomes a reactant (negative), and diopside is joined by zoisite as products. In the third quadrant, diopside also becomes a reactant along with anorthite to produce tremolite, calcite, quartz, and zoisite. Finally in the fourth quadrant, anorthite again becomes a product, along with tremolite, calcite, and quartz, while diopside and zoisite are reactants. Each quadrant, therefore, expresses a different way in which the mode of the rock could change at the isobaric invariant reaction point, depending on the nature of the infiltrating fluid. Figure 20.12(B) to (D) have been combined into a single diagram in Figure 20.12(E). The lines representing the minimum volumes of infiltrating H_2O or CO_2 fluids are applicable in each of the quadrants and indicate the composition of fluids that would give rise to the particular modal changes.

Because Figure 20.12 is constructed for 1000 cm^3 of rock, there is a limit to how far the reactions can progress before going to completion. For example, reaction (20.33) is complete when the entire 1000 cm^3 of rock is converted to diopside. Because the molar volume of diopside is 66.09 cm^3 (Table 7.1), 1000 cm^3 of diopside would correspond to 15.13 mol. According to the stoichiometry of reaction (20.33), this amount of diopside is produced once the reaction progress variable, ξ_{33}, reaches a value of 3.026. Conversion of the entire volume of rock into a mixture of anorthite and calcite in a 3:1 molar ratio limits ξ_{34} to a value <2.947. Conversion of the volume entirely into zoisite limits $\xi_{34} > -3.662$. Finally, conversion of the volume entirely into tremolite, calcite, and quartz in the molar ratio of 1:3:2 limits $\xi_{33} > -2.330$.

In Chapter 21 we will see that infiltrating fluids are rarely pure water or pure CO_2. As a consequence, the fluid:rock ratios calculated here can underestimate significantly the actual amount of infiltration that occurs during metamorphism. In order to obtain better estimates of fluid fluxes, it will be necessary to consider processes of fluid infiltration and their relationships to reaction. Nonetheless, a fluid:rock ratio is a valuable general indicator of whether or not infiltration has occurred.

20.7 PROBLEMS

20.1 Using the data in Table 20.1, calculate pressure versus molar volume graphs for water using the ideal gas law, the van der Waals equation, and the modified Redlich Kwong equation for a temperature of 750 °C and for molar volumes between 6×10^{-5} and 11×10^{-5} m^3.

20.2 Starting with the spreadsheet used for solving Problem 20.1, enter Eq. (20.8) for the fugacity coefficient of water, using separate columns for each group of terms in the equation. Then calculate the pressure, fugacity, and density of H_2O for molar volumes between 1.89×10^{-5} and 63×10^{-5} m^3 at 600 °C (873 K). Plot graphs of fugacity versus pressure and density versus pressure.

20.3 In Problem 8.8 the equilibrium pressures of water for the reaction of quartz + muscovite = sillimanite + K-feldspar + H_2O were calculated, assuming ideal behavior, for 600, 650, 700, 750, and 800 K to be 0.3, 1.3, 4.1, 12.2, and 28.7 MPa, respectively. However, H_2O is not an ideal gas, so the calculated pressures are in fact fugacities. Using the spreadsheet from Problem 20.2, select molar volumes for H_2O that give the fugacities above for the appropriate temperatures, and by so doing determine the actual equilibrium pressures. Graph the equilibrium curve for the reaction both for real and ideal behavior, noting how the fugacity affects the results. From the change in the fugacity coefficient of H_2O with pressure, how might you expect the real curve to be positioned with respect to the ideal curve at high pressures?

20.4 The object of this problem is to calculate the equilibrium pressure for the reaction quartz + calcite = wollastonite + CO_2 at 750 °C. The procedure is outlined for 700 °C in Section 20.4. First, starting with the same spreadsheet used in Problem 20.2 for H_2O, change the appropriate coefficients to those for CO_2 given in Table 20.1. Then create a table of fugacity coefficients and pressures for CO_2 over a range of pressures at 750 °C (1023 K). Take the free-energy change of the reaction and the volume change of the condensed phases for 750 °C and atmospheric pressure to be $-65\,085$ J mol^{-1} and -1.9692 m^3 mol^{-1}, respectively. By an iterative process, select a value of γ and calculate P using Eq. (20.13) until the values of γ and P match values in the tables you created for γ and P.

- **(a)** What is the equilibrium P and the value of γ at 750 °C?
- **(b)** What would the equilibrium P be if CO_2 were assumed to behave ideally?

20.5 Using the results from Problem 20.4, calculate the equilibrium pressure for the reaction quartz + calcite = wollastonite + CO_2 at 750 °C for the condition where the activity of CO_2 is buffered by the environment at **(a)** 0.75 and **(b)** 0.5.

20.6 Starting with Eq. (20.18), show that in an isobaric T–X_{CO_2} diagram, a reaction that evolves both H_2O and CO_2 has a temperature maximum when the ratio of the mole fractions of H_2O and CO_2 in the fluid are in the same proportions as the numbers of moles of H_2O and CO_2 produced by the reaction.

20.7 Starting with Eq. (20.18), determine a general equation for a reaction involving H_2O and CO_2 in terms of T and X P. Then plot graphs for the following reactions, assuming that ΔS_r remains constant over the temperature interval considered. You can use any value for ΔS_r, but you may wish to experiment with how this value affects the shape of the reaction curves.

(a) $A = B + 5H_2O + 3CO_2$
(b) $A = B + H_2O + 7CO_2$
(c) $A + H_2O = B + 2CO_2$
(d) $A + H_2O = B + 5CO_2$

20.8 The invariant point at 463 °C and $X_{CO_2} = 0.53$ in Figure 20.8 is missing one univariant line which was omitted because it involves a reaction that could not take place in the compositional range of metamorphosed sedimentary carbonate rocks. It can, however, occur in metaperidotites. Determine the missing reaction, and using Schreinemakers rules, draw the array of univariant lines around this invariant point. Check your answer by constructing SiO_2–CaO–CO_2 diagrams (Fig. 20.4) for each divariant field. Make certain that the missing reaction has the correct slope for the gases evolved.

20.9 Repeat Problem 20.8 for the invariant point at 575 °C and $X_{CO_2} = 0.92$.

20.10 Using Eq. (20.28), determine how many moles of tremolite would have to be consumed by the Tr + 3Cc = 4Di + Dol + H_2O + CO_2 reaction (Eq. 20.23) in order to move the univariant assemblage from the invariant point at 512 °C and $X_{CO_2} = 0.96$ to the invariant point at 575 °C and $X_{CO_2} = 0.92$.

20.11 The minerals enstatite (En), talc (Tc), magnesite (Mag), and quartz (Q) are related by the following reactions:

[En]	$3Mag + 4Q + H_2O = Tc + 3CO_2$
[Tc]	$Mag + Q = En + CO_2$
[Q]	$Mag + 3Tc = 4En + H_2O + CO_2$
[Mag]	$Tc = 3En + Q + H_2O$

(a) Show each reaction on a separate T–X_{CO_2} diagram, labeling each curve with reactants and products on the correct side.

(b) These reactions meet at an isobaric invariant point at $X_{CO_2} = 0.9$ on the isobaric T–X_{CO_2} diagram. Plot all four reactions on a T–X_{CO_2} diagram, and use Schreinemakers rules to obtain the correct arrangement of stable and metastable curves around the invariant point. Check your answer by placing mineral facies diagrams in each divariant field.

20.12 A metacarbonate rock contains calcite, quartz, and wollastonite (reaction (20.10)). Independent thermobarometry indicates that the minerals equilibrated at 550 ºC and 0.2 GPa. The value of the equilibrium constant for (20.10) at these conditions is 495.4, and the fugacity coefficient for pure CO_2 is 1.861 according to the Kerrick and Jacobs (1981) equation of state.

(a) Compute the mole fractions of CO_2 and H_2O in the equilibrium fluid assuming ideal mixing.

(b) Repeat part (a), but use nonideal mixing for the fluid. For very water-rich fluids, we can employ a simple Henry's law expression (Fig. 9.3) for the activity of CO_2 based on the mixing model of Kerrick and Jacobs (1981). The approximate relationship for these P–T conditions is: $a_{CO_2} = 2.2X_{CO_2}$. How do the ideal and nonideal mixing results compare?

20.13 Simple calculation using winTWQ (Berman, 2007). Here we will examine the wollastonite-producing reaction from Problem 20.12. This program is only available for Windows-based PCs.

(1) Start the program, and just click on **continue** for the batch run window.
(2) Make sure the TWQ v. 2.32 bullet is selected, and click on **continue**.
(3) If the program was installed properly, the next list of Global Options should be alright as is. Click on **next** to proceed.
(4) Select the T–X_{CO_2} bullet, then click **next**.
(5) The next window has many important options. Click on each window to edit or modify. On the top line, enter the chemical elements of interest without commas, periods, or dashes. For our problem, they are: Ca Si O C. Set the pressure to 2000.0 bars (be sure to use decimal points for all numeric values). Min and Max temperatures of 400.0 and 800.0 work well. The mole fraction H_2O-CO_2 option sets the x-axis limits on the T–X_{CO_2} graph; these can be adjusted or left as is. Select the nonideal (KJ, 1981) option, and the KJ (81) options for the water and CO_2 routines (KJ (81) refers to Kerrick and Jacobs, 1981). Defaults for the other options should work satisfactorily. When done, click on **next**.
(6) Hold down the Ctrl key and multiselect the end members of interest. For this problem, they are A-QUARTZ, WOLLASTONITE, CALCITE, and CARBON DIOXIDE. The other options should be alright as they are.
(7) Click on **next** in the Reaction Selection window.
(8) Click on **exit**.
(9) When done, the T–X_{CO_2} relations will be tabulated in the output file SUMMARY.DAT. Also, you can use the program winDXF to make a graphics .dxf file from the PLOT.DAT output file. The .dxf file can then be loaded directly into many common graphics programs. The wplot program will make quick plots of the results. How do your results from Problem 20.12 at 550 ºC compare with the winTWQ results? You will need to inspect the SUMMARY.DAT file to see the T–X_{CO_2} values in detail.

20.14 Impure carbonate rocks below the isograd marked by the isobaric invariant assemblage anorthite + calcite + quartz + tremolite + diopside + zoisite contain no diopside or zoisite, but those above the isograd contain 42.96 vol % diopside and 35.33% zoisite. If the reaction occurred at a pressure of 0.35 GPa ($T = 481$ °C), the molar volumes of H_2O and CO_2 would be 22.11 and 48.98 cm^3, respectively. The molar volumes of diopside and zoisite are 66.09 and 135.9 cm^3, respectively.

(a) What values of the reaction progress variables for reactions (20.33) and (20.34) would account for the modal change?

(b) What volume and composition of fluid is produced by these reactions in 1000 cm^3 of rock? Does the fluid composition produced by the reactions match that of the isobaric invariant point?

(c) What minimum volume and composition of fluid must have infiltrated the rock during reaction?

(d) What was the total fluid to rock volume ratio for the reaction?

(e) Finally, plot the reaction progress variables in Figure 20.12(E), and check that your calculations in parts (c) to (e) are correct.

21 Material transport during metamorphism

21.1 INTRODUCTION

Approximately 5 wt% H_2O must be driven from a shale to convert it to a high-grade metapelitic rock. Conversely, several wt% H_2O must be added to an anhydrous basalt to convert it to a chlorite schist. Clearly, metamorphism in these cases must involve the transport of considerable quantities of H_2O. Similarly, CO_2 released by decarbonation reactions must also be transported away from the rock. At the temperatures and pressures normally extant during metamorphism, H_2O forms a supercritical fluid, which has a density similar to that of water at room temperature and pressure. The fluid has a low viscosity, and can therefore flow through fractures or an interconnected pore network. This hot fluid, in equilibrium with a rock, is certainly not pure H_2O but, instead, contains considerable amounts of dissolved ions. The loss or addition of such a fluid phase during metamorphism can therefore change the bulk composition of a rock in ways other than simply modifying its content of H_2O and other volatiles.

When a metamorphic rock undergoes significant changes in composition, the process of change is referred to as *metasomatism*. In extreme cases, metasomatism can completely change the composition of a rock, as, for example, in the conversion of limestone to magnetite ore bodies at contacts with igneous intrusions. Despite clear evidence for such major compositional changes, metasomatism was the subject of much controversy during the first half of the twentieth century. The analysis of metamorphic mineral assemblages is firmly based on thermodynamic principles, but metasomatism, which was invoked to explain certain field relations (see Fig. 4.50(C)), seemingly provided little that could be tested. Commonly, the rock that was supposed to have been replaced no longer existed, nor did the solutions that brought about the change. The petrologic community became polarized into those who believed that metasomatism was capable of effecting enormous compositional changes over large regions and those who thought its effects were minor. The debate culminated with the controversy over the origin of granite (Read, 1957), with metasomatists arguing in favor of a metasomatic origin (*granitization*), and magmatists arguing for an igneous origin. The experimental work of Tuttle and Bowen (1958) decided the question in favor of the magmatists, and metasomatism fell out of favor. Indeed, a number of metamorphic petrology books went so far as to ignore the term completely. More recently, however, metasomatism has experienced a comeback, but this time it is firmly grounded in physical–chemical principles (e.g. Hofmann *et al*., 1974). Isotopic studies have also provided important means of monitoring changes in rock composition (e.g. Rye *et al*., 1976).

Today, it has become clear that fluid flow and mass transfer play fundamental roles in metamorphism (Fig. 21.1). Deformation and fluids, for example, are closely related. The development of high pore fluid pressures during metamorphic devolatilization reactions can lead to rock weakening, hydrofacturing, vein formation and, in tectonically active regions, seismic failure. Large fluid fluxes though shear zones may cause chemical changes leading to the growth of micas; this weakens the rock and thus changes fault behavior. As we saw earlier in Chapter 17, fluids also facilitate the development of certain cleavage fabrics.

Fluid infiltration during prograde heating is critical for driving many devolatilization reactions, particularly those that release CO_2. This release contributes to Earth's long-term carbon cycle. The amounts and timescales of metamorphic carbon transfer through geologic time are major research questions in the geosciences. The fluids that flow through metamorphic rocks are not just water and CO_2; they contain a host of dissolved ions as well, including ore metals. These metals, along with volatiles like CO_2, can be transported from metamorphic depths to the shallow hydrosphere, where precipitation in ore deposits occurs. If the flow rates are large enough, heat can be transported by fluid flow as well (Chapter 22). Establishing the connections between shallow and deep flow systems is an ongoing challenge in petrology.

Metamorphic dehydration of subducted oceanic crust produces the fluids that ascend into the mantle wedge to cause the partial melting that gives rise to arc magmas. Furthermore, the fluids transfer chemical elements to the mantle, exerting fundamental controls on the chemical and isotopic composition of arc magmas. Moreover, the fluid content of the magmas, derived mostly from metamorphic devolatilization, is the primary reason for explosive volcanism and the associated volcanic hazards above subduction zones.

How does this mass transfer take place? Material has two basic means of transport. An atom can behave passively, being moved only when its surroundings move, or it can move independently of its surroundings in response to a force developed by a potential gradient that affects that particular element. The first type of movement is referred to as *advection* and the second as *diffusion*. Ions carried in solution by a fluid that flows through a rock would be an example of advection, whereas diffusion of an ion down a concentration

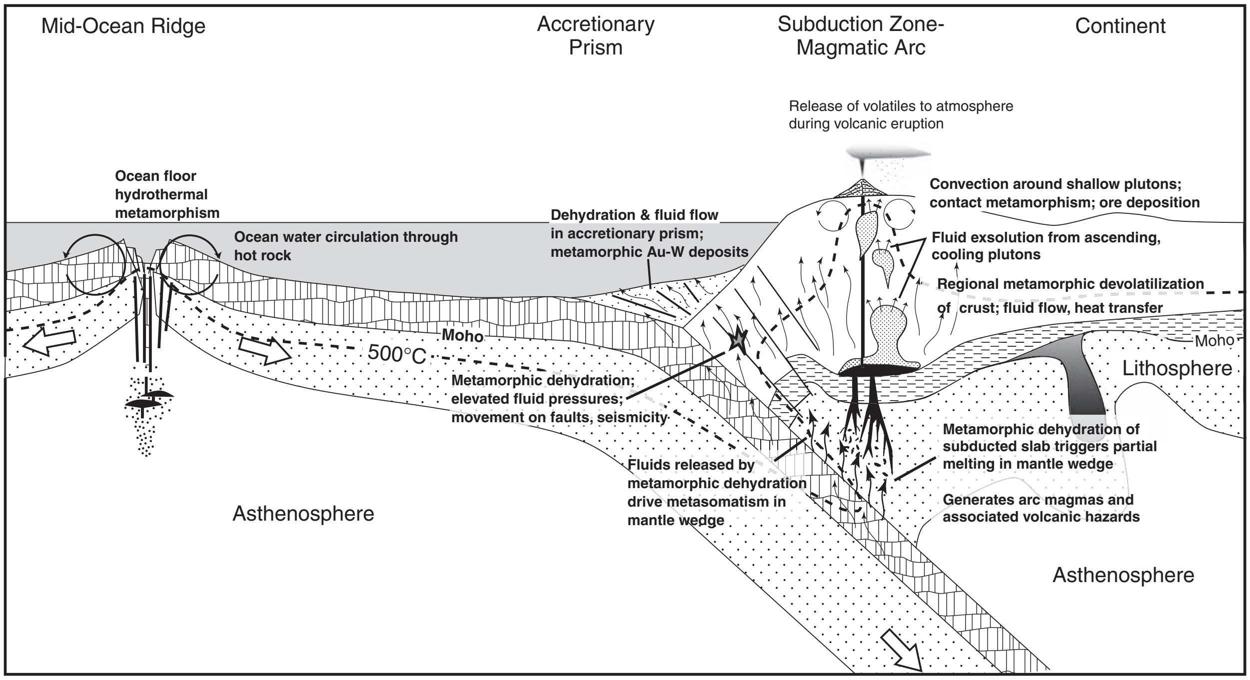


Fig. 21.1 Examples of metamorphic fluid flow and its effects.

gradient in a solution would be an example of the second type. Both forms of transport play important roles in metamorphism and both may operate on an atom simultaneously. Their rates, however, are very different, and thus they are capable of acting over very different distances in the time available during metamorphism.

Inspection of the photomicrographs in Figures 16.4 and 16.7 reveals that considerable redistribution of material is required to transform the mineral assemblages of one metamorphic facies to those of another. Several different transport mechanisms may be involved in a single reaction, such as fluid flow, diffusion through solids, and diffusion along grain boundaries. Each mechanism has a certain characteristic rate, the slowest of which controls the overall rate of the metamorphic reaction.

In this chapter we examine the evidence for mass transfer during metamorphism and the mechanisms by which it can occur. Knowledge of the transport mechanisms is of importance for understanding the amounts, length scales, and timescales of fluid infiltration, as well as the kinetics of metamorphism. If we are to interpret properly the record preserved in a metamorphic rock, it is necessary to know the rates of metamorphic reactions relative to other important geologic rates, such as those of plate motion and conductive heat transfer.

21.2 POROSITY

Transport of mass by fluid flow or by diffusion in a fluid phase takes place through an interconnected network of pores. The porosity (ϕ) is defined as the volume of pore space per unit volume rock. If the pores are completely filled with fluid, as would normally be the case in metamorphism, the porosity value gives the volume fluid per unit volume rock. As indicated in Section 16.2, lithification of sedimentary rocks eliminates much of their porosity. All rocks, however, do retain a small but finite porosity simply because crystal structures of adjoining grains cannot fit together perfectly. The degree of mismatch along grain boundaries depends on the disparity in the structures and orientations of juxtaposed grains. If lattice planes in adjoining crystals match, the grain boundary is said to be *coherent*; if only some lattice planes match, it is *semi-coherent*; and if none match, it is *incoherent*.

Figure 21.2 illustrates possible grain boundaries in a monomineralic rock composed of crystals having a structure based on close-packed spheres. The *intergranular region* consists in part of open spaces, which could be filled with a fluid phase, and zones of disordered structure (Brady, 1983). The width of these zones can range from about 0.1 to 100 nm or more (1 nanometer = 10^{-9} meter). A monomolecular layer of H_2O or CO_2 might be adsorbed onto the walls of the grain boundary channels, and this fluid would not be free to move advectively. Such layers would be approximately 0.5 nm wide. Consequently, advective flow would occur only in channels that are more than 1 nm wide. When the grain boundary width is less than this, movement would be by surface diffusion. Note that in Figure 21.2 channels between grains become more open where they meet other grain boundaries at triple junctions. Thus, although flow might occur between pairs of grains, most should occur in the channels between groups of three grains. Hiraga *et al*. (2001) document relict pores in the ~50 nm to ~500 nm size range at junctions between three or more grains in natural metapelitic schists.

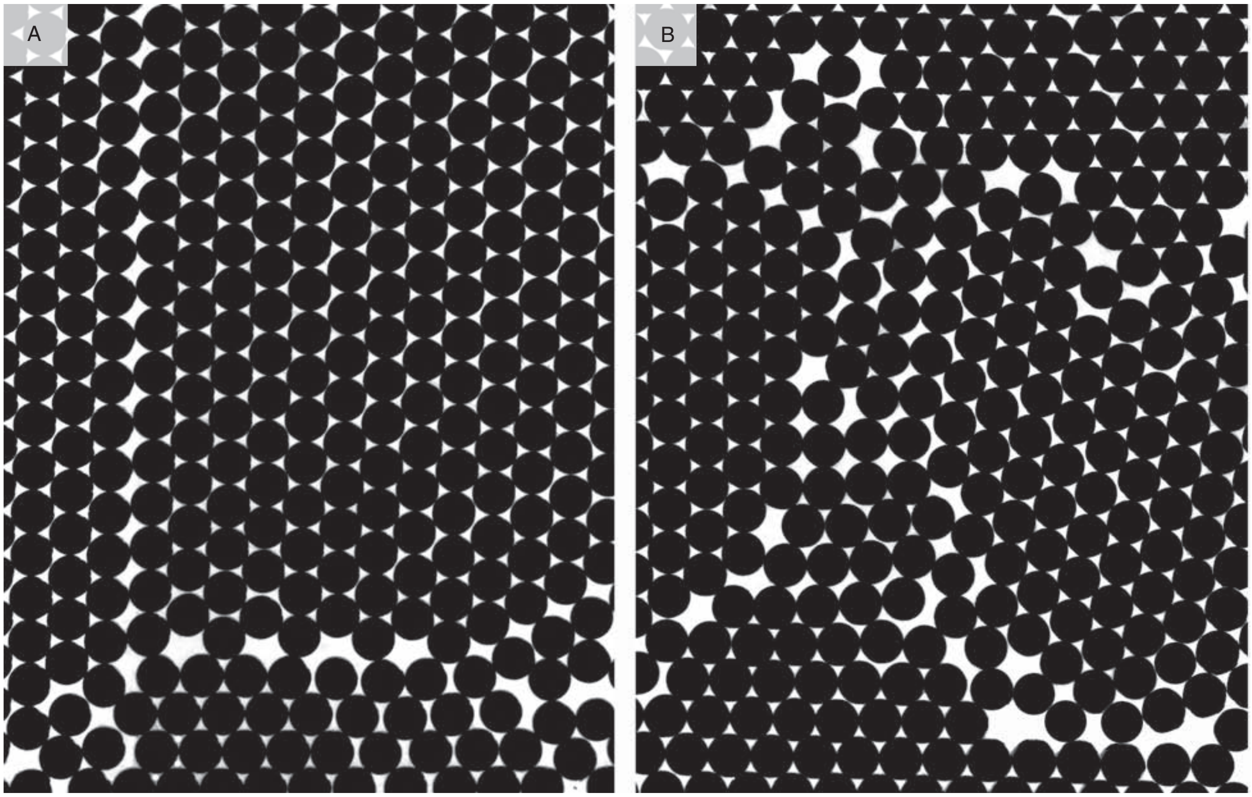


Fig. 21.2 Grain boundaries are regions of misfit between the regular crystal structures of adjoining grains. In this two-dimensional model, crystal structures are represented by regions of close-packed spheres. The amount of open space in intergranular regions depends on the degree of misfit across boundaries. The vertical boundary in (**A**) is a dislocation in the crystal occupying the upper part of the field. The boundary between this crystal and the one below is *semicoherent* because only every second row of atoms can be traced across the boundary; such a boundary is more open than a *coherent* one in which each row can be traced across the boundary. Boundaries in (**B**) are all *incoherent* and are marked by irregular, open intergranular regions. The more incoherent the boundary, the more space there is for intergranular fluids.

It will be recalled from Section 12.7, however, that whether a fluid flows between grains or not depends on the wetting property of the fluid, which can be expressed in terms of the surface free energy between the fluid and the solid (γ_{FS}) relative to the surface free energy between the solids themselves (γ_{SS}). The dihedral angle formed by the fluid in contact with two grains (Fig. 12.29) depends on the relative values of the two surface free energies (Eq. (12.48)). Only when the dihedral angle is less than 60° is the fluid able to penetrate along the entire length of channels between grain edges (Figs. 12.30, 23.13). When the dihedral angle is greater than 60°, the channels close off and the fluid is isolated in pockets at four grain junctions. This criterion applies when the fluid constitutes less than 1% of the volume of the rock. When there is 2% fluid, the critical angle becomes 65°. Dihedral angles are not a factor in low-temperature groundwater flow, because the solids are unable to change their shapes. Under metamorphic conditions, however, mineral grains in contact with a fluid would be expected to adjust their shapes so as to produce equilibrium dihedral angles by dissolving and reprecipitating the solids (Watson and Brenan, 1987; Watson, 1999).

Experiments reveal that dihedral angles produced with supercritical H_2O–CO_2 fluids are commonly large and vary with temperature, pressure, and fluid composition. The dihedral angle formed with quartz increases with temperature and/or fluid CO_2 content, ranging from about 60° to over 90° at 0.4 GPa (Holness, 1993). The dihedral angle for olivine ranges from 65° for H_2O-rich fluid to 90° for CO_2-rich fluid (Watson and Brenan, 1987). The dihedral angle formed with calcite is greater still, being about 140° for H_2O fluids and about 150° for CO_2 fluids (Hay and Evans, 1988). These large angles imply that fluids cannot exist as thin continuous films or channels between grains but, instead, must occur as isolated pockets at four-grain and three-grain junctions (Fig. 21.3). It seems likely, therefore, at least for H_2O-rich compositions, that fluids will penetrate channels along grain edges on quartz grains.

The large measured dihedral angles predict that residual pockets of fluid would be trapped at junctions between several grains, especially when the fluids are CO_2-rich and/or the grains are carbonates. Therefore, it is surprising that the rocks are not vesicular considering the large volumes of fluid that are evolved during metamorphism (Problem 16.2). Although some minerals, such as quartz, do contain small (<5 μm) fluid inclusions, these are typically much smaller than those in many igneous rocks. There is little doubt that most of the metamorphic fluid phase must be effectively purged from the rocks by some mechanism despite the large measured dihedral angles. Surface diffusion might play a role, but as will be shown below, this is a slow process. How else might fluids leave?

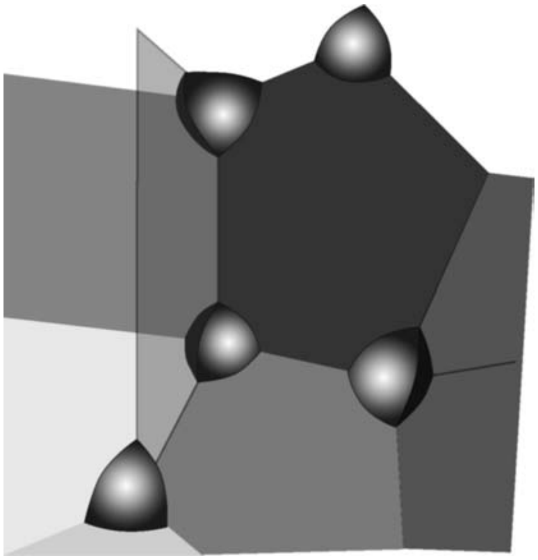


Fig. 21.3 According to experiments, the dihedral angle between minerals and fluids in some metamorphic rocks is greater than 60°. In such cases, fluid occurs as isolated beads at junctions between four grains and is not free to flow through the rock.

It is possible that minor constituents present in natural fluids can reduce dihedral angles to values below 60° and allow channels along grain edges to remain open. Very small amounts of impurities can dramatically affect the surface properties of phases. A small drop of detergent, for example, is all that is necessary to make a whole basin of water wet greasy dishes. Watson and Brenan (1987) found that addition of NaCl to water decreases the dihedral angle with quartz from 57° to 40°.

Another factor is temperature. The dihedral angle for quartz in contact with water decreases from >80° at 600 °C to about 60° at 450 °C (0.4 GPa; Holness, 1993). Study of natural chlorite zone samples by Hiraga *et al*. (2001) suggests that this general trend continues to lower temperatures such that dihedral angles are in the 40° to 50° range at 200–300 °C. Angles measured for quartz–albite contacts were even smaller, although those for a calcite-bearing rock were larger. Thus, dihedral angles <60° may be common in quartzofeldspathic rocks in the lower greenschist facies.

Finally, rocks can fracture as a result of both fluid pressure buildup during devolatilization reactions, and externally imposed tectonic stresses. Hiraga *et al*. (2001) illustrate healed microfractures that were originally filled with fluid. Metamorphic veins are evidence for fracturing on larger scales. The dihedral angles measured in experiments reveal the pore geometry attained when fluid and rock reach thermodynamic equilibrium. Fractures, on the other hand, are inherently nonequilibrium phenomena that can propagate through the rock matrix and connect pores that would otherwise be isolated. Fluid can thus flow when it occupies fractures.

The total porosity of a rock includes both matrix porosity and fracture porosity. Observations of relict pores as well as theoretical calculations suggest matrix porosities of 10^{-3} to 10^{-4} for regional metamorphic rocks (Connolly, 1997; Hiraga *et al*., 2001). The porosity associated with fractures probably lies in a similar range; values as great as $\sim 3 \times 10^{-3}$ have been reported (Ague, 1995). A key point is that only some of the porosity may be interconnected and be able to transmit fluid. For example, the total porosity of fractured rocks in shallow hydrothermal systems (relevant for some contact metamorphic settings) is in the range 0.01 to 0.2 (Norton and Knapp, 1977). The interconnected porosity available to transmit large-scale flow in such systems, however, is considerably smaller – 10^{-3} to 10^{-5} – and is associated primarily with fractures (Norton and Knapp, 1977). For simplicity, in the following discussion we will assume that the porosity, ϕ, is all interconnected.

The amount and nature of porosity will change as metamorphism proceeds. For example, porosity can be created by mineral dissolution, and filled by mineral precipitation. Furthermore, most dehydration reactions have negative volume changes for the solids involved, indicating that the products occupy less space than the reactants. Fracturing may be extensive after events like earthquakes, and this fracture porosity may collapse with time as the rock deforms and readjusts following rupture.

21.3 FLUID FLOW

Fluids migrate along grain boundaries or fractures. The transport rate is determined by the geometry of the channel network, the viscosity of the fluid, and the pressure differential causing flow. The equations governing such flow are similar to those derived in Chapter 3 for the flow of magma through conduits. The differences are that the fluids have much lower viscosities, and the conduits can be much smaller.

The viscosities of supercritical H_2O and CO_2 fluids between 400 and 600 °C at moderate pressures fall between 1×10^{-4} and 2×10^{-4} Pa s (Walther and Orville, 1982). These viscosities are not particularly sensitive to changes in temperature and pressure. Viscosities may also vary with dissolved solids, but these variations are also likely to be small compared with other factors affecting flow.

The fluid pressure gradients driving flow can vary widely. In the upper crust, rocks have considerable strength and can support pore and fracture networks that are connected over many kilometers. In this case, the fluid pressure gradient approximates the *hydrostatic* gradient, given by $-\rho_f g$. The sign is negative for a vertical z coordinate axis with z increasing upward (toward the surface). Because fluid is less dense than rock, the fluid pressure at a given depth is less than the pressure in the surrounding rock, and the difference increases with increasing depth. At 20 km depth, for example, hydrostatic fluid pressures would be roughly 0.4 GPa less than the rock pressure.

Rocks cannot support such large pressure differences at elevated temperatures over long timescales in the middle and lower crust. As a result of their limited strength, they will undergo ductile deformation and recrystallization which collapses the pore space around the fluids, restricts flow, and elevates fluid pressures to values approaching the rock pressure. In this case, the fluid pressure gradient is approximated by the lithostatic gradient: $-\rho_r g$, where ρ_r is the rock density.

Given values for fluid viscosity and the fluid pressure gradient, Eq. (3.16) can be used to estimate the average laminar flow velocity of fluid rising in a planar channel of width $2w$

$$\bar{\upsilon}_z = -\frac{(2w)^2}{12\eta}\left(\frac{\partial P}{\partial z} + \rho_f g\right)$$

We will assume a channel width of 100 nm, and that the fluid has a density of 1000 kg m^{-3} and a viscosity (η) of 10^{-4} Pa s. If the fluid pressure gradient is lithostatic, the terms in parentheses become $g(-\rho_r + \rho_f)$, where ρ_r can be taken as 2800 kg m^{-3}. This produces a pressure gradient of only -0.02 MPa m^{-1}. The calculated velocity is then ~5 m a^{-1}. If a metamorphic fluid originating by devolatilization at 30 km depth was to travel upward unimpeded at this velocity, it would take about 6000 years for it to reach the surface. If the channel width is reduced to 1 nm, the average flow velocity becomes 0.46 mm a^{-1} and it would take about 6×10^7 years for a fluid to travel 30 km. It is important to note that the flow velocity is proportional to the square of the channel width. Therefore, if flow occurs along schistosity planes or prominent fractures having larger $2w$, flow rates would be much greater (Problem 21.1).

If the rate of advance of metamorphic isograds is faster than the rate at which fluid can flow out of the rock, devolatilization reactions would cause the pressure on the fluid to rise above that of the lithostatic pressure. How much excess pressure could be generated in this way depends on the tensile strength of the rock. It was pointed out in Section 3.3 that rocks are able to withstand no more than about 0.03 GPa of excess pressure. This pressure, however, would be sufficient to increase greatly the flow rate of a fluid. Etheridge *et al.* (1984) estimate that pressure gradients of as much as 10 MPa m^{-1} can be generated in this way. This would cause fluid to flow through a 100-nm-wide channel at 2.6 km a^{-1} and through a 1-nm-wide channel at 26 cm a^{-1}. If fluid were not able to escape from the rock as fast as it was produced, the resulting increase in pressure would force open grain boundaries or cause *hydrofracturing* of the rock. The increased width of the channels would then easily accommodate the increased flux of fluids and thus lower the pressure on the fluid phase. In this way a metamorphic rock would behave as its own pressure release valve, which would prevent fluid pressure ever exceeding lithostatic pressure by more than the tensile strength of the rock (about 0.03 GPa).

Fluid flow occurs through a geometrically complex porosity network that includes tubular channels along grain edges, as well as fractures. Modeling the details of flow through these tortuous paths at the scale of individual pore spaces and cracks is a very difficult task. Instead, to obtain the fluid flux through bulk rock we use *Darcy's law* (Section 3.9), which states that the flux of fluid in a given direction through a permeable medium is proportional to the negative pressure gradient on the fluid in that direction. For the z direction (vertical) we can write

$$J_z = -\frac{k}{\eta}\left(\frac{\partial P}{\partial z} + \rho_f g\right) \tag{21.1}$$

where the constant of proportionality, k, is the *permeability*, which has units of square meters, and η is the viscosity. For horizontal flow in the x or y directions, the term in parentheses is simply $\partial P/\partial x$ or $\partial P/\partial y$, respectively. As shown in Chapter 3, Darcy's law has the same general form as Fourier's law of heat conduction (Eq. 5.3) and Fick's first law of diffusion (Eq. 5.51). The permeability of unfractured metamorphic and igneous rocks increases with decreasing grain size (higher percentage of channels) but is about 10^{-18} m^2. Based on a wide array of experimental and field-based data, Ingebritsen and Manning (1999) provide the permeability expression: $k \approx -3.2\log_{10}(Depth) - 14$ for the continental crust, where depth is in km and the uncertainties on permeability are roughly ± one order of magnitude.

The average flow velocity through a rock can be calculated from Darcy's law. The flux of fluid, J_z, is the volume of fluid passing a reference area, A, in a unit time; that is, V/At (e.g. m$^{-3}{}_{(\text{fluid})}$ m$^{-2}{}_{(\text{reference area})}$ a^{-1}; see Fig. 21.4). This flux represents an average over the area, but in detail the fluid

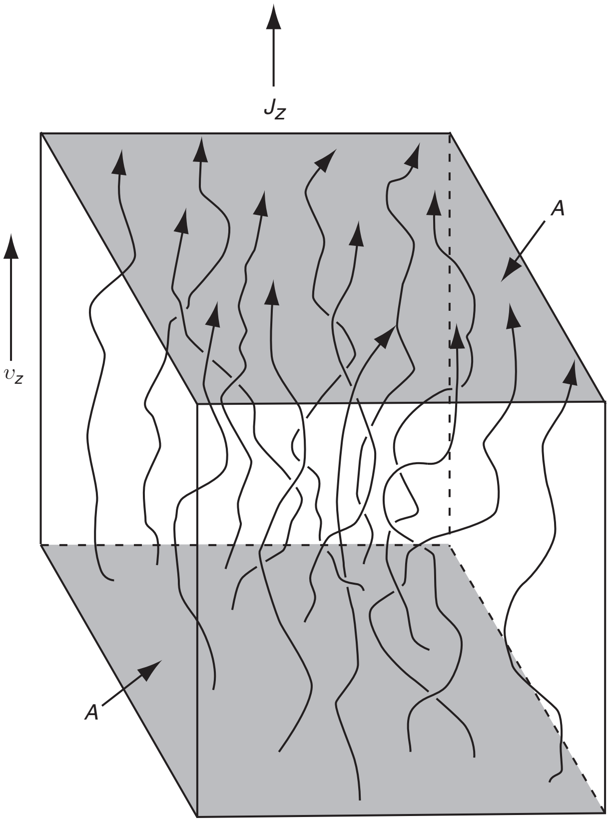


Fig. 21.4 Paths followed by fluid flowing through rock are tortuous and complex. It is more convenient, therefore, to refer to the flux of fluid in a given direction (z); flux is the volume of fluid passing through a reference area A in a unit of time. Dividing flux by porosity, which is the fraction of the area through which fluid travels, gives the average velocity ( ).

migrates through channels that constitute only a fraction of the total area *A*. If the porosity were 100%, the fluid would cross the reference area at every point, and the average velocity, $\bar{v}_z$, would have the same value as the flux (e.g. J_z in $m^3\ m^{-2}\ a^{-1} = \bar{v}_z$ in $m\ a^{-1}$). Thus the flux and the average velocity for fluid in a single planar channel of width $2w$ discussed above have the same value. But in a porous rock, the channels transmitting fluid constitute only the fraction ϕ of the reference area, so the average flow rate must be proportionately greater in these channels to account for the flux; that is,

$$\bar{v}_z = \frac{J_z}{\phi} = -\frac{k}{\phi\eta}\left(\frac{\partial P}{\partial z} + \rho_f g\right) \tag{21.2}$$

from which it follows that

$$J_z = \bar{v}_z\phi \tag{21.3}$$

If we assume a porosity of 0.1%, a permeability of $10^{-18}\ m^2$, and a lithostatic fluid pressure gradient, the average flow velocity (often called the *pore velocity*) is about $5\ m\ a^{-1}$, which is of the same order of magnitude as the calculated flow rate along a single 100-nm-wide channel. The calculation demonstrates that fluids are able to move advectively at significant rates through narrow channels.

Comparison of Darcy's law with the Hagen–Poiseuille law for flow through a pipe (Eq. 3.11)

$$\bar{v}_z = -\frac{r^2}{8\eta}\left(\frac{\partial P}{\partial z} + \rho_f g\right)$$

provides insight into the nature of permeability. Imagine an idealized porous "rock" containing many pipes of radius *r* that are parallel to each other and that run perpendicular to the reference area *A*. In order to obtain the fluid flux across *A* for this "rock," we multiply by the porosity due to the pipes

$$J_z = -\frac{r^2\phi}{8\eta}\left(\frac{\partial P}{\partial z} + \rho_f g\right) \tag{21.4}$$

Comparison with Darcy's law (Eq. 21.1) yields the relationship: $k = r^2\phi/8$. Using $r = 100$ nm and $\phi = 10^{-3}$, we obtain a permeability of about $10^{-18}\ m^2$, in reasonable agreement with measurements. Real porosity is considerably more complex than a collection of parallel pipes, however, so this relationship should not be used to model natural systems. Nonetheless, it reveals that permeability depends solely on the properties of the matrix, and is directly related to the average size and abundance of the through-going channels.

We have treated permeability as though it was constant in time and uniform throughout the rock. However, it can differ by orders of magnitude from one rock layer to the next in heterogeneous metamorphic rock packages. As a result, fluid fluxes can vary spatially during metamorphism, with flow being concentrated in the more permeable layers. In addition, permeability can be *anisotropic* such that it varies with direction in a rock; it is usually largest parallel to layering and metamorphic foliations. Consequently, layer-parallel flow is common in metamorphism due to both permeability heterogeneity and anisotropy (Section 21.9). We saw above that permeability is related to porosity, so changes in porosity due to deformation (e.g. fracturing) or reaction will change permeability with time.

The advective fluid flux transports mass at the regional scale through Earth's crust. The flux of some chemical species *s* due to advection, written here for the *x* direction, is simply the product of the fluid flux and the concentration

$$J_{s,x} = \bar{v}_x\phi c_s \tag{21.5}$$

where c_s is the concentration of *s* in the fluid and $J_{s,x}$ is the mass flux of *s*. For example, if the concentration of Na^+ ions in solution is 100 moles m^{-3}, the pore velocity is $0.1\ m\ a^{-1}$, and the porosity is 10^{-3}, then the flux of Na^+ is 10^{-2} moles m^2 a^{-1}. Note that the product $\bar{v}_x\phi$ is identical to the Darcy flux (Eq. 21.3).

The nature and directions of fluid motion vary with depth in Earth's crust. In the upper crust, fluid pressure gradients can approach the hydrostatic gradient. Under these conditions, thermal convection of fluid is a possibility where temperature gradients are large (Fig. 21.1). Direct measurements from drill holes indicate that near-hydrostatic pressures can persist down to at least 8–10 km depth in some areas (Walther, 1990). Example geologic settings where convection may be important include contact metamorphism, shallow regional metamorphism in areas of elevated heat flow such as rift zones, and hydrothermal ore deposition (e.g. Taylor, 1990; Cui *et al.*, 2001; Norton and Dutrow, 2001). Furthermore, convection is essential for driving seawater circulation during hydrothermal metamorphism at mid-ocean ridges. Large-scale circulation can have a significant effect on thermal evolution during metamorphism. Convection flattens the temperature gradient, and mineral assemblages reflecting a range of pressures at relatively constant temperatures result. The transfer of near-surface waters to depth within convection cells would not only cause a lowering of temperatures but would bring about retrograde hydration reactions.

In Section 14.4 we saw that a dimensionless number, the Rayleigh number (Eq. (14.10)), could be used to indicate whether a cooling sheet of magma would undergo convection. The Rayleigh number was defined as

$$\mathrm{Ra} \equiv \frac{\rho g d^3 \alpha (T_1 - T_2)}{\eta k_\mathrm{T}}$$

where ρ is the density of the magma, g the acceleration of gravity, d the depth or thickness of the sheet, α the coefficient of thermal expansion, T_1 and T_2 the temperatures at the bottom and top of the sheet respectively, η the viscosity of the magma, and k_T the thermal diffusivity of the magma. This equation can be modified so that it is applicable to convection in a porous medium by replacing the d^3 term with the product of the permeability and depth ($k \times d$), which has units of cubic meters. The Rayleigh number for convection in a porous medium therefore becomes

$$\mathrm{Ra} = \frac{\rho g k d \alpha (T_1 - T_2)}{\eta k_{\mathrm{T}}} \tag{21.6}$$

The critical Rayleigh number above which convection takes place in this situation is 40 (Problem 21.2). Note, however, that this equation does not consider the compressibility of the fluid nor its change in density and viscosity with depth. Furthermore, the initial fluid pressure gradient is assumed to be hydrostatic. If devolatilization reactions are proceeding rapidly, however, fluid pressures can rise and fluid pressure gradients can depart from the hydrostat. During prograde contact metamorphism, this will disrupt convection cells and tend to force fluid flow away from plutons or toward the surface (Hanson, 1992).

With these caveats in mind we can use Eq. (21.6) to make a general estimate of the permeability rocks would require for the intergranular fluid to convect. If the fluid has a density of 800 kg $\mathrm{m^{-3}}$, a viscosity of 10^{-4} Pa s, a coefficient of thermal expansion of 10^{-3} $\mathrm{K^{-1}}$, and a thermal diffusivity of 6×10^{-7} $\mathrm{m^2\ s^{-1}}$, and the temperature rises 500 K between Earth's surface and a depth of 10 km, the permeability would have to exceed 3×10^{-17} $\mathrm{m^2}$ for convection to occur. This permeability is not unreasonably large for rocks in the upper crust (Ingebritsen and Manning, 1999), so we can conclude that convection is possible. Indeed, there is strong stable isotopic evidence for circulation of near-surface waters to depths of at least 10 km in continental orogenic belts (Taylor, 1990), and hydrothermal convection of seawater at mid-ocean ridges is a well-known phenomenon.

In the middle and lower crust, fluid pressure gradients likely exceed hydrostatic and approach lithostatic. Under these conditions, flow of fluid upward toward the surface is strongly favored, but downward flow and convection are not. Nonetheless, recent field-based results suggest downward or horizontal flow may be possible in some cases (Wing and Ferry, 2007). Although advective flow under near-lithostatic pressure gradients can be a rapid means of moving material through intergranular porosity, the permeabilty of many metamorphic rocks may be very small if fluid–mineral dihedral angles are large and fluids fail to wet mineral grains. During devolatilization reactions, however, the evolution of volatiles causes a buildup in fluid pressure, which forces apart grains by *hydrofracturing* (Etheridge *et al*., 1984). Both the increased pressure gradient and the dilated channels result in rapid advective movement. This, however, releases the pressure that opens the channels to flow. Therefore, once reaction is complete and the fluids expelled, grain boundary channels heal themselves, and any remaining fluid is trapped as isolated pockets. The permeability of the rock then returns to small values. Fluid flow in many metamorphic rocks is therefore likely to be a pulsating phenomenon driven by devolatilization reactions. Fracturing due to rapid tectonic deformation would also be expected to open up channels and transiently enhance flow.

We can conclude that the depth at which the transition from near-hydrostatic to near-lithostatic fluid pressures takes place has major implications for fluid motion. Large-scale convection is possible above the transition, whereas below it flow will be mostly upward toward the surface (Walther and Orville, 1982). The depth of the transition will vary depending on numerous factors including temperature, rock type, heating rate, and strain rate. In many cases, it may coincide with the brittle–ductile transition (Section 17.5). England and Thompson (1984) argue that convection is not common in regional metamorphic terrains based on the general absence of large-scale retrograde metamorphism, and the lack of evidence for the low geothermal gradients that would be produced by circulating fluids. However, retrograde metamorphism is actually a fairly widespread phenomenon, and thermal gradients can be quite steep on the margins of convection cells. Clearly, more must be known about the properties of rocks and fluids under metamorphic conditions before questions about large-scale fluid flow can be resolved.

21.4 DIFFUSION

We now deal with another important mechanism for the transfer of matter, that of *diffusion*. As we saw in Chapter 5, diffusion occurs when random thermal motion of atoms in the presence of a chemical potential gradient produces a net transfer of atoms down the gradient. In many metamorphic rocks, diffusion is believed to have occurred predominantly along grain boundaries or in the intergranular region (Fig. 21.2). This is based on the knowledge that intergranular regions are zones of misfit between adjoining grains, where abundant defects, vacancies, and impurities all serve to increase diffusion rates. Moreover, reaction rims on minerals in metamorphic rocks commonly completely surround grains, indicating that diffusion is probably more rapid along boundaries than it is into grains (lattice diffusion).

Diffusion through a fluid is described by Fick's first law (Eq. 5.50)

$$J_{s,x} = -D_s \frac{\partial c_s}{\partial x} \tag{21.7}$$

where c_s is the concentration of chemical species s, D_s is the diffusion coefficient for s, and $J_{s,x}$ is the flux of s in the x direction. This equation treats diffusion through a free fluid phase. However, the flux through rock is considerably smaller, because diffusion in the fluid phase can only occur through interconnected pores. Furthermore, the pores and cracks through which diffusion takes place are not straight, but are instead convoluted and tortuous. Consequently, the diffusion flux must be reduced by a tortuosity factor that takes into account the difference between straight pathways and the actual pathways (Bear, 1988).

The effects of porosity and tortuosity are incorporated into the diffusion flux using

$$J_{s,x} = -\phi D_s \tau \frac{\partial c_s}{\partial x} \tag{21.8}$$

where τ is the tortuosity factor which ranges from 0 to 1. Some workers define an *effective* diffusion coefficient for the porous medium as $\phi D_s\tau$, whereas others define it as $D_s\tau$. Most diffusion coefficients for diffusion through a free fluid at metamorphic conditions are of the order of 10^{-8} m^2 s^{-1} (Fig. 5.13). Tortuosity typically varies between about 0.3 and 0.7 (Bear, 1988); we can take 0.5 as a representative value. Thus, for a porosity of 10^{-3}, we calculate an effective diffusion coefficient $\phi D_s\tau = 5 \times 10^{-12}$ m^2 s^{-1}. Coefficients for diffusion through solids are orders of magnitude smaller (Fig. 5.13). As a result, lattice diffusion is commonly neglected when calculating the total diffusive flux.

For intergranular transport through a porous rock, Fick's second law (Eq. 5.55) is written as

$$\frac{\partial(c_s\phi)}{\partial t} = \frac{\partial}{\partial x}\left[\phi D_s\tau\left(\frac{\partial c_s}{\partial x}\right)\right] \tag{21.9}$$

Note that we are interested in the time derivative of the mass of species s in the fluid per unit volume *rock*, so concentration is multiplied by porosity on the left-hand side of the equation. If both the porosity and the effective diffusion coefficient do not vary, then porosity cancels out and this equation reduces to

$$\frac{\partial c_s}{\partial t} = D_s\tau\frac{\partial^2 c_s}{\partial x^2} \tag{21.10}$$

Like other diffusion problems, the characteristic length scale for significant diffusion can be approximated by $x \sim 2\sqrt{(D_s\tau)t}$ (Section 5.6). For 1000 years and $D_s\tau$ = $(10^{-8} \times 0.5)$ m^2 s^{-1}, we calculate $x \sim 25$ m. Of course, the length scale could be considerably smaller if, for example, the diffusing species was consumed by some reaction during transport, or if the porosity of the rock was so small and unconnected that intergranular diffusion through a fluid phase was effectively prohibited. In the latter case, the rate of diffusion could be very slow and approach that for lattice diffusion through the solids.

Theoretical and field-based studies indicate that intergranular diffusion of H_2O and CO_2 can operate over ~1 to ~10 m length scales in devolatilizing rocks during metamorphism (Ague, 2000, 2002; Penniston-Dorland and Ferry, 2006). Length scales for Sr diffusion and reaction have been documented in the range 0.7 to 2 m (Bickle *et al.*, 1997; Baxter and DePaolo, 2002b). These transport length scales are certainly smaller than those for regional-scale advective flow. However, they are clearly of the same magnitude as bedding and layer thicknesses in many metamorphic sequences. Consequently, diffusive transport between layers with different fluid compositions *across* layer contacts may exert strong controls on reaction histories. In this way, large volumes of rock can be affected regionally by diffusion, even if the transport is on the scale of centimeters or meters.

21.5 MECHANICAL DISPERSION

Advection and diffusion are not the only ways to transport material in fluids. Another process, called *mechanical dispersion*, can be very important as well. Imagine that a drop of dye is introduced into a fluid moving through a porous rock at point A on Figure 21.5. The dye will disperse because the flow paths mix due to tortuosity, and because the fluid velocities in adjacent pores differ somewhat due to small differences in pore geometries. As a result, the dye spreads out both in the direction of flow, and transverse to it. The degree of spreading increases with distance from point A. It has been shown that the spreading also increases with increasing pore velocity.

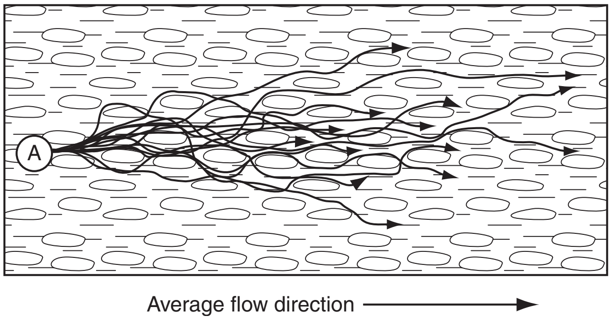


Fig. 21.5 Mechanical dispersion of solute introduced at point A. The mechanical dispersion acts in both the direction of flow (longitudinal dispersion) and transverse to it.

Although mechanical dispersion is the result of advection, its effects are like those of diffusion. Therefore, the process is treated mathematically like diffusion. In order to do this, coefficients of mechanical dispersion, analogous to diffusion coefficients, are defined. Considering flow and transport in one dimension (x), we can write

$$D_{\mathrm{MD,L}} = \alpha_{\mathrm{L}}|\bar{v}_x| \tag{21.11}$$

where α_{L} is the coefficient of dispersivity in the direction of flow (*longitudinal* dispersivity), $D_{\mathrm{MD,L}}$ is the corresponding coefficient of mechanical dispersion, and $|\bar{v}_x|$ is the absolute value of the pore velocity. Values for dispersivity can vary over orders of magnitude (Garven and Freeze, 1984). If we take a representative value of 5 m (Ague, 2000) and a pore velocity of 1 m a^{-1}, then $D_{\mathrm{MD,L}}$ is 1.6×10^{-7} m^2 s^{-1}, more than an order of magnitude greater than the typical value of $D_s\tau$ for diffusion discussed above (5×10^{-9}m^2 s^{-1}). Consequently, dispersion can be a very important mass transport process in active flow systems.

The combined effects of diffusion and mechanical dispersion are collectively referred to as *hydrodynamic dispersion*. The hydrodynamic dispersion flux is given as

$$J_{s,x} = -\phi\left(D_s\tau + D_{\mathrm{MD,L}}\right)\frac{\partial c_s}{\partial x} \tag{21.12}$$

The $D_s\tau + D_{\mathrm{MD,L}}$ term is commonly condensed into a single coefficient that describes both diffusion and mechanical dispersion. Note that in multiple dimensions, the treatment of mechanical dispersion becomes more complicated as dispersion occurs in the direction of flow and transverse to it (Bear, 1988). Transverse dispersivity coefficients may be as much as one to two orders of magnitude smaller than longitudinal dispersivity coefficients (Garven and Freeze, 1984).

21.6 DISSOLUTION OF MINERALS IN SUPERCRITICAL H_2O

Because the porosity of exhumed metamorphic rocks is small, the fluid phase formed during dehydration and decarbonation reactions must be largely lost. The amounts of other material transported during any such reaction depends on the solubilities of the various minerals in the fugitive fluids. In this section we examine the solubilities of some common minerals and the effects of pressure and temperature on their solubility. A detailed discussion of this topic is given by Fyfe *et al*. (1978).

Quartz is the most common vein mineral at most metamorphic grades. We can conclude, therefore, that quartz is the most soluble rock-forming mineral in most metamorphic fluids, a conclusion supported by experimental studies. The silica is present primarily as the neutral $H_4SiO_4^0$ species in solution (the superscript ^{0} indicates neutral). Although quartz is relatively insoluble in surface waters, its solubility in supercritical H_2O increases significantly with both increasing temperature and pressure (Fig. 21.6). For example, at the temperature and pressure of the Al_2SiO_5 triple point (500 °C and 0.38 GPa), H_2O can contain up to 0.7 wt% SiO_2; this increases to 3.4 wt% at 800 °C and 0.38 GPa, and to 6 wt% at 800 °C and 0.8 GPa. At about 1 GPa and 1200 °C, the system quartz–H_2O has a second critical endpoint where the solution contains about 10 wt% SiO_2. Because of the effect of temperature and pressure on the solubility of quartz, silica-saturated hydrous fluids rising toward Earth's surface invariably become supersaturated, which eventually results in precipitation of quartz. This undoubtedly is a factor controlling the abundance of quartz veins in metamorphic terranes.

In addition to temperature and pressure, the composition of the fluid is another important control on solubility. As shown in Figure 21.6(B), decreasing the water activity by addition of CO_2 or large amounts of dissolved salt (NaCl) decreases the solubility of quartz. The effect is particularly strong for CO_2. Thus, water-rich, quartz-saturated fluids that flow into a CO_2-rich environment would be expected to precipitate large amounts of quartz. This scenario might be realized, for example, at contacts between metapelitic and metacarbonate rocks. CO_2-rich fluids with greatly reduced water activities are also present in many granulite facies environments (e.g. Touret, 1985; Crawford and Hollister, 1986; Newton, 1995).

Next to quartz, calcite is the most common vein-forming mineral. Its solubility, however, decreases with increasing temperature, unlike that of quartz. Calcite's solubility does increase with increasing pressure, but this effect is small compared with that of temperature. At the moderate to high pressures and temperatures experienced during metamorphism, calcite and dolomite are relatively insoluble in H_2O fluids. As these fluids migrate toward the surface, their ability to dissolve carbonates continuously increases with falling temperature. Thus fluids released by dehydration reactions at depth could well have the ability to generate solution cleavage in overlying limestones, as long as the temperature of the fluid continues to drop.

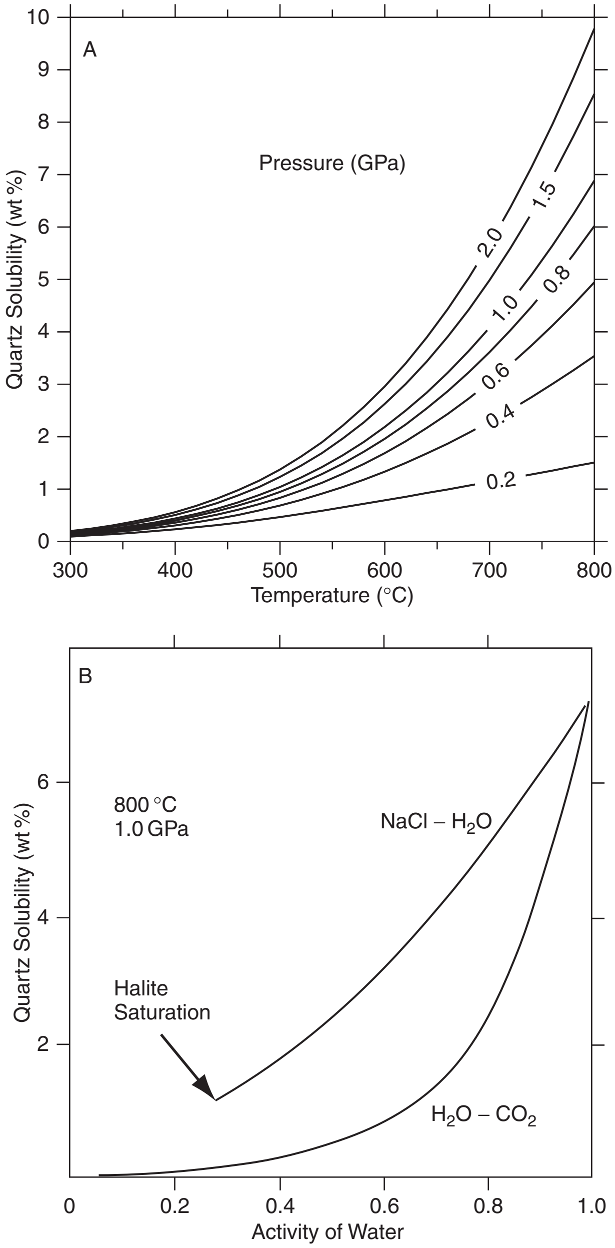


Fig. 21.6 (**A**) Solubility of quartz in supercritical H_2O as a function of pressure and temperature, calculated following Manning (1994). (**B**) Effect of NaCl and CO_2 on quartz solubility at 800 °C and 1.0 GPa (after Newton and Manning, 2000). (Published by permission of Elsevier and *Geochmica et Cosmochimica Acta*.)

Other common rock-forming minerals are not as soluble as quartz in H_2O under most metamorphic conditions. Although H_2O is able to dissolve 0.25 wt% SiO_2 at 500 °C and 0.1 GPa, it can dissolve only 0.08 wt% albite or microcline, and 0.06 wt% enstatite. Like quartz, the solubilities of these minerals increase with temperature and pressure. At pressures in excess of 0.5 GPa and at temperatures above 600 °C, the solubility of alkali feldspar may equal or exceed that of quartz. This could explain why quartz veins in upper amphibolite facies rocks may also contain feldspar. Such veins, however, could result from melting under these conditions (see Fig. 23.18). Vidale (1974), for example, found that veins in metapelitic rocks in Dutchess County, New York, and adjoining Connecticut exhibit a systematic variation in composition with metamorphic grade. Quartz veins are found in all metamorphic grades; quartz + albite veins, however, occur only in a narrow zone just below the biotite isograd; quartz + plagioclase ($An_{20–50}$) veins occur above the staurolite isograd; and quartz + plagioclase + orthoclase veins occur only above the sillimanite + K-feldspar isograd. Quartz + calcite veins also occur in rocks up to the staurolite isograd. These vein compositions probably reflect changes in mineral solubilities with temperature, but they also are controlled by the stability of minerals in the host rock. At low temperatures and high pressures, albite is more soluble than quartz, with the result that albite is a common vein mineral in blueschists.

The solution of most complex silicates takes place incongruently. For example, solutions in equilibrium with alkali feldspar are slightly enriched in alkalis relative to the feldspar composition. Dissolution of feldspars therefore leads to a residue enriched in aluminous minerals (muscovite, Al_2SiO_5, etc.).

Although Al concentrations in fluids are usually thought to be very small, the presence of minerals like kyanite in veins indicates that Al can be transported under some conditions. Manning (2007) concludes that Al and Si may form complexes like $HAlSiO_4^0$ in H_2O fluid. The amount of dissolved Al increases with increasing pressure. Typical metamorphic fluids probably have ~0.002 wt% Al or less at low to moderate pressures (Hauzenberger *et al*., 2001), but Manning (2007) predicts that Al concentrations in water coexisting with kyanite and quartz reach ~0.02 wt% at 1 GPa and 700 ºC. These concentrations may seem small, but Al could still be transported if fluid fluxes were large or if diffusion was able to operate over long timescales.

Metamorphic fluids can contain significant chlorine. Sources of Cl include relict surficial waters such as seawater trapped in pores or fluid inclusions, devolatilization of Cl-bearing minerals like biotite and amphibole, degassing igneous intrusions (e.g. Candela, 2003), and even metamorphosed evaporite sequences. The species in Cl-bearing aqueous solutions include charged ions like Na^+, K^+, and Cl^-, as well as neutral complexes such as $NaCl^0$, KCl^0, KOH^0, and $NaOH^0$. Concentrations are commonly expressed in *molality*, which is simply the moles of species per kg of solvent (H_2O). Total Cl concentrations in most metamorphic fluids are thought to be around one molal or less. These concentrations are relatively small and do not affect the activity of water greatly, so quartz solubility is essentially the same as in Cl-free systems. However, high-grade upper amphibolite and granulite facies rocks sometimes contain extremely Cl-rich minerals or fluid inclusions that indicate the presence of high-Cl fluids and greatly reduced water activities (e.g. Touret, 1985; Crawford and Hollister, 1986; Markl *et al*., 1998).

The amounts of dissolved cations increase with increasing Cl content, and their proportions are dependent on temperature and pressure. Consider a Cl-bearing fluid equilibrated with quartz, albite, muscovite, and kyanite (Fig. 21.7). As temperature decreases, the total amount of Na in solution increases, whereas the total K decreases. These changes become more pronounced at higher Cl contents. As we will discuss in Section 21.10, the behavior of Na and K has important implications for element transport.

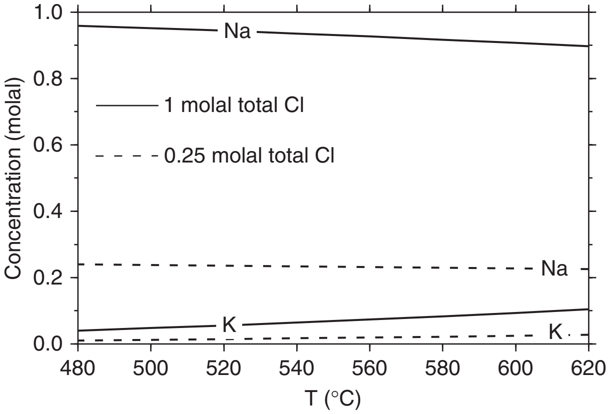


Fig. 21.7 Total Na and K concentrations in aqueous solution coexisting with quartz, albite, muscovite, and kyanite computed for a geothermal gradient of 20 °C per km. Total Cl concentrations are 1 molal (solid lines) and 0.25 molal (dotted lines). The total concentration of K increases as temperature increases, whereas the concentration of Na decreases. The changes with temperature become more pronounced with increasing Cl concentrations in the fluid. (After Ague, 2003b; published by permission of Elsevier.)

When a solution is in equilibrium with a mineral assemblage, its composition is fixed by the chemical potentials of the components in the solid phases (Helgeson, 1967). Consider, for example, a low-temperature metamorphic rock consisting of the four minerals albite, orthoclase, muscovite, and quartz. These minerals can be related through three different exchange reactions with a hydrous fluid:

$$\text{albite} + K^+ \;=\; \text{orthoclase} + Na^+ \tag{21.13}$$

$$3\ \text{orthoclase} + 2H^+ = \text{muscovite} + 6\ \text{quartz} + 2K^+ \tag{21.14}$$

$$3\ \text{albite} + K^+ + 2H^+ = \text{muscovite} + 6\,\text{quartz} + 3Na^+ \tag{21.15}$$

where the ions are in the hydrous solution. Thermodynamic data indicate that the equilibrium constants for these three reactions are 10, 5×10^7, and 10^{11}, respectively, at 300 °C. These equilibrium constants allow us to determine the activities of K^+, Na^+, and H^+ in equilibrium with the mineral assemblage. For Eq. (21.13) the equilibrium constant is written as

$$K = \frac{[a_{Or}][a_{Na^+}]}{[a_{Ab}][a_{K^+}]} = 10 \tag{21.16}$$

At these low temperatures the amount of solid solution in the two feldspars is minimal, so their activities can be taken to be unity. Equation (21.16) therefore reduces to

$$[a_{Na^+}] = 10[a_{K^+}] \tag{21.17}$$

Because this exchange reaction does not involve H^+ ions, it does not define the hydrogen ion activity. Indeed, according to this reaction, coexisting albite and orthoclase could form at any hydrogen ion concentration. Reaction (21.14), however, does involve H^+ ions. The equilibrium constant for this reaction is written as

$$K = \frac{[a_{\mathrm{Musc}}][a_{\mathrm{Qtz}}]^6[a_{\mathrm{K}^+}]^2}{[a_{\mathrm{Or}}]^3[a_{\mathrm{H}^+}]^2} = 5 \times 10^7 \tag{21.18}$$

Assuming unit activities for the minerals, we obtain

$$[a_{\mathrm{K}^+}] = 7071[a_{\mathrm{H}^+}] \tag{21.19}$$

Finally, the equilibrium constant for Eq. (21.15) is

$$K = \frac{[a_{\mathrm{Musc}}][a_{\mathrm{Qtz}}]^6[a_{\mathrm{Na}^+}]^3}{[a_{\mathrm{Ab}}]^3[a_{\mathrm{K}^+}][a_{\mathrm{H}^+}]^2} = 10^{11} \tag{21.20}$$

from which it follows that

$$\left(\frac{a_{\mathrm{Na}^+}}{a_{\mathrm{H}^+}}\right)^3 = 10^{11}\left(\frac{a_{\mathrm{K}^+}}{a_{\mathrm{H}^+}}\right) \tag{21.21}$$

These equilibria are best represented on an activity diagram where the $\log_{10}(a_{\mathrm{Na}^+}/a_{\mathrm{H}^+})$ is plotted against the $\log_{10}(a_{\mathrm{K}^+}/a_{\mathrm{H}^+})$ in the fluid (Fig. 21.8). This is because the reactions plot simply as straight lines using logarithmic axes. According to Eq. (21.19), the coexistence of muscovite and orthoclase occurs when the ratio of $[a_{\mathrm{K}^+}]/[a_{\mathrm{H}^+}] = 7071$; this ratio plots as a vertical line at $\log_{10}(a_{\mathrm{K}^+}/a_{\mathrm{H}^+}) = 3.8$. The line marking the coexistence of orthoclase and albite can be derived from Eq. (21.17). If we divide both sides of (21.17) by $a_{\mathrm{H}+}$, and then take logarithms of both sides we obtain:

$$\log_{10}(a_{\mathrm{Na}^+}/a_{\mathrm{H}^+}) = 1 + \log_{10}(a_{\mathrm{K}^+}/a_{\mathrm{H}^+})$$

Thus the albite–orthoclase reaction has a slope of one and a *y*-intercept of one. The muscovite–orthoclase line becomes metastable on the albite side of the albite–orthoclase reaction, and the albite–orthoclase reaction becomes metastable on the muscovite side of the muscovite–orthoclase reaction. Finally, the third reaction (Eq. (21.13)) passes through this same point of intersection, with a slope given by Eq. (21.21).

The intersection of these three reactions on the activity diagram is an isothermal, isobaric invariant point that defines the ratios of ions in the fluid in equilibrium with orthoclase, muscovite, and albite (+ quartz). In this fluid $\log_{10}(a_{\mathrm{Na}^+}/a_{\mathrm{H}^+}) = 4.9$ and $\log_{10}(a_{\mathrm{K}^+}/a_{\mathrm{H}^+}) = 3.8$ at 300 °C. Fluids emanating from a rock equilibrated with these minerals would have to have these ions in these ratios at this temperature; that is, the composition of the fluid is buffered by the rock.

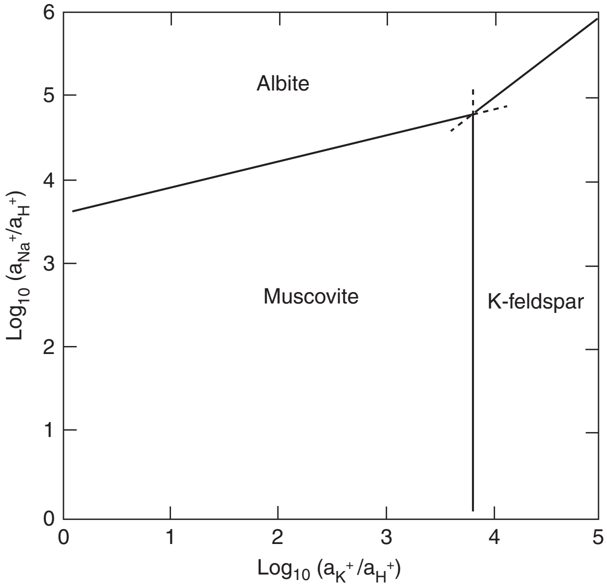


Fig. 21.8 Plot of the stability fields of albite, K-feldspar, and muscovite in terms of activities of Na^+, K^+, and H^+ ions in coexisting fluid.

21.7 MASS TRANSFER MECHANISMS AT A METAMORPHIC ISOGRAD

In previous chapters, metamorphic rocks have been discussed in terms of the development of mineral assemblages having the minimum free energy for a given set of conditions. Thermodynamics, however, does not provide information on how such assemblages are achieved. For this we must have kinetic information on the various paths by which a reaction can take place. Often, reaction paths are extremely circuitous; this is simply because more direct paths are kinetically slower.

The amount of material transport that accompanies a reaction depends on the reaction mechanism. Balanced reactions can be written by equating mineral assemblages above and below an isograd, but this does not tell us how the reaction actually took place. To determine this, we must interpret the textural changes in the rocks across the isograd.

Consider, for example, the reaction that defines the sillimanite isograd in a Barrovian metamorphic series. Below the isograd, rocks of appropriate composition contain kyanite, whereas those above contain sillimanite. The reaction is, therefore, the simple transformation between Al_2SiO_5 polymorphs, and one might expect that sillimanite would simply make its appearance by replacing the thermodynamically unstable kyanite. This, however, rarely happens. In Barrow's type locality, for example, Chinner (1961) showed that sillimanite usually appears first in biotite crystals, not in kyanite. He concludes that sillimanite nucleated more rapidly in biotite than in kyanite, and thus the reaction proceeded by transferring Al_2SiO_5 from the kyanite to nearby biotite grains. Exactly how the Al_2SiO_5 moved is uncertain, but it might have involved advection or diffusion through an aqueous pore fluid.

The first appearance of sillimanite in biotite is a common phenomenon and has been investigated by other workers. Carmichael (1969) has proposed that aluminum is relatively immobile during metamorphic reactions, and if this is the case, the sequence of textural changes across an isograd can be used to deduce the reaction mechanism. He does not believe that the reaction at the sillimanite isograd involves the transfer of aluminum. Instead, he proposes that three related reactions occur, which keep aluminum fixed while other components are transferred through the pore fluid.

The three reactions each take place in separate locations, which are determined by the positions of the minerals that contain the immobile aluminum. Thus, at kyanite grains the following reaction takes place:

$$\begin{aligned}&3\ \text{kyanite} + 3\ \text{quartz} + 2\text{K}^+ + 3\text{H}_2\text{O}\\&\quad \rightleftharpoons 2\ \text{muscovite} + 2\text{H}^+\end{aligned} \qquad (21.22)$$

and at biotite grains the reaction is

$$\text{biotite} + \text{Na}^+ + 6\text{H}^+ \rightleftharpoons \text{albite} + \text{K}^+ + 3(\text{Mg, Fe})^{2+} + 4\text{H}_2\text{O} \qquad (21.23)$$

and where muscovite and albite grains are close, the reaction is

$$\begin{aligned}&2\ \text{muscovite} + \text{albite} + 3(\text{Mg, Fe})^{2+} + \text{H}_2\text{O} \rightleftharpoons\\&\quad \text{biotite} + 3\ \text{sillimanite} + 3\ \text{quartz} + \text{K}^+ + \text{Na}^+ + 4\text{H}^+\end{aligned} \qquad (21.24)$$

The sum of these three reactions is simply *kyanite = sillimanite*, the polymorphic transformation. By having the transformation proceed via the three reactions the textural changes in the rock can be explained (Fig. 21.9). Kyanite is typically rimmed by muscovite at the sillimanite isograd; and biotite is embayed by plagioclase. Finally, sillimanite and biotite are produced together as reaction products at the expense of muscovite and plagioclase. Note that in this interpretation sillimanite and biotite are both products of reaction, whereas in Chinner's interpretation sillimanite is the only product, which grows epitaxially on preexisting biotite.

Two important conclusions can be drawn from this discussion. First, regardless of the simplicity of a metamorphic reaction, the actual mechanism by which a reaction proceeds may be complex and involve circuitous reaction paths that are quite different from the net reaction. The subreactions forming these paths are favored because they provide the fastest route by which the net reaction can proceed. Second, these kinetically favored reaction paths typically involve considerable exchange of material, presumably either by diffusion or advection through an intergranular fluid. Indeed, the cation exchange is so important to the overall progress of the reaction that were a fluid phase not available to transfer the cations or increase diffusion rates, the reaction might not take place or at least it would be forced to proceed by a more sluggish mechanism. As will be seen in Chapter 22, once the fluid phase has been expelled from a metamorphic rock, retrograde reaction rates may become negligible. This, in part, is responsible for the preservation of metamorphic mineral assemblages.

21.8 METASOMATIC ZONATION

In the reactions considered by Carmichael (1969) in Section 21.7, considerable exchange of material was necessary to account for the reaction mechanisms, but no change in the bulk composition of the rock took place, except perhaps for the loss of volatiles. Many reactions, however, involve a change in rock composition, and these are referred to as *metasomatic*. The change may involve large volumes of rock, such as that of a contact metasomatic ore body of magnetite, or it may involve small volumes of rock, such as millimeter- or centimeter-thick zones separating chert nodules or metapelitic layers from enclosing marble (e.g. Fig. 20.3). The scale of metasomatic processes therefore varies considerably.

A rock undergoes a change of composition in order to eliminate chemical potential gradients (Eq. (5.51)). These gradients may result from compositional differences inherited from the protolith, or they may be induced by changes in the composition of the fluid phase. Chemical potential differences also produce reactions between minerals, with the result that many metasomatic rocks are typified by series of sharply defined reaction layers, which are characterized by rocks with relatively small numbers of minerals. The metasomatic transfer of material and the presence of reaction

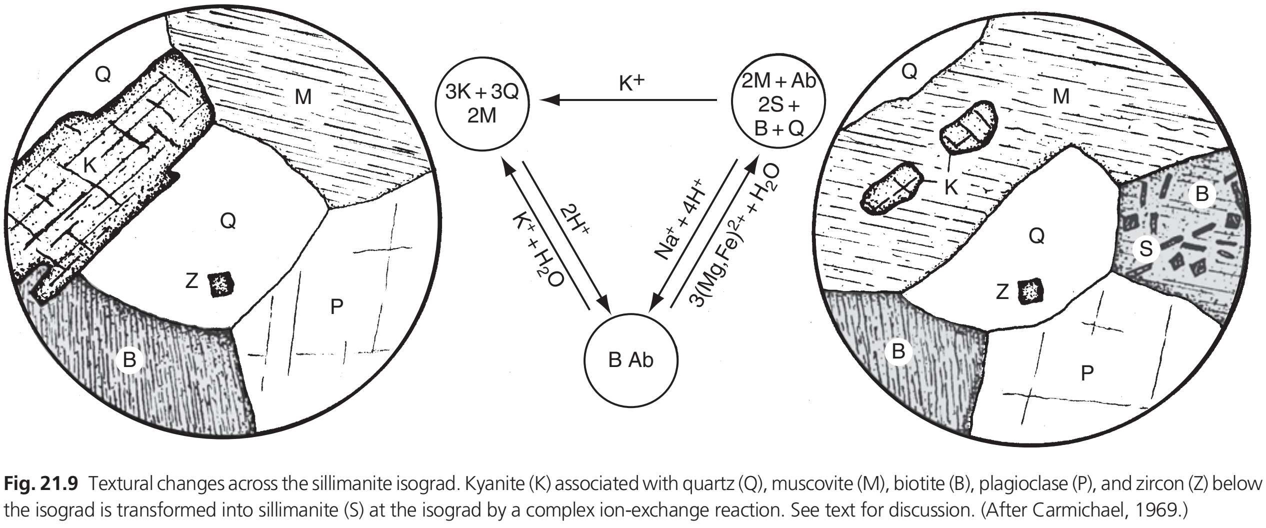


Fig. 21.9 Textural changes across the sillimanite isograd. Kyanite (K) associated with quartz (Q), muscovite (M), biotite (B), plagioclase (P), and zircon (Z) below the isograd is transformed into sillimanite (S) at the isograd by a complex ion-exchange reaction. See text for discussion. (After Carmichael, 1969.)

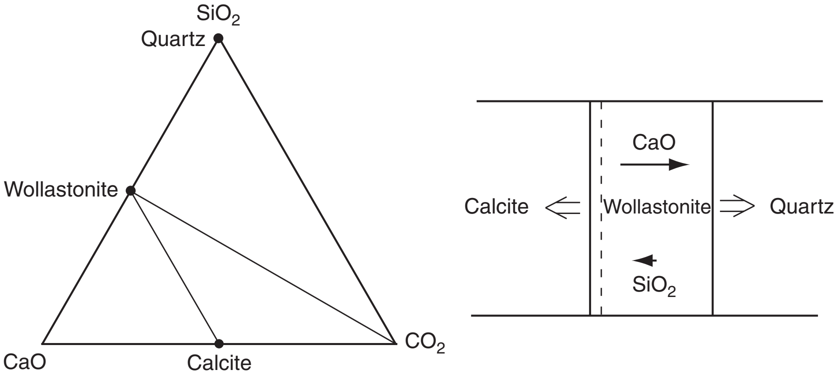


Fig. 21.10 Heating of chert nodules in limestone results in the formation of a reaction rim of wollastonite on the nodules. Growth of the wollastonite depends on the diffusion of CaO and SiO_2 through the reaction rim. Because CaO diffuses more rapidly than SiO_2, wollastonite grows more rapidly on the quartz side of the rim than it does on the calcite side; the lengths of the two arrows indicate the relative magnitudes of the diffusion coefficients for CaO and SiO_2. The original boundary between the chert nodule and the limestone is indicated by the dashed line. (After Joesten and Fisher, 1988.)

layers are clear evidence that metasomatic rocks as a whole do not represent equilibrium products. Equilibrium thermodynamics, however, can be used to interpret the mineral assemblages in these rocks as long as sufficiently small volumes of rock are considered. The idea of *local equilibrium* is central to the interpretation of metasomatic mineral assemblages proposed by Korzhinskii (1970) and Thompson (1959).

Let us consider the metasomatic layers that develop around a chert nodule in a contact metamorphosed limestone, such as that of the Christmas Mountains, Texas, shown in Figure 20.3 (Joesten, 1976; Joesten and Fisher, 1988). On being heated, coexisting quartz and calcite become unstable, and a reaction layer of wollastonite forms between them. For simplicity, we will assume that the fluid phase is pure CO_2 ($X_{CO_2} = 1$). With time, the incompatible phases, quartz and calcite, are separated by a progressively thickening layer of wollastonite (Fig. 21.10). The equilibrium assemblage quartz + wollastonite + CO_2 is found on one boundary of this reaction layer, and the equilibrium assemblage calcite + wollastonite + CO_2 is found on the other boundary. Thus, although the system as a whole is not at equilibrium, local equilibrium exists at the boundaries on either side of the reaction layer. Diffusion through the reaction layer allows more wollastonite to form, but it does not change the equilibrium assemblages on either side of the reaction layer.

Diffusion through the reaction layer is caused by the chemical potential gradient set up by the compositional difference between the chert and the calcite (Joesten, 1977, 1979). All minerals can be considered to be in local equilibrium with an intergranular phase, which is largely the fluid but would also include the disordered regions on grain boundaries; we will refer to this as the fluid phase. The chemical potentials of components in the fluid are determined by the chemical potentials of the components in the local minerals. These chemical potentials can be defined using simple reactions between the minerals and the fluid. They are as follows:

$$\begin{array}{ccc}\text{(mineral)} & & \text{(fluid components)}\\ CaCO_3 & \rightleftharpoons & CaO + CO_2\end{array} \tag{21.25}$$

$$CaSiO_3 \rightleftharpoons CaO + SiO_2 \tag{21.26}$$

$$SiO_2 \rightleftharpoons SiO_2 \tag{21.27}$$

If each mineral is in local equilibrium with the fluid, we can use Eq. (9.5) to write

$$\begin{array}{cc}\text{(mineral)} & \text{(fluid)}\\ \mu_{\text{calcite}} = \mu^f_{CaO} + \mu^f_{CO_2}\end{array} \tag{21.28}$$

$$\mu_{\text{wollastonite}} = \mu^f_{CaO} + \mu^f_{SiO_2} \tag{21.29}$$

$$\mu_{\text{quartz}} = \mu^f_{SiO_2} \tag{21.30}$$

The stability of these minerals can, therefore, be expressed in terms of the chemical potentials of CaO, SiO_2, and CO_2. Because the fluid phase is taken to be pure CO_2, the chemical potential of CO_2 is determined only by temperature and pressure. Thus, at a given temperature and pressure, only the chemical potentials of CaO and SiO_2 are variables. We therefore show the stability relations between the minerals in an isothermal, isobaric plot of μ_{CaO} versus μ_{SiO_2} (Fig. 21.11).

For a fluid to be in equilibrium with quartz, $\mu^f_{SiO_2}$ must have the value defined by Eq. (21.30). At a temperature of 700 °C and a pressure of 35 MPa, which corresponds to the depth at which wollastonite layers developed around the chert nodules in the contact metamorphic aureole of the Christmas Mountains intrusion (Joesten, 1976), $\mu^f_{SiO_2}$ would, according to the data in Table 7.1 and Eq. (7.45), be -977 kJ mol^{-1}. This chemical potential plots as a vertical line in Figure 21.11. For a fluid to be in equilibrium with calcite, μ^f_{CaO} must have a value defined by Eq. (21.28). From the data in Table 7.1, μ_{calcite} is -1345 kJ mol^{-1}. The $\mu^f_{CO_2}$ is calculated from the data in Tables 7.1 and 20.1, and Eq. (8.7) to be -574 kJ mol^{-1}. Consequently, the μ^f_{CaO} in the fluid in equilibrium with calcite under these conditions must be $\mu_{\text{calcite}} - \mu^f_{CO_2}$; that is, -771 kJ mol^{-1}. This plots as a horizontal line in Figure 21.11. Finally, fluids in equilibrium with wollastonite can have a range of chemical potentials of CaO and SiO_2 as long as they satisfy Eq. (21.29). For example, $\mu_{\text{wollastonite}}$ is -176 kJ mol^{-1}, and if we take the chemical potential of

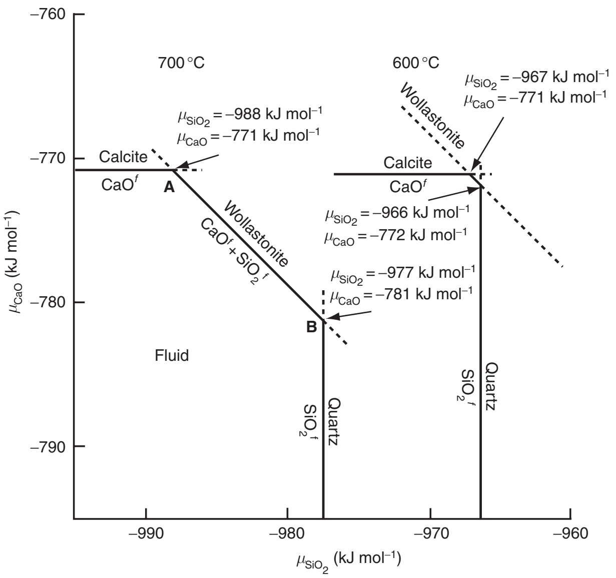


Fig. 21.11 Isobaric (P = 35 MPa), isothermal (T = 700 °C) plot of the chemical potentials of CaO and SiO_2 in fluids in equilibrium with calcite, wollastonite, and quartz. Chemical potentials of CaO and SiO_2 are uniquely defined where calcite and wollastonite are in contact and where wollastonite and quartz are in contact. Differences in chemical potential between these two boundaries drive the diffusion through the wollastonite layer. At 600 °C the wollastonite saturation surface only just intersects the surfaces for calcite and quartz. At temperatures below this, wollastonite is unstable with respect to calcite and quartz, and then the wollastonite saturation surface does not produce a stable intersection with the saturation surfaces of calcite and quartz. (After Joesten and Fisher, 1988.)

SiO_2 to be that which would be in equilibrium with quartz (–977 kJ mol^{-1}), then μ^f_{CaO} is –781 kJ mol^{-1}. On the other hand, if the μ^f_{CaO} is that which is in equilibrium with calcite (–771 kJ mol^{-1}), then $\mu^f_{SiO_2}$ is –988 kJ mol^{-1}. These two points define a line in Figure 21.11 indicating fluids with chemical potentials of CaO and SiO_2 that would be in equilibrium with wollastonite. From the stoichiometry of Eq. (21.26), the slope of this line must be 45°. Indeed, the slope of any reaction line in an isothermal, isobaric μ_i versus μ_j plot can be determined from the stoichiometry of the reaction as follows:

$$\frac{d\mu_i}{d\mu_j} = \frac{-v_j}{v_i} \tag{21.31}$$

where v_i and v_j are the stoichiometric coefficients for the constituents i and j, respectively, in the balanced reaction, with the coefficients being positive if they are products and negative if they are reactants.

Note in Figure 21.11 that the line representing the equilibrium between fluid and calcite is metastable once it crosses the line marking the fluids coexisting with wollastonite, because here the chemical potential of SiO_2 in the fluid is sufficiently high to react with CaO to produce wollastonite. Furthermore, the wollastonite line is metastable once it crosses the calcite line, because at these high values of μ_{CaO} calcite forms in place of CaO. Similar arguments apply to the intersection of the wollastonite and quartz lines. The intersecting lines delineate a *saturation surface*, which must be a convex polyhedron. Within the surface, fluid is undersaturated with respect to solids, but at the surface, the chemical potentials are high enough to precipitate a mineral or minerals. At point A, where the calcite and wollastonite saturation surfaces intersect, fluid with a $\mu^f_{CaO} = -771$ kJ mol^{-1} and $\mu^f_{SiO_2} = -988$ kJ mol^{-1} is in equilibrium with both calcite and wollastonite; at point B, where $\mu^f_{CaO} = -781$ kJ mol^{-1} and $\mu^f_{SiO_2} = -977$ kJ mol^{-1}, fluid is in equilibrium with wollastonite and quartz.

At different temperatures (and pressures) the position of the saturation surface and the points of intersection of individual saturation lines change. In Figure 21.11, a saturation surface is also shown for 600 °C. Note that at this temperature, the calcite saturation line is approximately the same as that at 700 °C, but the quartz saturation line shifts to higher values (less negative). More important, however, the wollastonite saturation line also shifts to higher values and almost does not produce a stable intersection with the calcite and quartz saturation lines. At temperatures just below 600 °C (at P = 35 MPa), the wollastonite saturation line plots above the intersection of the calcite and quartz saturation lines and is, therefore, entirely metastable. This simply means that these conditions are below the stability field of wollastonite, and instead, calcite + quartz is stable.

We are now in a position to evaluate the driving force causing diffusion through the wollastonite layer separating the calcite and quartz at 700 °C. On one side of this layer, coexisting calcite and wollastonite buffer the chemical potentials of SiO_2 and CaO at –988.0 and –770.7 kJ mol^{-1}, respectively (A in Fig. 21.11), whereas on the other side, coexisting quartz and wollastonite buffer these potentials at –977.4 and –781.3 kJ mol^{-1}, respectively (B in Fig. 21.11). The chemical potentials of SiO_2 and CaO both decrease by 10.6 kJ mol^{-1} through the wollastonite layer. The gradients of these two components are therefore equal but in opposite directions.

When diffusion occurs in multicomponent systems, interdiffusion coefficients can be written for each pair of constituents. These are referred to as cross coefficients and are represented by D_{ij}, which refers to the diffusion of component i through a gradient of j. If the gradient is expressed in terms of chemical potential, the coefficient is represented by L_{ij}. These terms are then incorporated into Onsager's extension of Fick's equation by summing the effects of all the components (Fisher, 1973); this gives

$$J_i = \sum_{j=1}^{n} L_{ij}\left(-\frac{d\mu_i}{dx}\right) \tag{21.32}$$

If the Onsager diffusion coefficients for CaO and SiO_2 through the intergranular channels in the wollastonite were the same, the identical chemical potential gradients would cause equal amounts of CaO and SiO_2 to diffuse through the wollastonite layer. This would result in equal amounts of growth on both sides of this layer. Joesten and Fisher

(1988), however, have shown that almost all of the growth takes place on the wollastonite/quartz boundary; that is, CaO diffuses much more rapidly than SiO_2. Separate diffusion coefficients for CaO and SiO_2 cannot be determined from the field data. The thickness of reaction layers is certainly controlled by diffusion rates, but these rates would have varied as the rocks were heated and then cooled during the magmatic episode, and diffusion coefficients may also have varied as a function of composition. The layer thickness therefore represents the cumulative growth of the wollastonite formed while the rock was above 600 °C. Joesten and Fisher, however, were able to extract an Onsager diffusion ratio of $L_{\mathrm{CaO}}/L_{\mathrm{SiO_2}} = 42$. With this ratio, the reactions on the two boundaries of the wollastonite layer can be written as follows (see Fig. 21.10): at the calcite/wollastonite boundary

$$26\ \text{calcite} + 0.6\mathrm{SiO_2} \rightleftharpoons 0.6\ \text{wollastonite} + 25.4\mathrm{CaO} + 26\mathrm{CO_2} \tag{21.33}$$

and at the wollastonite/quartz boundary

$$26\ \text{quartz} + 25.4\mathrm{CaO} \rightleftharpoons 25.4\ \text{wollastonite} + 0.6\mathrm{SiO_2} \tag{21.34}$$

Nearer the igneous contact, where temperatures exceeded 941 °C, the mineral tilleyite formed between calcite and wollastonite around the chert nodules (Fig. 20.3). The position and orientation of the saturation surface for this mineral in the μ_{CaO}–$\mu_{\mathrm{SiO_2}}$ diagram can be evaluated from the stoichiometry of the reaction relating the mineral to its components

$$\mathrm{Ca_5Si_2O_7(CO_3)_2} = 5\mathrm{CaO} + 2\mathrm{SiO_2} + 2\mathrm{CO_2} \tag{21.35}$$

from which it follows that $\mathrm{d}\mu_{\mathrm{CaO}}/\mathrm{d}\mu_{\mathrm{SiO_2}} = -2/5$. Its slope is shallower than that of the saturation surface of wollastonite (Fig. 21.12). Consequently, the tilleyite layer forms between the calcite and wollastonite; this is also evident from the position of the mineral with respect to calcite and wollastonite in the $CaO–SiO_2–CO_2$ mineral facies diagram (Fig. 20.3 (D)). This shallower slope also means that the chemical potential difference of CaO across this layer is less than it is across the wollastonite layer. It is not surprising, therefore, to find that although the wollastonite layer grows mainly at the boundary with quartz, considerable growth of the tilleyite layer occurs at the boundary with calcite (Fig. 21.12).

Although two components (plus CO_2) are necessary to describe the minerals in the various layers around the chert nodules in the limestone, individual reaction layers are composed of just one mineral. For metasomatically layered rocks in general, the maximum number of phases in any local thermodynamic system (individual layer) is equal to the number of components necessary to define all phases present in all the layers less the number of independent compositional gradients. In the case of the chert nodules, two components, CaO and SiO_2, are necessary to describe all the phases (we have assumed that a fluid phase of pure CO_2 is present), and there is one independent compositional gradient, either CaO or SiO_2. Consequently, only one phase is present in each layer. At layer boundaries, reactions buffer gradients, so that there are no independently variable gradients, and consequently, two minerals can coexist.

21.9 ESTIMATING FLUID FLUXES DURING METAMORPHISM USING GEOCHEMICAL FRONTS

The discussion in previous sections of this book should make it clear that mass transfer is an important aspect of metamorphism. Regional scale advection through mountain belts is capable of changing the composition of rocks and, it may, if great enough, be able to transfer sufficient heat to become an important agent causing metamorphism. This latter topic is dealt with in Chapter 22. But first it is necessary to be able to estimate the volumes of fluid flux during metamorphism. The basic idea is that if fluids flow through the rock and react with it, then the amount of reaction is in some way proportional to the amount of flow. If the amount of reaction can be determined, therefore, then fluid fluxes can be estimated. These estimates require quantitative expressions that relate fluid transport processes (advection, diffusion, mechanical dispersion) and fluid–rock reaction. Flux estimates are usually given on a time-integrated basis – that is, the total volume of fluid that has passed over a unit area of rock – because the timescales of flow are not always known. This section contains some fairly intense derivations, but the practical results, such as Eq. (21.60), are straightforward.

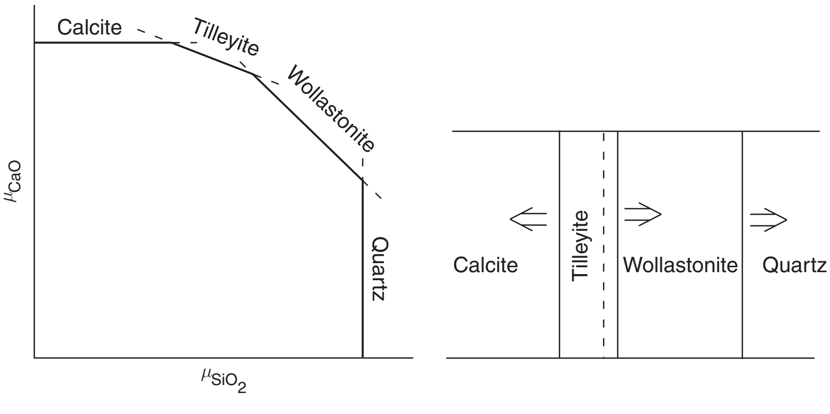


Fig. 21.12 At temperatures above 941 °C at $P = 35$ MPa, the saturation surface for the mineral tilleyite intersects the saturation surfaces for calcite and wollastonite in a μ_{CaO} versus $\mu_{\mathrm{SiO_2}}$ plot. Consequently, tilleyite forms a reaction rim between the wollastonite and calcite on metamorphosed chert nodules in limestone. The dashed line represents the original calcite/chert boundary. (After Joesten and Fisher, 1988.)

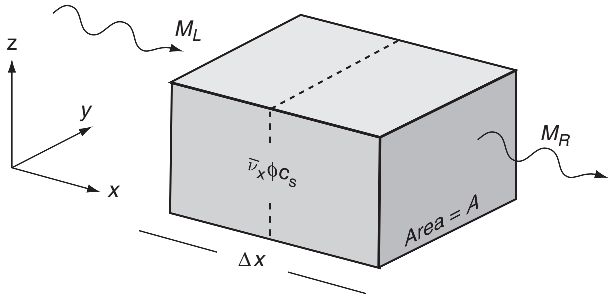


Fig. 21.13 Control volume of length Δ*x* and cross-sectional area *A*. M_L is the total mass flux of chemical species *s* (mole time^{-1}) entering the volume from the left due to fluid advection, and M_R is the mass flux leaving on the right. The flux *per area* across the plane in the center of the volume (dashed outline) is given by $\bar{v}_x \phi c_s$(mole area^{-1} time^{-1}). The total mass flux across this plane is thus given by: $\bar{v}_x \phi c_s \cdot A$ (mole time^{-1}). See text for further discussion.

We begin by deriving an expression that describes how the concentration of some species *s* in the fluid evolves in space and time during advective flow through a porous rock. This can be done using a derivation very similar to that used for Fourier's equation (Eq. (5.11)). Fourier's equation is based on the conservation of energy, whereas here we consider conservation of mass. Consider a tiny block of porous rock with length Δ*x* and cross-sectional area *A* through which fluid flows only in the *x* direction (Fig. 21.13). The flux of some species *s* at the center of the block is $\bar{v}_x \phi c_s$ (mole area^{-1} time^{-1}; Eq. (21.5)), and there is a gradient in the flux through the block given by $\partial(\bar{v}_x \phi c_s)/\partial x$. The total mass flux of *s* (mole time^{-1}) in through the left side of the block, M_L, is thus

$$M_\mathrm{L} = \left(\bar{v}_x \phi c_s - \frac{1}{2} \Delta x \frac{\partial(\bar{v}_x \phi c_s)}{\partial x} \right) A \tag{21.36}$$

and the total mass flux out through the right side is

$$M_\mathrm{R} = \left(\bar{v}_x \phi c_s + \frac{1}{2} \Delta x \frac{\partial(\bar{v}_x \phi c_s)}{\partial x} \right) A \tag{21.37}$$

The amount of mass of *s* that accumulates in, or is lost from, the block per time is then given by the difference between the inflow and the outflow

$$M_\mathrm{L} - M_\mathrm{R} = -\Delta x A \frac{\partial(\bar{v}_x \phi c_s)}{\partial x} \tag{21.38}$$

where mass gain is positive and mass loss is negative. We can come at the problem another way by remembering that the mass of *s* in the fluid per unit volume rock, m_s, is equal to porosity times the concentration: $m_s = \phi c_s$. Taking the time derivative of this relationship yields

$$\frac{\partial m_s}{\partial t} = \frac{\partial(\phi c_s)}{\partial t} \tag{21.39}$$

Equation (21.39) is written per unit volume rock; we can also write it specifically in terms of our model block (Fig. 21.13) by just multiplying by the block's volume

$$\Delta x A \frac{\partial m_s}{\partial t} = \Delta x A \frac{\partial(\phi c_s)}{\partial t} \tag{21.40}$$

Equation (21.40) gives the rate of mass accumulation or loss within the block, and is thus equivalent to Eq. (21.38). Setting these two equations equal yields the final result

$$\frac{\partial(\phi c_s)}{\partial t} = -\frac{\partial(\bar{v}_x \phi c_s)}{\partial x} \tag{21.41}$$

which we can solve to investigate how the mass of species *s* changes with time due to advection of fluid through rock. For advection in three dimensions, we write an expression analogous to Eq. (5.11)

$$\frac{\partial(\phi c_s)}{\partial t} = -\frac{\partial(\bar{v}_x \phi c_s)}{\partial x} - \frac{\partial(\bar{v}_y \phi c_s)}{\partial y} - \frac{\partial(\bar{v}_z \phi c_s)}{\partial z} \tag{21.42}$$

To avoid having to list all the terms, the right hand side of this equation is usually written as

$$\frac{\partial(\phi c_s)}{\partial t} = -\nabla \cdot (\vec{v} \phi c_s) \tag{21.43}$$

in which $\nabla \cdot$ is the *divergence* from multivariable calculus (also written as: div), and $\vec{v}$ is the pore velocity vector.

Relationships similar to Eq. (21.41) hold for diffusion and mechanical dispersion transport as well, so we can write the general equation for mass conservation due to transport in the *x* direction as

$$\frac{\partial(\phi c_s)}{\partial t} = -\frac{\partial(J_{s,x})}{\partial x} \tag{21.44}$$

where $J_{s,x}$ is the flux of *s* due to advection, diffusion, and/or mechanical dispersion (Haase, 1990). The total flux due to these processes is simply their sum (Eqs. (21.5), (21.12))

$$J_{s,x} = \bar{v}_x \phi c_s - \phi D_\mathrm{HD} \frac{\partial c_s}{\partial x} \tag{21.45}$$

where D_HD is the coefficient of hydrodynamic dispersion incorporating the effects of both diffusion and mechanical dispersion. Substituting this equation into Eq. (21.44) yields

$$\frac{\partial(\phi c_s)}{\partial t} = -\frac{\partial(\bar{v}_x \phi c_s)}{\partial x} + \frac{\partial\left(\phi D_\mathrm{HD} \dfrac{\partial c_s}{\partial x}\right)}{\partial x} \tag{21.46}$$

which is the full expression for conservation of mass due to transport in the *x* direction. Note, however, that chemical reaction is not included, a point we will return to below.

If porosity, pore velocity, and the coefficient of hydrodynamic dispersion don't vary, then (21.46) simplifies considerably to

$$\frac{\partial c_s}{\partial t} = -\bar{v}_x \frac{\partial c_s}{\partial x} + D_\mathrm{HD} \frac{\partial^2 c_s}{\partial x^2} \tag{21.47}$$

As we saw in Chapter 5, in order to solve partial differential equations like Eq. (21.47), we need to specify initial and boundary conditions. Consider a rock mass whose pore

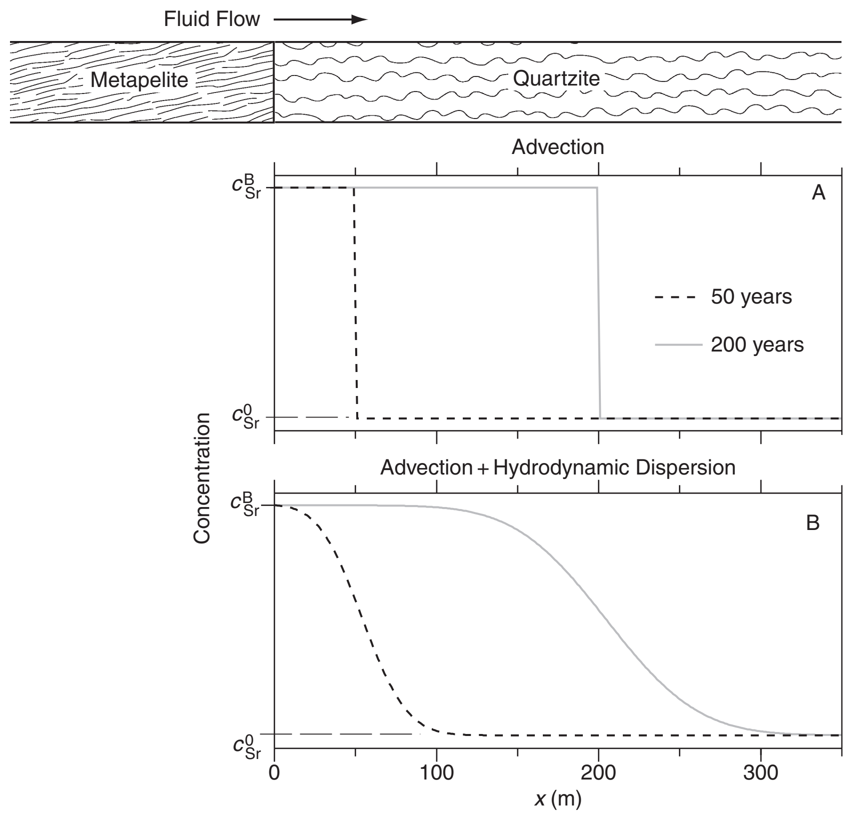


Fig. 21.14 Hypothetical concentration profiles resulting from advection and hydrodynamic dispersion of Sr from a model metapelitic rock into a pure quartzite computed using Eq. (21.48). (**A**) Concentration profiles in the fluid for pure fluid flow (advection): $c^{\mathrm{B}}_{\mathrm{Sr}}$ is the input concentration of Sr in the fluid at the contact at x=0, and c^{0}_{Sr} is the initial concentration in the quartzite ($x > 0$). Solute front propagates as a sharp "step." (**B**) Transport by advection and hydrodynamic dispersion. Note broadening of concentration profiles relative to part (A). The degree of broadening increases with increasing transport distance.

space contains a fluid that has an initial concentration of s equal to c_s^{o} (Fig. 21.14). For the boundary condition at $x = 0$, fluid of constant composition c_s^{B} is input into the rock; flow and hydrodynamic dispersion proceed in the positive x direction (to the right in Fig. 21.14). The solution for this case is well known (Carslaw and Jaeger, 1959, p. 388; Ogata and Banks, 1961) and can be calculated using a spreadsheet program

$$c_s^{x,t} = c_s^{\mathrm{o}} + \left(c_s^{\mathrm{B}} - c_s^{\mathrm{o}}\right)\frac{1}{2}\left(\operatorname{erfc}\left[\frac{x - \bar{v}_x t}{2\sqrt{D_{\mathrm{HD}} t}}\right] + \exp\left[\frac{\bar{v}_x x}{D_{\mathrm{HD}}}\right]\operatorname{erfc}\left[\frac{x + \bar{v}_x t}{2\sqrt{D_{\mathrm{HD}} t}}\right]\right) \tag{21.48}$$

where $c_s^{x,t}$ is the concentration of s at position x and time t, and erfc is the complimentary error function erfc(y) = 1 − erf(y); see Chapter 5 for a discussion of the error function. One could envision, for example, fluid infiltration from a dehydrating metapelitic schist into a pure quartzite (metasandstone) across a lithologic contact. Plagioclase that contains a small amount of Sr is present in the metapelite; the metapelitic fluid will thus have a trace amount of Sr dissolved in it as well. Because quartz will not consume any of the Sr, Eqs. (21.47) and (21.48) for infiltration without reaction are appropriate. Consequently, we can use the Sr as a tracer to track infiltration into the quartzite, and see how fluid compositions can change with time due to infiltration.

Although all real systems will have some amount of hydrodynamic dispersion, let us first consider the simple case of transport by advection only. The concentration of the Sr tracer is shown for two different times in Figure 21.14(A) for a pore velocity of 1 m a^{-1}. The infiltrating fluid displaces the original pore fluid; the boundary separating the two is often called the *hydrodynamic front*. This boundary coincides with a sharp change in concentration known as the *solute front*. For pure advection, the distance of hydrodynamic and solute front propagation, d_{F}, is simply given by the product of the pore velocity and the total time t of flow (distance = velocity × time)

$$d_{\mathrm{F}} = \bar{v}_x t \tag{21.49}$$

So, for our example, the fronts would travel 50 m in 50 years, and 200 m in 200 years (Fig. 21.14(A)). Furthermore, the total flux of fluid that has passed across the boundary at $x = 0$ at time t can be calculated for this case by taking into account the porosity

$$q_{\mathrm{TI}} = \bar{v}_x t \phi \tag{21.50}$$

Here q_{TI} is the *time-integrated fluid flux* ($\mathrm{m}^3_{\mathrm{(fluid)}}\ \mathrm{m}^{-2}_{\mathrm{(rock)}}$), about which we will have much more to say below.

The shape of the solute concentration profiles changes when hydrodynamic dispersion operates. We consider a diffusion coefficient of 10^{-8} $\mathrm{m}^2\ \mathrm{s}^{-1}$, a tortuosity factor of 0.5, and a coefficient of longitudinal dispersivity of 5 m (Eq. (21.12)), together with the advective velocity of 1 m a^{-1} used above. The shape of the solute concentration profile broadens noticeably when hydrodynamic dispersion is present (Fig. 21.14(B)). Note as well that the amount of broadening increases with front propagation distance. The simple expression for d given above is still useful for estimating the general length scale of solute infiltration, although the front broadening means that some solute will be transported farther than for the case of advection alone.

Here we have considered trace element transport, but Eq. (21.46) is appropriate for major fluid constituents as well. However, treatment of the full fluid chemistry is more complicated as one needs to solve coupled sets of equations of the form of (21.46) involving the various species *s*; this task usually requires numerical methods.

The infiltration behavior will be quite different if chemical reactions between fluid and rock occur. An additional term must be incorporated into Eq. (21.46) that accounts for how the fluid composition changes due to liberation or consumption of chemical species by reaction. Thus, the general mass conservation expression in one space dimension incorporating the effects of chemical reaction can be written as

$$\frac{\partial(\phi c_s)}{\partial t} = -\frac{\partial(\bar{v}_x \phi c_s)}{\partial x} + \frac{\partial\left(\phi D_{\text{HD}} \dfrac{\partial c_s}{\partial x}\right)}{\partial x} + \phi \sum_l R_{s,l} \quad (21.51)$$

where $R_{s,l}$ is the rate at which reaction *l* produces (positive) or consumes (negative) species *s* (mole $\text{m}^{-3}{}_{\text{(fluid)}}$ time^{-1}). To obtain the total amount of *s* produced or consumed, one must sum over all reactions *l* occurring in the rock mass. In nature, the reaction rates will not be instantaneous, so departures from local fluid–rock equilibrium may occur. Nonetheless, if we assume for the moment that rates are fast enough to maintain equilibrium, then an analytically tractable advection–dispersion–reaction equation can be derived (Fletcher and Hofmann, 1974). This treatment would be appropriate if, for example, rates of mineral crystallization and/or intracrystalline diffusion were fast relative to advection and hydrodynamic dispersion.

To begin, we must specify a relationship between the concentration of the tracer in the fluid and the solid at equilibrium. This is commonly done using a simple "isotherm" relationship

$$c_s = K_v c_{s,\text{Sol}} \quad (21.52)$$

where c_s and $c_{s,\text{Sol}}$ are the concentrations of species *s* in the fluid and solid, respectively, and K_v is the equilibrium fluid/solid *partition coefficient* by volume (e.g. $[\text{mole m}^{-3}]_{\text{fluid}}/[\text{mole m}^{-3}]_{\text{solid}}$). For example, if the concentration of *s* in the fluid increases, then at equilibrium it must also increase in the solid according to Eq. (21.52). Note that the increase in concentration in the solid requires that *s* be consumed from the fluid; the mass needed to satisfy Eq. (21.52) is obtained by advection and/or hydrodynamic dispersion through the system. Normally, the partition coefficient will vary with temperature, pressure, and fluid composition, but we will assume that it is constant here. The partition coefficient is also commonly expressed as the solid/fluid ratio by mass, K_d, such that $K_d = \rho_f/(\rho_{\text{Sol}}K_v)$; here, ρ_f and ρ_{Sol} are the fluid and solid densities.

If $\bar{v}_x$, D_{HD}, and ϕ don't vary, and if the chemical exchange is treated as one overall reaction ($l = 1$), then Eq. (21.51) can be rewritten as

$$\phi\frac{\partial c_s}{\partial t} - \phi R_s = -\bar{v}_x \phi \frac{\partial c_s}{\partial x} + \phi D_{\text{HD}} \frac{\partial^2 c_s}{\partial x^2} \quad (21.53)$$

In order to use the local equilibrium relationship of (21.52), we need to express R_s in terms of the rate of reaction of the solid

$$-\phi R_s = (1-\phi) R_s^{\text{Sol}} \quad (21.54)$$

where R_s^{Sol} is the solid reaction rate (e.g. moles $\text{m}^{-3}{}_{\text{(solid)}}$ time^{-1}), and the term in parentheses gives the $\text{m}^3{}_{\text{(solid)}}$ $\text{m}^{-3}{}_{\text{(rock)}}$. The signs are opposite because *s* liberated to the fluid by reaction is lost from the solid, and vice versa. The right-hand side of (21.54) can also be expressed as

$$(1-\phi) R_s^{\text{Sol}} = (1-\phi)\frac{\partial c_{s,\text{Sol}}}{\partial t} \quad (21.55)$$

where it is understood that the time derivative pertains to changes in solid composition due to reaction. Assuming local equilibrium and constant K_v (Eq. (21.52)), the right-hand side of Eq. (21.55) can be rewritten

$$(1-\phi)\frac{\partial c_{s,\text{Sol}}}{\partial t} = \frac{(1-\phi)}{K_v}\frac{\partial c_s}{\partial t} \quad (21.56)$$

Substitution back into Eq. (21.53), remembering the change in sign (Eq. 21.54), yields

$$\phi\frac{\partial c_s}{\partial t} + \frac{(1-\phi)}{K_v}\frac{\partial c_s}{\partial t} = -\bar{v}_x \phi \frac{\partial c_s}{\partial x} + \phi D_{\text{HD}} \frac{\partial^2 c_s}{\partial x^2}$$

Collecting terms and rearranging gives

$$\frac{\partial c_s}{\partial t}\left(\phi + \frac{(1-\phi)}{K_v}\right) = -\bar{v}_x \phi \frac{\partial c_s}{\partial x} + D_{\text{HD}} \phi \frac{\partial^2 c_s}{\partial x^2} \quad (21.57)$$

Finally, for small porosity, the term in parentheses on the left-hand side is approximately $1/K_v$, so we can write

$$\frac{\partial c_s}{\partial t} \approx -K_v \bar{v}_x \phi \frac{\partial c_s}{\partial x} + K_v D_{\text{HD}} \phi \frac{\partial^2 c_s}{\partial x^2} \quad (21.58)$$

Conveniently, Eq. (21.58) and, with slight rearrangement, Eq. (21.57) have the same form as Eq. (21.47), so Eq. (21.48) can be used to solve them. For example, for (21.58), one substitutes $K_v \bar{v}_x \phi$ for $\bar{v}_x$ and $K_v D_{\text{HD}} \phi$ for D_{HD}. The result gives the composition of the fluid, but the composition of the solid is easily computed from (21.52). Many common tracers partition more strongly into the solid and thus have K_v less than one, and ϕ will always be less than one for flow through a porous rock matrix. In such cases, $K_v \bar{v}_x \phi$ will be smaller than $\bar{v}_x$ and $K_v D_{\text{HD}} \phi$ will be smaller than D_{HD}. Therefore, the rate of solute front propagation given by Eq. (21.58) will be slower than for nonreactive flow (Eq. (21.47)), as we will see below.

It is instructive to revisit the model case of Sr infiltration from metapelite to metasandstone (Fig. 21.14). In the earlier example we assumed that the metasandstone layer was composed of nonreactive quartz, but now consider metasandstone that contains a large amount of plagioclase feldspar, as well

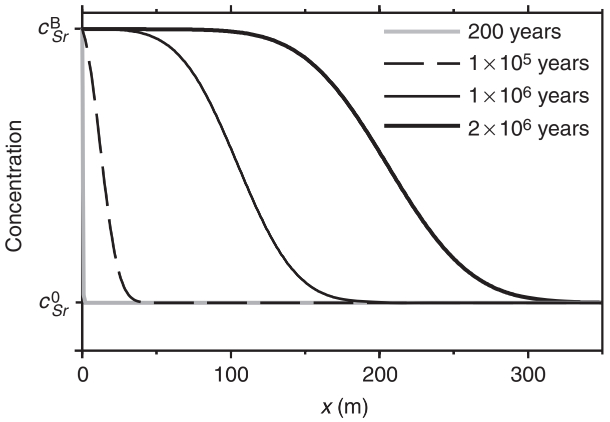


Fig. 21.15 Concentration profiles in fluid for advection–hydrodynamic dispersion reaction computed using Eq. (21.48) for model infiltration problem described in text. c^B_{Sr} is the input concentration of Sr in the fluid at the model lithologic contact (boundary at x=0), and c^0_{Sr} is the initial concentration in the flow region ($x > 0$). Concentration profiles for the solid would have the same shapes, but the absolute concentration values would be different (Eq. (21.52)).

as quartz. Let us further stipulate that the Sr content of the feldspar in the metasandstone is significantly less than that in the metapelite. Therefore, as the fluid from the metapelite enters the metasandstone, the plagioclase in the metasandstone will react to become more Sr rich according to Eq. (21.52). The K_v for Sr can vary widely (orders of magnitude) depending on such factors as the Cl content of the fluid; we use 0.1 for illustration purposes here but caution that no single value of K_v is applicable to all cases.

Results for reactive flow shown in Figure 21.15 were obtained using the values of fluid velocity and hydrodynamic dispersion from Figure 21.14 described above. For the reactive case, porosity must also be specified and was set to 10^{-3}. For K_v = 0.1, the solute front propagates much more slowly than for nonreactive flow, because the rock is now consuming Sr along the flow path. The difference in front position between the nonreactive and reactive cases is considerable. In the nonreactive case the front propagates ~200 m in 200 years (Fig. 21.14(B)), whereas in the reactive case it takes 2 million years to move the same distance! (Fig. 21.15). In Eq. (21.49) we saw that the distance of hydrodynamic front propagation was simply equal to the pore velocity multiplied by time of flow. Equation (21.58) shows that the effective "velocity" of solute front migration is reduced by the factor $K_v\phi$. Thus, the average distance of solute front migration during reactive flow is

$$d = K_v \bar{v}_x \phi t \tag{21.59}$$

But $\bar{v}_x \phi t$ is also equal to the time-integrated fluid flux q_{TI} (Eq. (21.50)). Thus, the distance of solute front migration, which is also equal to the length scale of rock alteration, d, is directly related to the time-integrated fluid flux (Bickle, 1992)

$$q_{\mathrm{TI}} = \frac{d}{K_v} \tag{21.60}$$

When studying metamorphic rocks in the field, the fluid has migrated away (except for fluid inclusions), but the distance d can be measured from the chemically or isotopically altered rock left behind. Such fronts preserved in rocks are commonly termed *geochemical fronts*. Equation (21.60) is very powerful. If the distance d of chemical alteration in a rock mass can be measured and if K_v is known, then the time-integrated fluid flux can be estimated.

The value of $1/K_v$ in Eq. (21.60) has units of $\mathrm{m^3}_{(\mathrm{fluid})}$ $\mathrm{m^{-3}}_{(\mathrm{rock})}$ and is equivalent to the classical *fluid to rock ratio* or *water to rock ratio* that we encountered first in Chapter 20. This ratio does not express the total amount of fluid that has passed through a rock, but rather only that amount which has caused the change in composition due to reaction. Consequently, fluid:rock ratios underestimate the total fluid flux in nearly all cases. However, the fluid:rock ratio is a useful index of the general extent of alteration, and it can be calculated without specific knowledge of the transport processes (advection, diffusion, or mechanical dispersion).

Equations like (21.48) and (21.60) are applicable to isotopic systems if the change in ratio during alteration is small (e.g. ^{87}Sr/^{86}Sr) or if the concentration of the tracer isotope is small (e.g. ^{18}O). For these cases, one could use the K_v for Sr or O to model the isotopic transport (Lassey and Blattner, 1988). Of course, for Sr isotopes, the decay of ^{87}Rb to ^{87}Sr would also need to be accounted for (Baxter and DePaolo, 2002a, b). The K_v for oxygen in typical fluid–rock systems is ~0.6 and is normally larger than that for Sr (Bickle, 1992). Thus, under local equilibrium conditions, oxygen isotopic fronts will propagate farther than Sr isotopic fronts for a given q_{TI} (Eq. (21.60)). In general, different tracers will have different characteristic transport distances and, therefore, measurements of multiple tracers can place strong constraints on time-integrated fluid fluxes and flow processes.

Bickle *et al.* (1997) investigated isotopic transport across metapelite–metacarbonate contacts in greenschist facies rocks of the Waterville Formation, Maine. Carbonate minerals and quartz in limestones and dolostones normally have heavier oxygen isotope signatures than carbonates and quartz in shales, providing an isotopic difference that can be propagated across lithologic contacts by fluid infiltration during metamorphism. Bickle *et al.* (1997) found that the advective geochemical front displacement from metapelite into metacarbonate across one contact was about 2 m. A time-integrated fluid flux of ~3.3 $\mathrm{m^3}$ $\mathrm{m^{-2}}$ is estimated using a K_v of 0.6 for oxygen and Eq. (21.60). This means that about 3.3 $\mathrm{m^3}$ of fluid flowed across each square meter of rock during fluid–rock interaction. The estimated value is rather small, and is much smaller than the volumes expected for devolatilization of thick metasedimentary sequences (~10^3 $\mathrm{m^3}$ $\mathrm{m^{-2}}$; Problem 16.2). Therefore, much of the flow was probably parallel to the layering, rather than across it. In addition, the hydrodynamic dispersion distance can be estimated from the broadening of the front. It is substantial, being over 6 m for this case (Bickle *et al.*, 1997), indicating that a large amount of diffusion and/or mechanical dispersion took place across the contact.

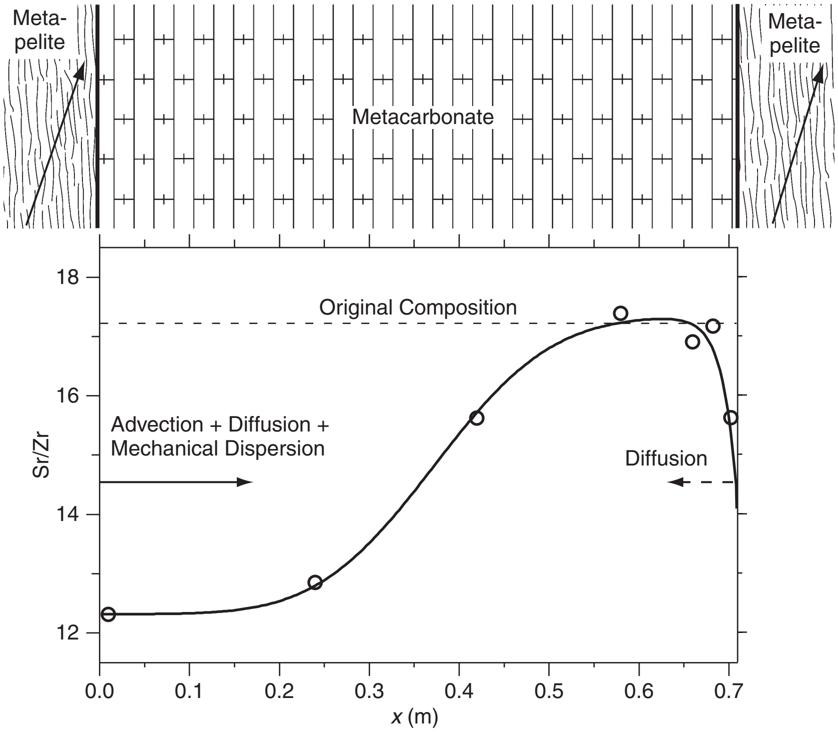


Fig. 21.16 Sr concentration profile across metacarbonate layer from south-central Connecticut (Ague, 2003a). Solid line is best-fit advection–hydrodynamic dispersion-reaction solution obtained from Eq. (21.48). The shape of the profile indicates a component of flow and associated dispersion entered the metacarbonate layer at the left contact ($x = 0$). The overall time-integrated fluid flux across the contact is small, however, suggesting that most of the flow was actually layer parallel. This is shown schematically by the large arrows, which indicate overall flow direction in the upper cartoon. The shape of the profile on the right indicates diffusion from the metapelite into the metacarbonate across the contact at $x = 72$ cm. Sr concentrations normalized to Zr concentrations to provide geochemical reference frame, although in this case plotting Sr alone gives a similar profile shape (see Section 21.13).

For another example, we can investigate the Sr elemental profile across a greenschist facies (chlorite zone) metacarbonate layer in contact with metapelitic rocks in the Wepawaug Schist, Connecticut (Fig. 21.16; Ague, 2003a). Fluids equilibrated with the metacarbonate would have had higher Sr contents than those in the metapelite, owing to the relatively high Sr concentrations in calcite. So reactive fluid infiltration from metapelite to metacarbonate acted to lower metacarbonate Sr contents. Bickle *et al.* (1997) estimated fluid Sr contents in the range 75 to 400 ppm for similar metasedimentary rocks in Maine. For a typical whole-rock Sr content of 1000 ppm (Ague, 2003a), the corresponding values of K_v are 0.025 to 0.13. The reader should be cautioned that these values are subject to rather large uncertainties.

Inspection of Figure 21.16 indicates that the displacement of the geochemical front is about 0.4 m from the contact at $x = 0$ m, so the estimated time-integrated fluid fluxes across the layering are in the range of 3 to 16 $\mathrm{m^3\ m^{-2}}$ with the above K_v values (see Eq. (21.60)). Once again, these values are rather small, suggesting that most of the metamorphic fluid flow was parallel to the layering. Although small, these fluxes predict that the length scale for oxygen isotope alteration due to advection will be at least 1.8 m, larger than the thickness of the layer ($K_v = 0.6$; Eq. (21.60)). So oxygen isotopes in the metacarbonate would be expected to be largely equilibrated with the surrounding metapelite, as observed for similar thin layers in the Wepawaug Schist by Palin (1992).

Note that the Sr concentrations drop at the lithologic contact on the right side of Figure 21.16. This drop reflects diffusion from the metapelite into the metacarbonate. The affected zone is relatively thin because the diffusion is operating in a direction opposite to that of the cross-layer fluid flow and is thus working "against" the advection (Fig. 21.16). The geometry of the Sr profile highlights an important point, namely, that the asymmetry of geochemical fronts reveals fluid flow directions (e.g. Bickle, 1992).

21.10 FLUID FLUXES ALONG GRADIENTS IN TEMPERATURE AND PRESSURE

Our discussion thus far has focused mainly on the transport of geochemical fronts across contacts under isothermal, isobaric conditions, and has not considered the effects of temperature and pressure. The solubilities of minerals, however, are dependent on temperature and pressure. Consequently, metasomatic mass transfer can occur if reactive fluids flow through rocks along gradients in temperature and pressure. For example, the solubility of quartz decreases with both decreasing temperature and pressure (Fig. 21.6(A)), such that fluids in high-grade, quartz-bearing metamorphic rocks will contain more dissolved silica than fluids in low-grade rocks. A fluid flowing upward toward the surface will normally be cooling and decompressing. If the ascending fluid reacts with surrounding rocks that contain quartz, the amount of silica dissolved in solution will progressively decrease along the flow path. The concentration is decreasing because silica is continually removed from the fluid as quartz precipitates in the rock. In contrast, fluid flow in a direction of increasing temperature and pressure will tend to dissolve quartz from the rock, and increase the silica content of the fluid.

Quartz veins are zones where silica has been precipitated in fractures. The amount of fluid required to precipitate the quartz can be estimated if the precipitation was caused by flow in a direction of decreasing temperature and pressure (Yardley, 1986a). Ferry and Dipple (1991) present a method for quantification of the time-integrated fluid fluxes based on the reaction-transport equation (21.51). To derive the appropriate expression, it is convenient to recast Eq. (21.53) as

$$\phi\frac{\partial c_s}{\partial t}+(1-\phi)\frac{\partial c_{s,\mathrm{Sol}}}{\partial t}=-\bar{v}_{x'}\phi\frac{\partial c_s}{\partial x'}+\phi D_{\mathrm{HD}}\frac{\partial^2 c_s}{\partial x'^2}\quad(21.61)$$

where x' denotes in a general way the direction of fluid motion which can be parallel to x, y, z, or any other direction. In writing (21.61) we have made use of the relationships in (21.54) and (21.55). Note that the second term on the left-hand side has units of moles of s per unit volume rock (solid + fluid) per unit time so, assuming constant porosity, we will represent it here as

$$(1-\phi)\frac{\partial c_{s,\mathrm{Sol}}}{\partial t}=\frac{\partial c_{s,\mathrm{Rock}}}{\partial t}$$

where $c_{s,\mathrm{Rock}}$ is now the concentration of species s per unit volume rock. If the change in the concentration in the fluid with time is negligible, and the effects of advection dominate those of diffusion along regional flow paths, then (21.61) simplifies to

$$\frac{\partial c_{s,\mathrm{Rock}}}{\partial t}=-\bar{v}_{x'}\phi\frac{\partial c_s}{\partial x'}\quad(21.62)$$

This equation is for *steady state* with respect to fluid composition. The fluid composition doesn't change with time even though the solid composition does (it is a good exercise to think about how this is possible). Rearrangement yields

$$\frac{\bar{v}_{x'}\phi}{\left(\dfrac{\partial c_{s,\mathrm{Rock}}}{\partial t}\right)}=-\frac{1}{\left(\dfrac{\partial c_s}{\partial x'}\right)}\quad(21.63)$$

If the concentration gradient $(\partial c_s/\partial x')$ remains constant, this expression can be readily integrated over the total time interval of flow and reaction

$$\bar{v}_{x'}\phi\int_0^t \mathrm{d}t=-\frac{1}{\left(\dfrac{\partial c_s}{\partial x'}\right)}\int_{c^0_{s,\mathrm{Rock}}}^{c^t_{s,\mathrm{Rock}}}\mathrm{d}c\quad(21.64)$$

where $c^0_{s,\mathrm{Rock}}$ and $c^t_{s,\mathrm{Rock}}$ are the concentrations of s in the rock at the beginning and end of reaction progress, respectively. The integrated expression is

$$\bar{v}_{x'}\phi t=-\frac{c^t_{s,\mathrm{Rock}}-c^0_{s,\mathrm{Rock}}}{\left(\dfrac{\partial c_s}{\partial x'}\right)}\quad(21.65)$$

The left-hand side is the time-integrated fluid flux (q_{TI}). The numerator on the right-hand side gives the change in moles in the solid due to reaction; we will refer to it as $m_{s,\mathrm{Rock}}$ (e.g. moles m^{-3}). The denominator is the concentration gradient in the direction of flow. If the concentration gradient is due only to precipitation or dissolution caused by flow along T or P gradients, the denominator can be expanded using the chain rule (Baumgartner and Ferry, 1991)

$$\frac{\partial c_s}{\partial x'}=\left(\frac{\partial c_s}{\partial T}\right)_P\frac{\partial T}{\partial x'}+\left(\frac{\partial c_s}{\partial P}\right)_T\frac{\partial P}{\partial x'}\quad(21.66)$$

Thus, making the appropriate substitutions, (21.65) can be rewritten as

$$q_{\mathrm{TI}}=-\frac{m_{s,\mathrm{Rock}}}{\left(\dfrac{\partial c_s}{\partial T}\right)_P\dfrac{\partial T}{\partial x'}+\left(\dfrac{\partial c_s}{\partial P}\right)_T\dfrac{\partial P}{\partial x'}}\quad(21.67)$$

where q_{TI} is the time-integrated fluid flux in m$^3{}_{\mathrm{(fluid)}}$ m$^{-2}{}_{\mathrm{(rock)}}$.

Equation (21.67) looks rather formidable, but it is actually straightforward to apply. Let us model precipitation of a pure quartz vein due to the cooling and decompression of an ascending fluid. The coordinate axis is parallel to z (vertical), and it increases upward. $m_{s,\mathrm{Rock}}$ is the moles of SiO$_2$ precipitated per unit volume, and is equal to the inverse of the molar volume of quartz (4.41×10^4 mole m^{-3}; Table 7.1). If local fluid–rock equilibrium is maintained, then the derivatives of concentration with respect to temperature and pressure can be evaluated using solubility expressions for quartz (e.g. Manning, 1994). For example, consider amphibolite facies vein formation at 600 ºC and 0.8 GPa. To obtain the T derivative, one can calculate the concentration of dissolved aqueous silica over a T range, say 598 to 602 ºC. Then, the derivative can be estimated using a finite difference

$$\frac{\partial c_{\mathrm{SiO_2,aq}}}{\partial T}\approx\frac{c^{602}_{\mathrm{SiO_2,aq}}-c^{598}_{\mathrm{SiO_2,aq}}}{602-598}\quad(21.68)$$

Analogous estimates can be made for the P derivative using a narrow range of P values, such as 0.79 to 0.81 GPa. For this example, the derivatives are small, about 1.9 mole m^{-3} K^{-1}, and 3×10^{-7} mole m^{-3} Pa^{-1}. The T gradient term can be estimated using a crustal geotherm: –25 ºC km^{-1} = -0.025 K m^{-1}. Assuming a lithostatic P gradient gives -2.8×10^4 Pa m^{-1}. The minus signs on these gradients simply indicate that temperature and pressure are decreasing upward in the direction of flow along z.

With the above values, the time-integrated fluid flux can be estimated using Eq. (21.67)

$$q_{\mathrm{TI}}=-\frac{4.41\times10^4}{(1.9)(-0.025)+(3\times10^{-7})(-2.8\times10^4)}=7.9\times10^5\ \mathrm{m^3\ m^{-2}}\quad(21.69)$$

The flux value of ~8×10^5 m^3 m^{-2} is immense. It implies that a column of water nearly 10^6 m (1000 km) long passed through each square meter of rock now occupied by quartz vein! Similar large fluxes would also be needed to dissolve quartz from rocks by the flow of fluids in a direction of increasing T and P (Feehan and Brandon, 1999). The main reason the fluxes are so large is that aqueous silica concentrations change little over small ranges of T and P, leading to small $\partial c_{\mathrm{SiO_2,aq}}/\partial T$ and $\partial c_{\mathrm{SiO_2,aq}}/\partial P$ terms in the denominator of Eq. (21.67). The necessary fluxes would be somewhat smaller if the gradients in T and P had larger magnitudes in the direction of flow, but there are limits on how large such gradients can be in nature. For example, *T* gradients in excess of ±100 ºC km^{-1} are unlikely except in some contact aureoles.

Do the results of Equation (21.69) mean that all quartz veins are conduits for enormous volumes of fluid? The short answer is no. For example, the flux expression neglects other transport processes like diffusion. When a fracture opens, the pressure is somewhat lower in the fracture than in the adjacent wall rocks. The concentration of silica in the wall rock fluids will thus be larger than in the fracture, so local concentration gradients can drive silica to the cracks by diffusion. Quartz precipitation then occurs, acting to seal the cracks (e.g. Yardley, 1975, 1986a; Ramsay, 1980). For such locally derived veins, fluid fluxes need not be large – in fact, the diffusive transport could occur through an essentially static pore fluid. Furthermore, *T* and *P* gradients are not the only things that control the aqueous silica content of fluids. Silica solubility decreases significantly as water activity decreases (Fig. 21.6(B)). Decreases in water activity due to, for example, increased fluid CO_2 content near marbles could dramatically lower silica solubility, increase concentration gradients, and thus produce quartz veins with a much lower flux than predicted by Eq. (21.69). Finally, if sluggish reaction rates slow quartz precipitation or dissolution significantly, then flux estimates cannot be made using the assumption of local equilibrium (Bolton *et al*., 1999). Despite these considerations, Eq. (21.69) shows that very large time-integrated fluxes are required if quartz is to be precipitated or dissolved by aqueous fluids flowing along gradients in *T* and *P* through quartz-bearing rocks.

In addition to silica, Eq. (21.61) can be used to estimate the fluid fluxes required to cause metasomatic changes for other elements due to flow along gradients in *T* and *P*. One example is the K–Na exchange reaction

$$\mathrm{NaCl}^0_{\mathrm{aq}} + \text{K-feldspar} = \mathrm{KCl}^0_{\mathrm{aq}} + \text{albite} \tag{21.70}$$

where $\mathrm{NaCl}^0_{\mathrm{aq}}$ and $\mathrm{KCl}^0_{\mathrm{aq}}$ are neutral aqueous chloride complexes in the fluid. In nature, fluids will contain a host of other dissolved species as well, such as $\mathrm{Na}^+_{\mathrm{aq}}$ and $\mathrm{K}^+_{\mathrm{aq}}$ ions. For a wide array of quartz and feldspar bearing crustal rocks, the total amount of Na dissolved in aqueous solution decreases, whereas the total K increases, with increasing temperature (Fig. 21.7). This means that fluids flowing in a direction of increasing temperature (up-*T*) will produce sodic phases like albite and destroy K-rich phases like muscovite or K-feldspar, thereby increasing the Na/K of the rock (Orville, 1962). Fluids moving down-*T* will do the opposite. The amount of dissolved material, as well as the temperature derivatives of concentration, increase in magnitude with increasing Cl content (Fig. 21.7).

The presence of Na and/or K metasomatism has been documented in a number of geologic environments. For example, hydrothermally metamorphosed basalts on the seafloor ("spillites") altered by fluid flow up-*T* toward mid-ocean ridges are commonly enriched in albite. Convective systems around cooling plutons often have *propylitic* alteration characterized by abundant albitic plagioclase on their flanks where fluids flow up-*T* inward toward the intrusion, and *potassic* alteration (muscovite, K-feldspar) in their cores, where fluids flow down-*T* on the upwelling limbs of convection cells (Brimhall, 1977). Dipple and Ferry (1992a) document large amounts of down-*T* fluid flow in metamorphic ductile shear zones leading to K addition and the growth of micas. The micas are relatively soft and deform easily, so the metasomatism may in fact exert major controls on fault strength.

The advective reaction–transport equation written for Na (or K) is analogous to Eq. (21.67)

$$q_{\mathrm{TI}} = -\frac{m_{\mathrm{Na,Rock}}}{\left(\dfrac{\partial c_{\mathrm{Na,tot}}}{\partial T}\right)_P \dfrac{\partial T}{\partial x'} + \left(\dfrac{\partial c_{\mathrm{Na,tot}}}{\partial P}\right)_T \dfrac{\partial P}{\partial x'}} \tag{21.71}$$

where $m_{\mathrm{Na,Rock}}$ is the total number of moles of Na added to (positive) or lost from (negative) the rock per unit volume rock, and $c_{\mathrm{Na,tot}}$ is the total amount of Na dissolved in the fluid. A significant difference between this expression and that for quartz veins is that silica solubility gradients are uniquely determined at a given *P* and *T* in water equilibrated with quartz, but the gradients in Na or K concentration are dependent on Cl concentration (Fig. 21.7) as well as the concentrations of other elements in the fluid. In general, more Cl-rich fluids will produce more extensive metasomatism for a given flux.

One way to approach Eq. (21.71) is to determine independently the key compositional variables that control the fluid speciation, like fluid Cl content. For example, mica Cl contents can be used to quantify fluid Cl concentrations. Furthermore, fluid inclusions may provide direct samples of metamorphic fluids (e.g. Vityk and Bodnar, 1995; Whitney *et al*., 1996). Calculations can then be done to estimate changes in fluid composition over narrow ranges of *T* and *P* for fluids coexisting with a given mineral assemblage, so that the derivatives of concentration with respect to *T* and *P* can be estimated using finite differences, similar to the silica example discussed above. Mass balance analysis (Section 21.13) tells us the amount of Na or K gained or lost by the rock, quantities needed to evaluate $m_{\mathrm{Na,Rock}}$ or $m_{\mathrm{K,Rock}}$. Finally, given estimates of the *T* and *P* gradients along the flow path, Eq. (21.71) can be solved for the time-integrated flux. Note that equations like (21.71) can be written for other elements too, including K, Ca, Mg, and Fe.

Dipple and Ferry (1992a) estimated that time-integrated fluid fluxes causing major element metasomatism in ductile shear zones were around 2×10^4 m^3 m^{-2}, and reflected down-*T* flow in four of the five cases they examined. Ague (1997) estimated similar flux magnitudes for flow in fractures in the Barrovian sequence in northeastern Scotland. The fluxes required for major element metasomatism seem to average about 1–2 orders of magnitude less than those required to make quartz veins by advection, but are still very large. Note, however, that these estimates require that advection along temperature and/or pressure gradients is the dominant mass transfer reaction mechanism. If diffusion were important, or if flow occurred across contacts between rock types with very different fluid compositions, then the flux estimates will be in error (generally they will be too large). To test assumptions, it is a good idea to evaluate expressions like (21.71) for a range of elements to check if all elements provide similar flux estimates. For cases where Eq. (21.71) is applicable it is extremely valuable, because it constrains both the time-integrated fluid flux and the direction of fluid motion with respect to regional *T* and *P* gradients.

Mass transfer along gradients in *T* and *P* can also be critical for driving reactions among minerals and H_2O–CO_2 fluids. Here we focus on H_2O and CO_2, neglecting other compositional changes. The appropriate expression for steady-state reaction and advective mass transfer is similar to Eq. (21.62)

$$\frac{\partial c_{s,\mathrm{Rock}}}{\partial t} = -\frac{\partial(\bar{v}_{x'}\phi c_s)}{\partial x'} = -\bar{v}_{x'}\phi\frac{\partial c_s}{\partial x'} - c_s\frac{\partial(\bar{v}_{x'}\phi)}{\partial x'} \qquad (21.72)$$

The difference is that because fluid is produced or consumed along the flow path, the gradient in the flux (rightmost term in Eq. (21.72)) is nonzero. Because phase equilibria among minerals and H_2O–CO_2 fluids are commonly depicted on T–X_{CO_2} diagrams, it is convenient to recast (21.72) in terms of mole fractions ($X_s = c_s\bar{V}$). Baumgartner and Ferry (1991) show that the resulting expression for the time-integrated fluid flux is

$$q_{\mathrm{TI}} = \frac{\bar{V}(m_s - X_s[m_{\mathrm{CO_2}} + m_{\mathrm{H_2O}}])}{\left(\frac{\partial X_s}{\partial T}\right)_P \frac{\partial T}{\partial x'} + \left(\frac{\partial X_s}{\partial P}\right)_T \frac{\partial P}{\partial x'}} \qquad (21.73)$$

where $\bar{V}$ is the molar volume of the fluid, the subscript *s* denotes either H_2O or CO_2, m_s is the number of moles of *s* added to (positive) or removed from (negative) the fluid, and $m_{\mathrm{CO_2}}$ and $m_{\mathrm{H_2O}}$ are the amounts of CO_2 and H_2O added to (positive) or removed from (negative) the fluid. Note that the m_s values are related to the reaction progress ξ (Chapter 20) by $m_s = n_s\xi$, where n_s is the stoichiometric coefficient for species *s* in the reaction.

For example, consider flux estimation for the following reaction at amphibolite facies conditions of 550 ºC and 0.78 GPa: 5 dolomite + 8 quartz + H_2O = 3 calcite + tremolite + 7 CO_2. This reaction proceeded at or near these conditions in many metacarbonate rocks in New England (Ferry, 1992; Ague and Rye, 1999; Ague, 2002). Let's say we compared a low-grade rock containing no tremolite to a reacted rock in which tremolite was produced. We find that the tremolite-bearing rock lost on average 7.04×10^2 mole m^{-3} CO_2 and gained 1.01×10^2 mole m^{-3} H_2O relative to its low-grade counterpart. Thus, the fluid *gained* 7.04×10^2 mole m^{-3} CO_2 and *lost* 1.01×10^2 mole m^{-3} H_2O. At 550 ºC and 0.78 GPa, the equilibrium $X_{\mathrm{CO_2}}$ is 0.12, and $\partial X_{\mathrm{CO_2}}/\partial T$ is 2.77×10^{-3} K^{-1}. Note that $\partial X_{\mathrm{CO_2}}/\partial T$ is positive for most reactions involving CO_2 under water-rich conditions at low $X_{\mathrm{CO_2}}$ (Fig. 20.8). The *P* gradient term is often small and is ignored here. The equilibrium fluid molar volume can be computed using the expressions of Kerrick and Jacobs (1981) and is 2.046×10^{-5} m^3 mole^{-1}. Using CO_2 as the "*s*" species, Eq. (21.73) evaluates to

$$q_{\mathrm{TI}} = \frac{2.046\times10^{-5}\left(7.04\times10^2 - 0.12\left[7.04\times10^2 - 1.01\times10^2\right]\right)}{2.77\times10^{-3}\left(\frac{\partial T}{\partial x'}\right)} \qquad (21.74)$$

or

$$q_{\mathrm{TI}} = \frac{1.29\times10^{-2}}{2.77\times10^{-3}\left(\frac{\partial T}{\partial x'}\right)} \qquad (21.75)$$

Because the coordinate framework is defined so that x' is positive in the direction of flow, the value of q_{TI} must be positive as well. As a consequence, the temperature gradient in the direction of flow for this example must also be positive. In other words, the reaction would have to be driven by flow of fluid in a direction of increasing temperature. For a gradient of 25 ºC km^{-1}, the q_{TI} estimate is 186 m^3 m^{-2}. The fluid would get more CO_2-rich along the flow path, as the reaction produces CO_2 and consumes H_2O.

Metacarbonate rocks lost substantial CO_2 during metamorphism in New England (Ferry, 1992; Ague, 2003a), and CO_2 loss is a general consequence of prograde heating. Local equilibrium fluid flow along gradients in *T* and *P* is one way to drive CO_2 loss, but infiltration of more water-rich fluids from external sources is another (Sections 20.5, 20.6, and 21.12). Thus Eq. (21.73) is best used to estimate q_{TI} values in settings dominated by thick metacarbonate sequences, far removed from external sources of water such as dehydrating metapelitic rocks or degassing magmas. If such external fluids are introduced by diffusion or advection, then Eq. (21.73) will tend to overestimate q_{TI} values and the inferred flow directions may be incorrect (Ague and Rye, 1999).

Fluid flow along temperature gradients will also change the stable isotopic composition of rocks and fluids. These isotopic changes can be used to estimate time-integrated fluid fluxes (e.g. Dipple and Ferry, 1992b; Bowman *et al.*, 1994; Cui *et al.*, 2001).

21.11 REACTION KINETICS

Our discussion of mass transport and reaction has been based on the assumption that fluid and rock are always in local chemical or isotopic equilibrium. However, equilibrium is not always maintained. For example, expressions like Eq. (21.58) require that minerals be homogeneous, but trace element or isotopic zonation is not uncommon in nature (Kohn, 2003). Refractory minerals, like the Al_2SiO_5 polymorphs and garnet, can have sluggish rates of nucleation and growth. In some cases, the rates of reaction may be so slow that they are unable to keep pace with changing temperature, pressure, or fluid composition (e.g. Baxter and DePaolo, 2002a, b; Carlson, 2002; Lüttge *et al.*, 2004; Müller *et al.*, 2004). In this section, we will examine some of the consequences of reaction rates that depart from local equilibrium conditions.

The rate at which a metamorphic reaction progresses depends on both thermodynamic and kinetic factors. The thermodynamic potential must be large enough to overcome kinetic factors – this involves *overstepping* the equilibrium temperature, pressure, and/or fluid composition conditions by some finite amount. Kinetic factors are particularly sensitive to temperature, but the presence and composition of fluids can be equally important. Strain can also affect the rate of a reaction.

As we saw in Section 21.7, reactions between fluids and minerals involve coupled processes, including dissolution of reactant mineral surfaces, transport of nutrients away from dissolving minerals to the surfaces of product minerals, and precipitation on the surfaces of product minerals. The rates of these processes are generally too complex to treat individually, so the rate of the overall reaction is normally cast in terms of a *rate-limiting step* – the slowest step in the overall reaction that transforms reactants to products (e.g. Berner, 1980). If the rate of reaction is slow relative to the rate of transport to and from precipitating and dissolving minerals, the reaction is said to be *surface controlled*. The opposite case of fast reaction relative to transport is called *transport controlled* or sometimes *diffusion controlled*. Because pore sizes are likely to be small and diffusion rates through fluid are relatively rapid, surface-controlled kinetics may dominate in many rocks, although both kinds of kinetics are known (e.g. Sanchez-Navas, 1999; Lüttge *et al*., 2004), and more complex reaction mechanisms are possible (Dohmen and Chakraborty, 2003).

The driving force for a metamorphic reaction is provided by the lowering of free energy of the reaction. For a reaction to occur spontaneously, ΔG must be negative. With changing temperature and pressure, the change in ΔG is given by (Eqs. (7.33) and (7.34))

$$\left(\frac{\partial \Delta G}{\partial T}\right)_P = -\Delta S$$

and

$$\left(\frac{\partial \Delta G}{\partial P}\right)_T = \Delta V$$

Over the range of temperature that is likely to be involved in the overstepping of a reaction, ΔS and ΔV of reaction will be approximately constant; that is, the change in ΔG with change in T and P is roughly linear. Thus the rate at which ΔG becomes increasingly negative with overstepping of a reaction should be about the same in the prograde and retrograde directions. Changes in temperature and pressure are therefore likely to provide equal driving potentials for prograde and retrograde reactions.

Departures from $\Delta G = 0$ are necessary to drive reaction, but the value of ΔG alone does not tell us what the rate of reaction will be. Experimental studies indicate that rates are critically dependent on the reactive surface area of minerals, the rate "constant" of reaction, and the activities of catalyzing or inhibiting agents in solution or on mineral surfaces. A general rate law that accounts for these factors and is applicable to surface-controlled reactions is (Steefel and Lasaga, 1994, and references therein)

$$R = \kappa \cdot A \cdot f(a_s) \cdot f(\Delta G) \tag{21.76}$$

where κ is an intrinsic rate constant of reaction, A denotes reactive mineral surface area, $f(a_s)$ is a function of the activity of any species s which catalyzes or inhibits reaction, and $f(\Delta G)$ is a function of the ΔG of reaction. From general chemistry, the intrinsic reaction rate constant can be written as (e.g. Oxtoby *et al*., 1999)

$$\kappa = \kappa^0 \exp\left(-\frac{E_\mathrm{a}}{R}\left[\frac{1}{T} - \frac{1}{T^0}\right]\right) \tag{21.77}$$

where T^0 is a convenient reference temperature (often 298.15 K), κ^0 is the rate constant at the reference temperature, R is the gas constant, and E_a is the activation energy (J mole^{-1}).

The activation energy can be thought of as an energy barrier that must be crossed before the reaction can progress (Fig. 21.17). With increasing temperature, an increasing number of atoms involved in the reaction have the requisite

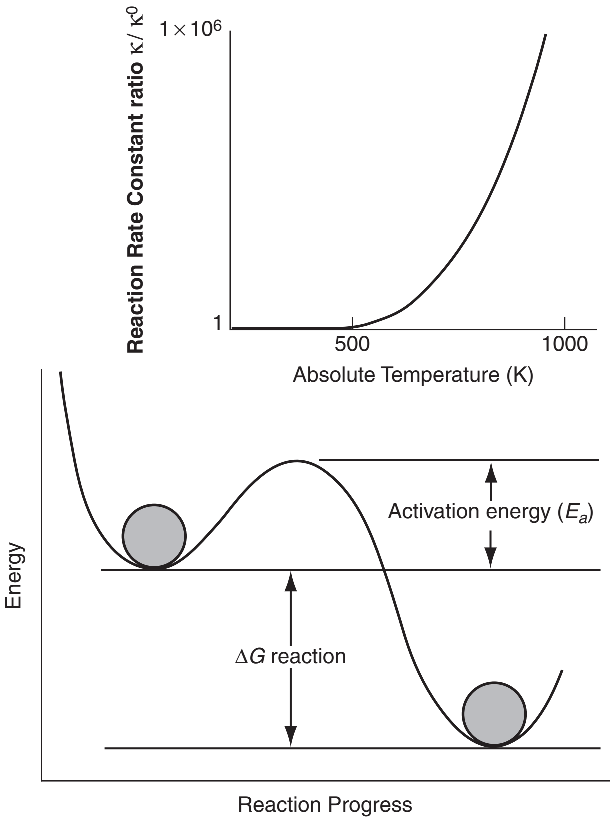


Fig. 21.17 Relation between free-energy change of a reaction, ΔG, and activation energy, E_a, or energy barrier that must be overcome for the reaction to progress. Reaction rate constants commonly obey an Arrhenius-type relation (inset), where a small increase in temperature results in an exponential increase in magnitude. κ^0 is the value of the intrinsic reaction rate constant at the reference T (298.15 K here), and κ is the value at the T_1

energy to jump the barrier. Equation (21.77) can be thought of as consisting of two parts, the exponential factor and the preexponential constant. With increasing temperature, the exponential term gets progressively larger (Fig. 21.17). This strong temperature dependence means that for typical values of E_a (~40–300 kJ mole^{-1}), reaction rates will increase markedly with increasing T along the geothermal gradient (holding other rate parameters constant). As a result, high temperatures favor faster reaction rates and closer approaches to chemical equilibrium, although the value of other variables including ΔG and surface area also play critical roles (Problem 21.11).

The A term in Eq. (21.76) specifies how much mineral surface area is available for reaction with the surroundings and is another major control on rates. The conversion of kyanite to sillimanite in the Barrovian type area was sluggish and involved circuitous reaction pathways (Section 21.7; Fig. 21.9). In fact, some sillimanite zone rocks still contain abundant relict kyanite and little or no sillimanite (Chinner, 1961), even though the equilibrium T for the kyanite → sillimanite transition was overstepped by some 50 °C (Ague *et al*., 2001). On the other hand, retrograde conversion of sillimanite back to kyanite was fairly common (Fig. 16.2(A); Chinner, 1961). Part of the reason for the sluggish prograde kinetics is that the kyanite → sillimanite phase transition is *reconstructive*; that is, it requires a major rearrangement of the mineral structure and hence has a relatively high activation energy. Furthermore, the free energy difference between kyanite and sillimanite at amphibolite facies conditions is rather small, so the ΔG term in the rate expression is also small. Moreover, the presence or absence of fluids had a major impact on the rate of transformation, as we will explore in more detail later on.

But these considerations apply to both the prograde and retrograde paths, so why was the direct transformation of one polymorph to another so much more common during retrogression? The reactive surface area may provide the answer. The sillimanite grows mainly as the "fibrolite" variety comprising masses of very fine-grained needles (Fig. 16.2(A)), whereas the prograde kyanite is found as larger, blocky or prismatic crystals. Consider a cube of kyanite 1.0 mm on a side (kyanite is triclinic, but let us assume a cube for simplicity). It will have a surface area of $6 \times 10^{-6}\,\text{m}^2$. Now consider a cylindrical sillimanite needle in fibrolite 20 μm in diameter and 50 μm long. It will have a surface area of $3.8 \times 10^{-9}\,\text{m}^2$. About 64 000 needles fit into a 1mm cube, so the total surface area of fibrolite needles in a 1mm cube is $2.4 \times 10^{-4}\,\text{m}^2$. This value is a factor of 40 larger than that calculated for the kyanite crystal. Consequently, the area term in the rate expression would have been large on the retrograde path because the surface area of the reactant sillimanite is very large. The increase in the surface area term would have had to outweigh the decrease in κ attending retrograde cooling. The blocky kyanite grains on the prograde path had much lower surface areas relative to their volumes, and this factor almost certainly contributed to the sluggish prograde rates.

The fluid phase also plays an important kinetic role during metamorphism through its effect on diffusion. Diffusion rates in intergranular fluid are orders of magnitude greater than in dry rocks (Sections 5.6, 21.4). Reaction rates in the presence of these fluids are therefore greater than in rocks that lack fluid. The fluid could be provided internally by devolatilization reactions, or could infiltrate the rock from an external source. As discussed in Section 21.7, the exchange of chemical species via transport through a fluid phase can be critical for the kyanite → sillimanite transition. For example, Figure 21.18 shows metasandstone and metapelite from the sillimanite zone of the Barrovian type locality, Scotland, cut by an externally derived pegmatite vein. Where the vein fluids have come in contact with schist, kyanite has been converted to sillimanite. However, outside the vein margins, relict kyanite persists. So the fluids were able to facilitate reaction, in spite of the relatively small surface area/volume ratio of the large kyanite crystals. An example of the impact of fluids on the kinetics of eclogite facies metamorphism is described in Section 22.6.

The rate of another solid–solid transformation, that of calcite to aragonite, has been experimentally calibrated. Hacker *et al*. (1992) studied natural Carrara marble from Italy with an average grain size of 115 μm. The reaction was investigated at a variety of temperatures for pressure oversteps ranging from 0.1 to 1.0 GPa. No free fluid phase occupied the porosity when the experiments started, although traces of H_2O may have been present and some CO_2 could have been evolved during reaction.

Hacker *et al*. (1992) found that the new aragonite crystals nucleated predominantly on the grain boundaries of existing calcite grains along calcite–calcite contacts, once again illustrating the importance of reactive surface area. As expected from the generalized rate constant expression (Eq. (21.76)), the rate of transformation increases markedly with temperature (Fig. 21.19). At temperatures below about 200 °C, the transformation will take a million years or more, whereas at 600 °C it will take days (Fig. 21.19). The systematics of the reverse reaction, aragonite → calcite, will be somewhat different, but the results of Figure 21.19 imply that very rapid exhumation would be required to preserve prograde aragonite if peak temperatures were much in excess of a few hundred degrees.

Reactions involving more than two phases, such as devolatilization reactions, are also modeled using Eq. (21.76) as a basis. A general rate expression for devolatilization reactions can be cast in terms of the product of the rate constant, reactive surface area, and the Gibbs free energy of overstep (Lüttge *et al*., 2004, and references therein). A form of this law generalized for reactions of any stoichiometry is

$$R_{s,l} = \left(\frac{1}{\phi}\right)(\text{sgn})\kappa_l n_{s,l} A_{p,l} |\Delta G_l|^{N_l} \tag{21.78}$$

where $R_{s,l}$ is the rate of release or consumption of species s (e.g. H_2O, CO_2) in mole m$^{-3}{}_{(\text{Fluid})}$ time^{-1} for reaction l, κ_l is the rate constant, $n_{s,l}$ is the stoichiometric coefficient of s in reaction l (positive for products, negative for reactants), $A_{p,l}$ is the rate-limiting reactive surface area of mineral per volume

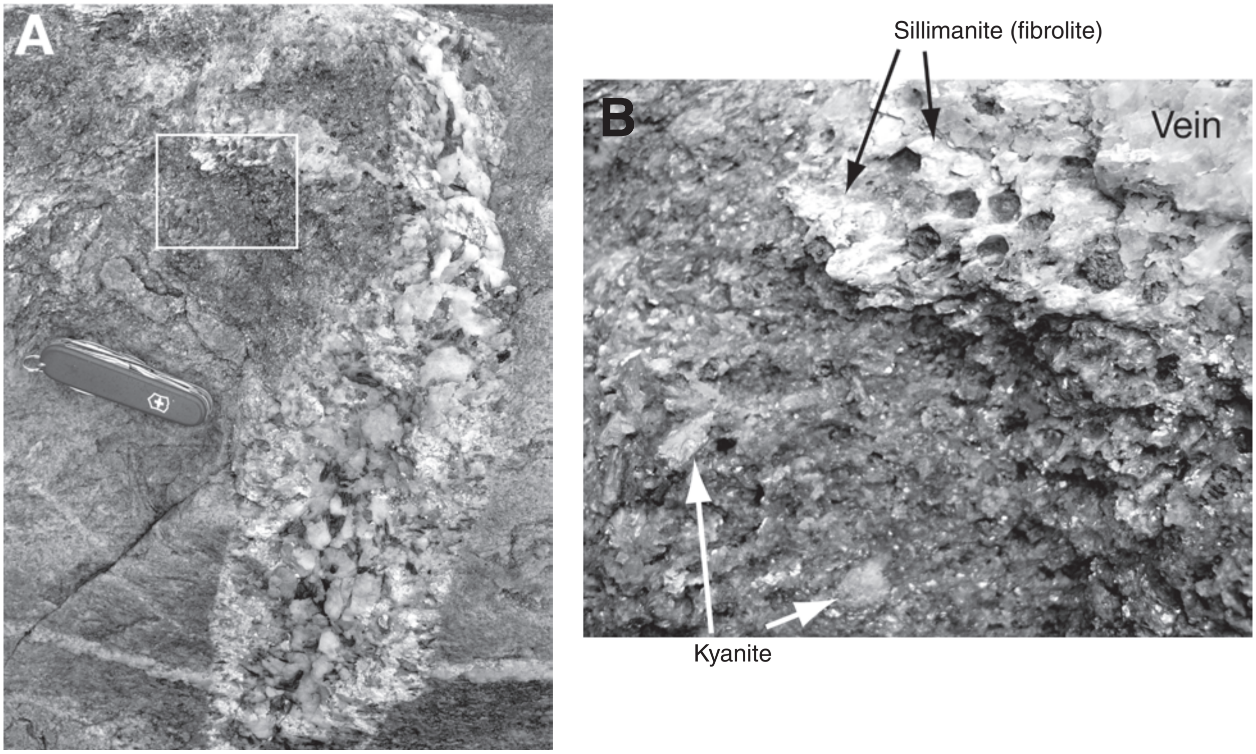


Fig. 21.18 Example of fluid-mediated kyanite to sillimanite reaction, Barrovian type locality, Glen Clova, Scotland. (**A**) Pegmatitic vein cutting metapsammitic rock (bottom) and metapelitic rock (top). Area shown in part (B) denoted with square. (**B**) Where vein fluids came in contact with metapelite, kyanite reacted to form sillimanite. However, relict kyanite is preserved a short distance away from the vein. Equant crystals in sillimanite-rich area are garnets.

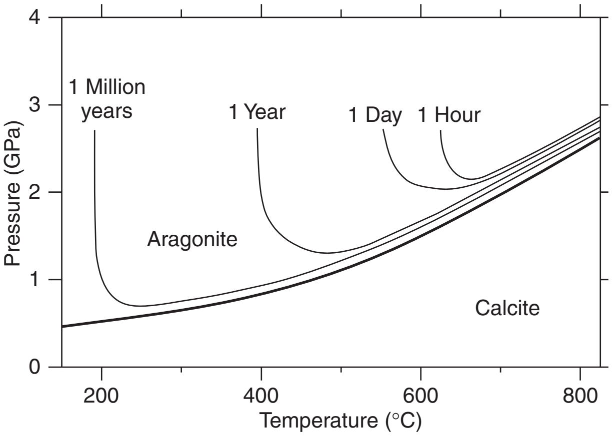


Fig. 21.19 Times for transformation of calcite to 50% aragonite in marble. Heavy line is equilibrium calcite = aragonite boundary. Rate of transformation increases strongly with temperature. Note, however, that transformation is slow near the equilibrium curve, due to the small ΔG of reaction. Uncertainties on time curves below 500 °C are roughly ±100 °C. (From Hacker *et al*., 1992; published by permission of *Science*.)

rock in reaction l, $|\Delta G_l|$ is the absolute value of the Gibbs free energy change for reaction l, N_l is the reaction order, and (sgn) is, by convention, +1 if ΔG_l is negative and -1 if ΔG_l is positive. If the reaction order is unity the kinetics are said to be *linear*, otherwise they are *nonlinear*. The rate-limiting surface area is typically given by the surface area of the slowest reacting mineral, or by the mineral with the smallest overall surface area. This rate expression can be used directly in the overall advection–hydrodynamic dispersion–reaction equation (21.51).

As we discussed above, for a reaction to start, the equilibrium temperature, pressure, or fluid composition must be overstepped in order to provide a sufficiently large free-energy drop to overcome the activation energy of the reaction. The magnitude of the overstepping depends on the type and mechanism of reaction. Temperature oversteps may vary from a few degrees to perhaps as much as 100 K, equivalent to <1 to ~10 kJ mole^{-1} (Walther and Wood, 1984; Putnis and Holland, 1986; Ridley and Thompson, 1986). Devolatilization reactions probably have the smallest oversteps, whereas solid–solid transformations with small entropy and Gibbs free energy changes have the largest oversteps. Pressure oversteps may exceed 0.1 GPa (Ridley and Thompson, 1986; Ernst and Banno, 1991; Rubie, 1998). The size of the temperature or pressure overstep will increase, for example, if reactant minerals are refractory and don't dissolve near equilibrium, or if product minerals are difficult to nucleate. Of course, equilibrium can also be overstepped if a fluid in equilibrium with a given mineral assemblage is displaced by a reactive, disequilibrium fluid.

We can speculate on what the *P*–*T*–*t* evolution of a rock undergoing devolatilization might be like (Fig. 21.20). Consider a dehydration reaction that is being overstepped in

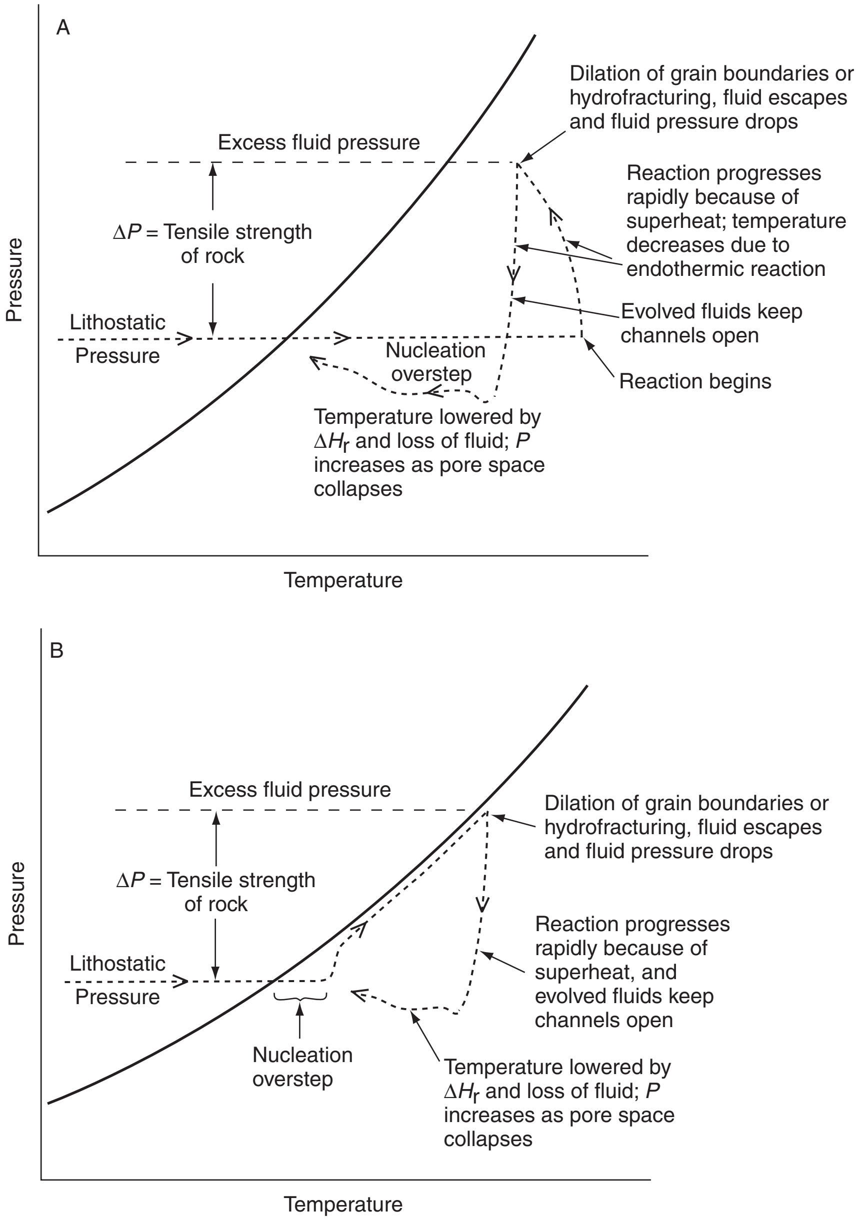


Fig. 21.20 Possible P–T–time paths for a metamorphic rock encountering a devolatilization reaction. Once reaction begins, evolved fluids are unable to escape from the rock if it has low permeability. The fluid pressure thus rises above the lithostatic pressure until it exceeds the tensile strength of the rock at which point fracturing occurs. This allows volatiles to escape and lower their pressure. Once channelways seal themselves, fluid pressures rise and the cycle repeats itself. (**A**) is for a larger T overstep than (**B**).

temperature. The fluid pressure is assumed to be equal to lithostatic. As soon as devolatilization begins, the evolved volatiles are added to the fluid in the pores. If the pores are isolated and the rock permeability is low, the fluid cannot escape from the rock and so causes the fluid pressure to rise above the lithostatic pressure. Devolatilization will be endothermic, but because rocks are reasonably good conductors of heat, temperature will not decrease substantially unless reaction rates are very large (Ague *et al*., 1998; Chapter 22). Consequently, the system tries to reach equilibrium ($\Delta G_l = 0$) mainly by increasing the fluid pressure. Because most dehydration reactions have steep dP/dT slopes at moderate and high pressures, the necessary pressure increases can be very large, even if the amount of overstepping is relatively small (Ague *et al*., 1998). For example, for a typical dP/dT reaction slope of 0.009 GPa K^{-1}, the pressure increase needed to attain equilibrium is ~0.027 GPa for an overstep of only 3 degrees. This pressure is near the tensile strength of rocks.

Initially, ΔG_l and reaction rates will be relatively large due to the overstepping, so pressure may increase rapidly (Eq. (21.78)). Rapid reaction rates can leave a record behind in the rock if, for example, they produce disequilibrium textures in product minerals like garnet (Fig. 12.27). The reaction rates will slow as pressure increases, ΔG_l decreases, and equilibrium is approached. Eventually, the excess fluid pressure in the pores exceeds the tensile strength of the rock (about 0.03 GPa), and grain boundaries are forced apart allowing the fluid to escape and, more importantly, lowering the fluid pressure (Fig. 21.20(A)). With the sudden drop in fluid pressure, the rock finds itself well above the equilibrium temperature, ΔG is once again large, and the reaction progresses rapidly with the evolution of more volatiles, which keep open the channels. Eventually, the rock returns to equilibrium conditions, either because the reaction has consumed heat and the rock mass has cooled, or because the rock deforms, collapsing porosity and elevating fluid pressures (Sibson *et al*., 1975). Once mineral grains reestablish their equilibrium dihedral angles, fluid flow becomes restricted and the fluid pressure again begins to increase, and the cycle is repeated.

The situation is similar even if the amount of initial overstepping is small (Fig. 21.20(B)). Here, once the reaction begins after the initial overstep, devolatilization will occur near the equilibrium curve as the rock is heated. If the fluid phase does not wet grain boundaries and/or the permeability is very small, the fluid pressure will slowly increase until grain boundary dilation or hydrofracturing occurs. Then, fluid pressure drops and the remainder of the cycle proceeds as described above.

Fluids escaping from the source rock flow along grain boundaries, but as they rise into overlying rocks their flow is likely to be channelized. Pulses of fluid pressure cause hydrofracturing of the overlying rocks. Because of the decreasing solubility of silica with falling temperature, fractures are likely to seal themselves with quartz. Successive pulses of fluid therefore have to reopen these veins. Extensional veins with repeated fillings of quartz are common in metamorphic terranes (e.g. Etheridge *et al*., 1984; Section 21.12.3) and attest to this pulsing flux of fluids.

In this scenario, a devolatilization reaction is a pulsating phenomenon, the frequency of which is determined to a large degree by the tensile strength of the rock (Walder and Nur, 1984). Once a reaction overstepped by more than a few degrees begins, the fluid pressure increases can be large enough and fast enough to cause hydrofracturing on decade to century timescales (Ague *et al*., 1998). If these fracturing events occur in tectonically active areas, they may promote seismogenic failure. In such cases, the development of high fluid pressures and the resultant rock failure would be the direct result of active metamorphism. It is even possible that the recurrence intervals of earthquakes on some faults are related to the timing of pulsating reactions. Much more research is needed to explore the relationships between metamorphic devolatilization and seismic hazard, both in the crust and in subduction zones.

Fluid composition is another key factor in kinetics. The treatment of fluids in Chapter 20 is based on equilibrium thermodynamics. This treatment implies that reaction rates are fast relative to changes in temperature, pressure, and composition. But what if the rates are slow with respect to these changes so that minerals and fluids do not reach equilibrium? Let us examine the calcite–dolomite–quartz–tremolite–talc isobaric invariant point from Figure 20.8. A rock containing calcite, dolomite, and quartz is heated along the dashed path in Figure 21.21. When the first reaction begins, T–X_{CO_2} conditions are not constrained to stay on the equilibrium curve because the reaction rate is slow relative to the rate of heat input. Consequently, the T–X_{CO_2} path oversteps the equilibrium curve before any of the reactants are used up. In the example case the degree of overstepping is large due to the sluggish reaction kinetics.

If all the reaction rates are slow relative to the heating, then by the time the T–X_{CO_2} path reaches point A in Figure 21.21, all the reactions (including any metastable extensions) can be proceeding (Lüttge *et al*., 2004). None of the reactants will have been consumed completely, and the rock will contain calcite–dolomite–quartz–tremolite–talc, identical to the equilibrium isobaric invariant point assemblage. But here the mineral assemblage at point A is not in chemical equilibrium, and the $T-X_{CO_2}$ path never actually reached the invariant point conditions. An important implication is that mineral assemblages alone may not tell us whether rocks followed disequilibrium $T-X_{CO_2}$ reaction paths.

Are slow reaction rates important in nature? Recent theoretical calculations and field studies suggest that disequilibrium in metacarbonate rocks may be significant, particularly in contact aureoles where heating rates are rapid (Lüttge *et al*., 2004; Müller *et al*., 2004; Nabelek, 2007). In their study of regionally metamorphosed metapelite and amphibolite, Baxter and DePaolo (2002a, b) found that rates of reaction were too slow to keep pace with

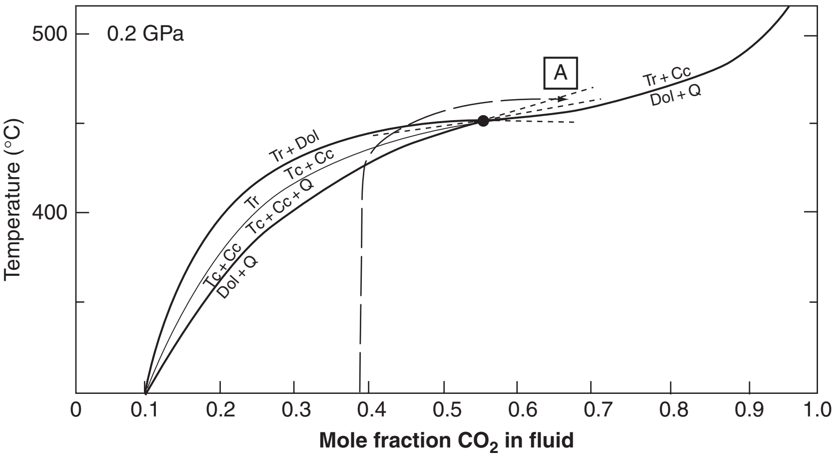


Fig. 21.21 Hypothetical kinetic T–X_{CO_2} reaction path denoted by dashed line does not follow isobaric univariant curves or reach the isobaric invariant point because of slow reaction rates. Nonetheless, the rock would contain the invariant point mineral assemblage when it reached point A. Dol, dolomite; Q, quartz; Tc, talc; Tr, tremolite; Cc, calcite. (Modified from Lüttge *et al*., 2004.)

evolving temperature, pressure, and composition conditions. Determining whether or not large reaction oversteps are common and distinguishing between kinetic reaction paths and equilibrium are important challenges in petrology today.

Reaction kinetics also have important consequences for mass transfer during fluid infiltration. The propagation of geochemical fronts across lithologic contacts discussed in Section 21.9 can be used to illustrate the impact of kinetics. From Eqs (21.53) and (21.54) we can write

$$\phi\frac{\partial c_s}{\partial t} = -\bar{v}_x\phi\frac{\partial c_s}{\partial x} + D_{\mathrm{HD}}\phi\frac{\partial^2 c_s}{\partial x^2} - (1-\phi)R_s^{Sol} \tag{21.79}$$

where, as before, R_s^{Sol} is the reaction rate for the solid. We can rewrite this term using a very simple rate expression that is commonly used in the literature

$$R_s^{Sol} = \kappa'\left(c_{s,Sol}^{equil} - c_{s,Sol}\right) = \kappa'\left(\frac{c_s}{K_v} - c_{s,Sol}\right) \tag{21.80}$$

where κ' is a rate constant (time^{-1}), c_s is the concentration of species s in the fluid, $c_{s,Sol}$ is the concentration of s in the solid, and $c_{s,Sol}^{equil}$ is the concentration of s in the solid at equilibrium. The equilibrium value for the solid (c_s/K_v) is given by the partition coefficient expression (21.52).

The difference in concentration between the actual and equilibrium values acts as a proxy for the driving force (ΔG) of reaction; when the actual concentration approaches the equilibrium concentration, the rate goes to zero. Note that in this simplified treatment, the value of κ' is somewhat different from κ in Eq. (21.77), as κ' includes factors such as the reactive surface area. Furthermore, we are assuming a constant κ', but in nature it will vary with temperature and changes in reactive surface area accompanying reaction. Substitution of (21.80) into (21.79) and rearranging yields

$$\frac{\partial c_s}{\partial t} = -\bar{v}_x\frac{\partial c_s}{\partial x} + D_{\mathrm{HD}}\frac{\partial^2 c_s}{\partial x^2} + \kappa'\frac{1-\phi}{\phi}\left(c_{s,Sol} - \frac{c_s}{K_v}\right) \tag{21.81}$$

which allows one to calculate the evolution of fluid composition in time and space due to infiltration and kinetically controlled reaction. The composition of the solid is easily found from

$$\frac{\partial c_{s,Sol}}{\partial t} = -\kappa'\left(c_{s,Sol} - \frac{c_s}{K_v}\right) \tag{21.82}$$

Let us consider infiltration of a reactive fluid across a lithologic contact as we did before in Section 21.9, but now with kinetic constraints on reaction rates. All real flow systems will have diffusion and mechanical dispersion, but it is easiest to illustrate concepts using an advection-only example. An analytical solution for this case is given by Lassey and Blattner (1988). If advection is very slow, there will be considerable time for the fluid to react with the rock and, thus, there is a good chance that local equilibrium will be attained. On the other hand, if flow is fast, there will be less time for the fluid at any particular place to react with the rock, and so significant disequilibrium may result. Thus, the rate of reaction relative to the rate of transport is a key factor in determining how far a system will be from equilibrium. This concept can be summarized using the dimensionless *Damköhler-I* number (Da-I; Boucher and Alves, 1959)

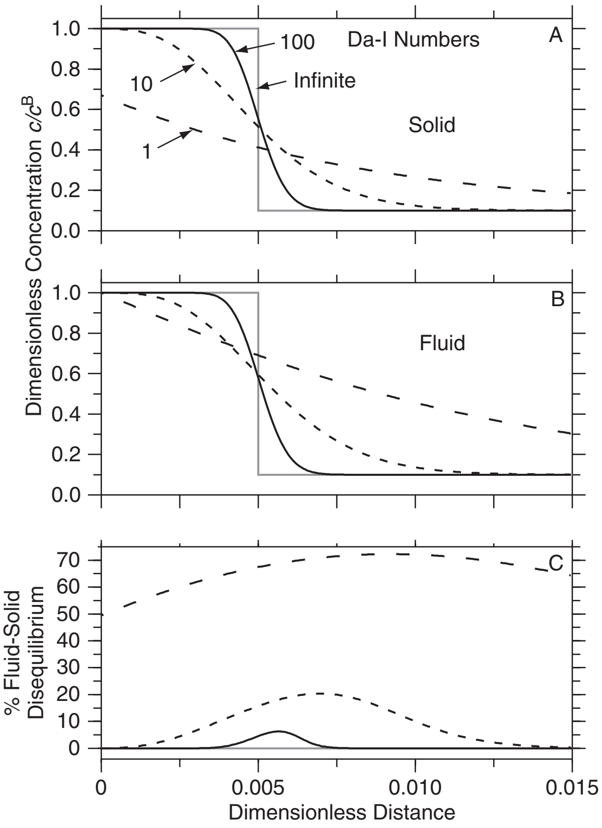


Fig. 21.22 Advection reaction with kinetically controlled rates for a porosity of 10^{-3}. Concentration (c) normalized to equilibrium value at fluid input boundary (c^B) plotted against normalized distance. (**A**) Solid composition. Degree of fluid-solid disequilibrium increases with decreasing Damkhöler-I number (Da-I). Large Da-I corresponds to fast rates near equilibrium; small Da-I corresponds to slow rates farther from equilibrium. Note how profiles broaden with increasing disequilibrium. (**B**) Fluid composition. (**C**) Percentage disequilibrium between fluid and solid.

$$\text{Da-I} = \frac{\kappa' L}{\bar{v}_x} \tag{21.83}$$

in which L is a length scale of flow and reaction. It is convenient to take $L = \bar{v}_x t$, where t is the total time of flow (Lassey and Blattner, 1988). A different Damköhler number can be defined if transport is by diffusion.

Results for several different Da-I numbers are shown in Figure 21.22. If perfect fluid–rock equilibrium is maintained, Da-I is infinitely large and the geochemical front propagates as a sharp step function. As Da-I decreases, the rate of reaction decreases relative to transport and the geochemical front becomes increasingly "smeared out" (Fig. 21.22(A)). Consequently, the distance over which fluid and rock are out of equilibrium gets larger as rates get slower and Da-I decreases (Fig. 21.22(B), (C)). Furthermore, the degree of disequilibrium increases with decreasing rates (Fig. 21.21

(C)). Nonetheless, as shown by Bickle (1992), the distance of propagation of the geochemical front is still directly related to the time-integrated fluid flux by Eq. (21.60). This is shown in Figure 21.22(A) by the fact that the midpoints in concentration for the Da-I = 100 and 10 cases coincide with the geochemical front distance for complete fluid–rock equilibrium. Thus, time-integrated fluid fluxes can still be estimated from geochemical fronts, even if kinetics are important. However, it is also clear that geochemical fronts become increasingly difficult to identify when fluid–rock disequilibrium is large, such as for the Da-I = 1 case in Figure 21.22(A).

The smoothing of geochemical fronts in Figure 21.22(A) looks superficially like the smoothing effects of hydrodynamic dispersion (e.g. Figs. 21.14, 21.15). Indeed, in the field, it may be difficult to discern if front smoothing is caused by hydrodynamic dispersion, kinetic factors, or both. If two contacts can be found, however, much tighter constraints can be placed on infiltration and reaction. For example, if advection and slow reaction rates operate in the absence of hydrodynamic dispersion, the smoothing will only take place in the direction of flow. This is not the case for the natural Sr data of Figure 21.16, for here diffusion from metapelite to metacarbonate clearly took place at the right side contact in a direction opposing that of the flow. In fact, numerical results indicate that this profile can be fit well with a single coefficient of hydrodynamic dispersion, which suggests that, for this case at least, local equilibrium was approached. Otherwise the geochemical front would be more smeared out in the direction of flow, and the profile would be inconsistent with a single coefficient of hydrodynamic dispersion. Relatively fast rates may have been favored in these rocks because their grain size is rather small (~200 μm or less) and, thus, reactive surface areas were large. In another example, however, strong disequilibrium for diffusion–reaction of Sr across a metapelite–amphibolite contact was found by Baxter and DePaolo (2002a, b). These contrasting results show that rates of reaction relative to transport can vary significantly during metamorphism.

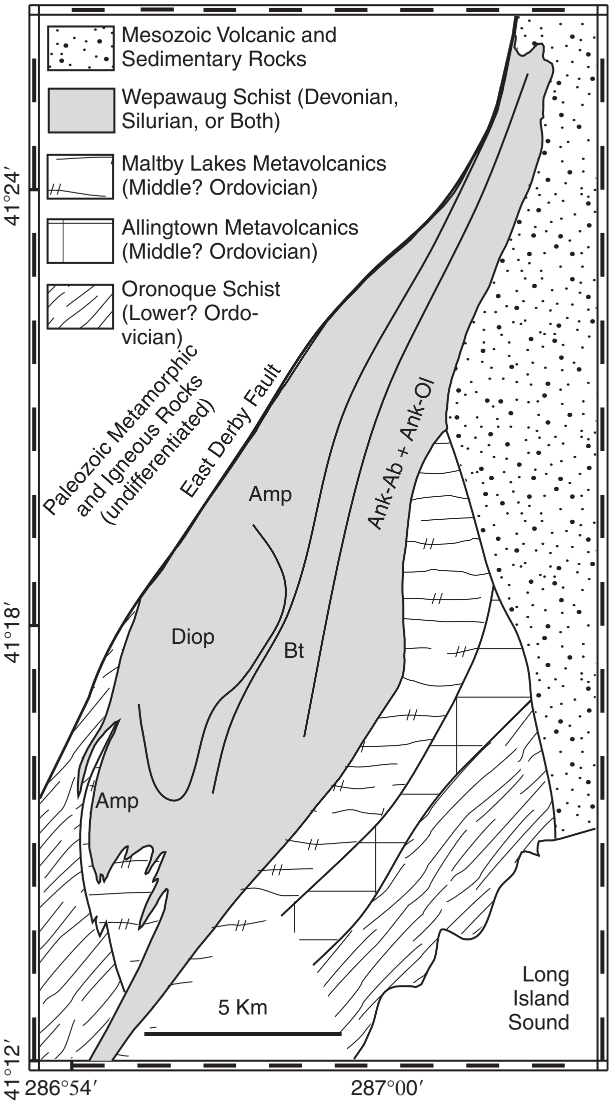


Fig. 21.23 Simplified geologic map of the Wepawaug Schist and surrounding area in south-central Connecticut, USA. The Wepawaug is dominated by metapelitic rocks, but also contains some interbedded metacarbonate layers. Metamorphic grade increases from east to west. Ankerite–Albite and Ankerite–Oligoclase (Ank-Ab + Ank-Ol), Biotite (Bt), calcic amphibole (Amp), and Diopside (Diop) index mineral zones shown for metacarbonate layers. (After Ague, 2003a.)

21.12 MULTIDIMENSIONAL TRANSPORT AND REACTION

Diffusion and mechanical dispersion operate in the direction of advective flow, but can also act transverse to the flow as well. This fact has considerable implications for how rock packages composed of chemically and isotopically different rock types react during metamorphism. We will examine three cases where multidimensional transport and reaction can be important: devolatilization of metacarbonate rocks, subduction zone mélange, and fracturing and vein formation.

21.12.1 Intercalated metacarbonate and metapelitic rocks

The examples of Section 21.9 indicate that a large component of fluid flow during regional metamorphism of the metacarbonate rocks studied was parallel to layering. Diffusion and mechanical dispersion (collectively termed hydrodynamic dispersion) operated in the direction of flow and across the layers as well. When metacarbonate rock layers are intercalated with dehydrating metapelitic rocks, the transport of volatiles across layers has to be considered in order to assess reaction processes because, as we saw in Chapter 20, the infiltration of water can drive decarbonation.

In a number of metasedimentary sequences in New England, such as the Wepawaug Schist of Connecticut (Fig. 21.23; Hewitt, 1973), one to ten meter thick metacarbonate layers are intercalated with metapelitic rocks. During prograde heating, the schists produced large amounts of water, whereas reactions in metacarbonate layers involved CO_2 and water. Major minerals in the lowest grade metacarbonate rocks

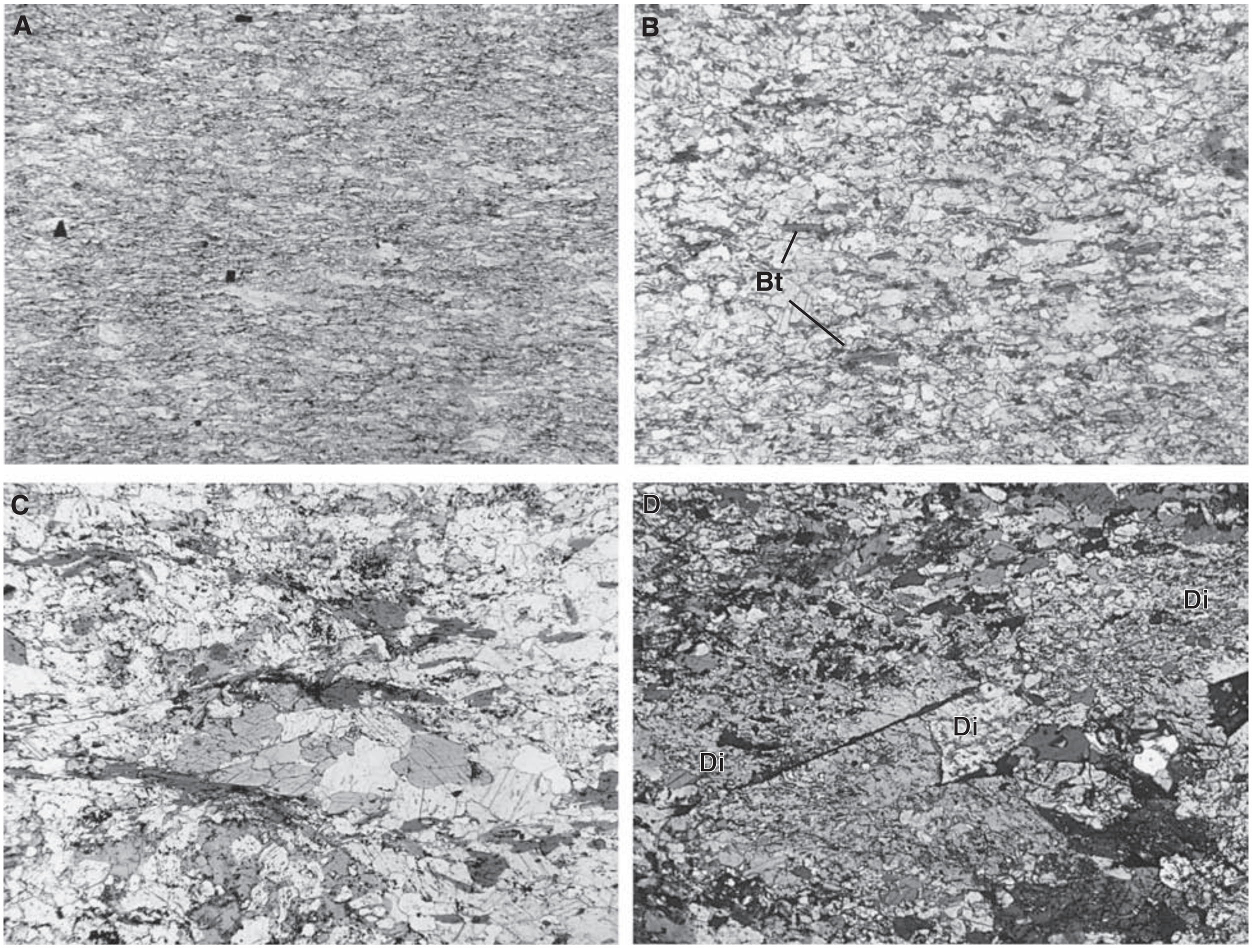


Fig. 21.24 Photomicrographs of metacarbonate rocks of the Wepawaug Schist. Fields of view are 4.5 mm. Crossed-polarized light for part (D), plane-polarized light for others. (**A**) Ankerite–Albite zone. Fine-grained matrix consists of calcite, dolomite, muscovite, quartz, and albite, with accessory pyrite and organic matter. (**B**) Biotite zone. (**C**) Amphibole zone. Hornblende in center of photo with 120° cleavage, together with biotite. (**D**) Diopside zone. Twinned, inclusion-rich diopside crystal (Di) spans the width of the photo.

are calcite, ankerite (dolomite with some Fe^{2+} substitution), quartz, muscovite, and albite. Index mineral zones in the metacarbonate layers can be defined based on the appearance of oligoclase, biotite, amphibole, and diopside with increasing grade (Figs. 21.23, 21.24). Simplified model reactions that represent the typical prograde sequence are (Hewitt, 1973; Ferry, 1992)

Ankerite–oligoclase zone:

$$\text{paragonite} + \text{calcite} + 2\ \text{quartz} = \text{albite} + \text{anorthite} + CO_2 + H_2O \qquad (21.84)$$

Biotite zone:

$$\text{muscovite} + 3\ \text{dolomite} + 2\ \text{quartz} = \text{phlogopite} + \text{anorthite} + 2\,\text{calcite} + 4CO_2 \qquad (21.85)$$

Amphibole zone:

$$5\ \text{dolomite} + 8\ \text{quartz} + H_2O = \text{tremolite} + 3\ \text{calcite} + 7CO_2 \qquad (21.86)$$

$$6\ \text{calcite} + 5\ \text{phlogopite} + 24\ \text{quartz} = 5\ \text{K-feldspar} + 3\ \text{tremolite} + 2H_2O + 6CO_2 \qquad (21.87)$$

Diopside zone:

$$\text{tremolite} + 3\ \text{calcite} + 2\ \text{quartz} = 5\ \text{diopside} + 3CO_2 + H_2O \qquad (21.88)$$

$$\text{phlogopite} + 3\ \text{calcite} + 6\ \text{quartz} = 3\ \text{diopside} + \text{K-feldspar} + 3CO_2 + H_2O \qquad (21.89)$$

In addition, reaction (20.34), which produces clinozoisite or zoisite, was important in the Biotite zone and at higher grades. Note that all of these reactions liberate CO_2. Hypotheses of across-layer transport involving these reactions have been tested using numerical models of advection and hydrodynamic dispersion incorporating kinetic rates of reaction (Ague and Rye, 1999; Ague, 2000, 2002).

Let us consider metacarbonate layers undergoing devolatilization according to the biotite-producing reaction (Fig. 21.25, 21.26) (Eq. (21.85)). During heating, CO_2 is produced in the metacarbonate layers, whereas water is produced in the

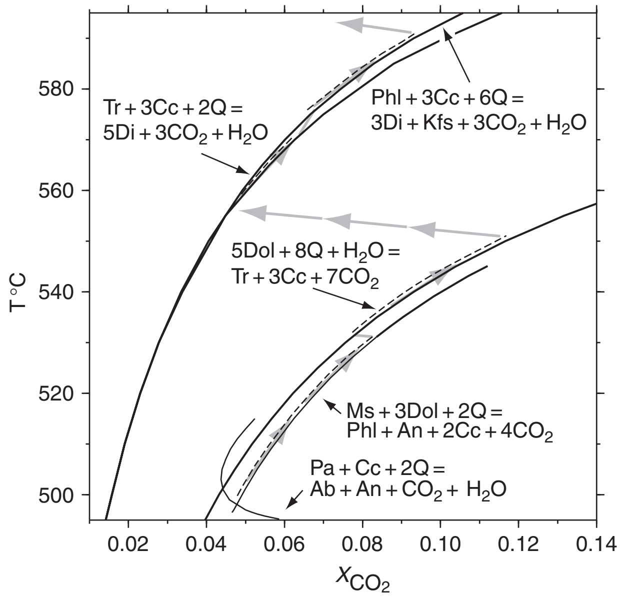


Fig. 21.25 T–X_{CO_2} diagram illustrating reactions (21.84) through (21.89). T–X_{CO_2} path for heating of a thin metacarbonate layer intercalated with dehydrating metapelitic rocks shown with gray arrows. T–X_{CO_2} values for layer interior denoted with solid lines, values for layer edges shown with dotted lines (Fig. 21.26). See text for discussion. (After Ague and Rye, 1999; published by permission of Oxford University Press.)

surrounding metapelitic layers. Thus, fluids in the metacarbonate layers will have more CO$_2$ than fluids in the metapelites, and concentration gradients will be set up at lithologic contacts between the two rock types (Fig. 21.26(A)). Diffusion and mechanical dispersion operate when such gradients exist (Eq. (21.12)). As a consequence, water will be transported into the metacarbonate layers and CO$_2$ will leave.

At a given temperature, input of a fluid that is more water-rich than the equilibrium fluid composition will push the reaction into the stability field of the prograde products (biotite and plagioclase feldspar in this case; Fig. 21.26(B)). Reaction (21.85) is therefore driven to produce more CO$_2$ in order to return to the equilibrium fluid composition. In this way, the input of water and release of CO$_2$ lead to large amounts of CO$_2$ loss from the layer during heating. Conversely, decreasing the water activity in the metapelitic rocks by input of CO$_2$ will decrease the equilibrium dehydration temperature (Fig. 16.8), pushing reactions into the stability field of the reactants and thus driving more dehydration. Therefore, dehydration of metapelite and decarbonation of metacarbonate are coupled by mass transfer of volatiles between layers. The volatiles are ultimately transported out of the system by fluid flow along the layering (Fig. 21.27).

The T–X_{CO_2} path of the fluid bears a superficial resemblance to the internally buffered situation (Section 20.5). For internal buffering, the fluid is generated by reactions taking place in the local rock; no fluid infiltrates, but some may leave. The critical difference is that, for the open-system case discussed here, water continually infiltrates and CO$_2$ continually leaves the rock layer during reaction, so a large amount of devolatilization takes place along the heating path, rather

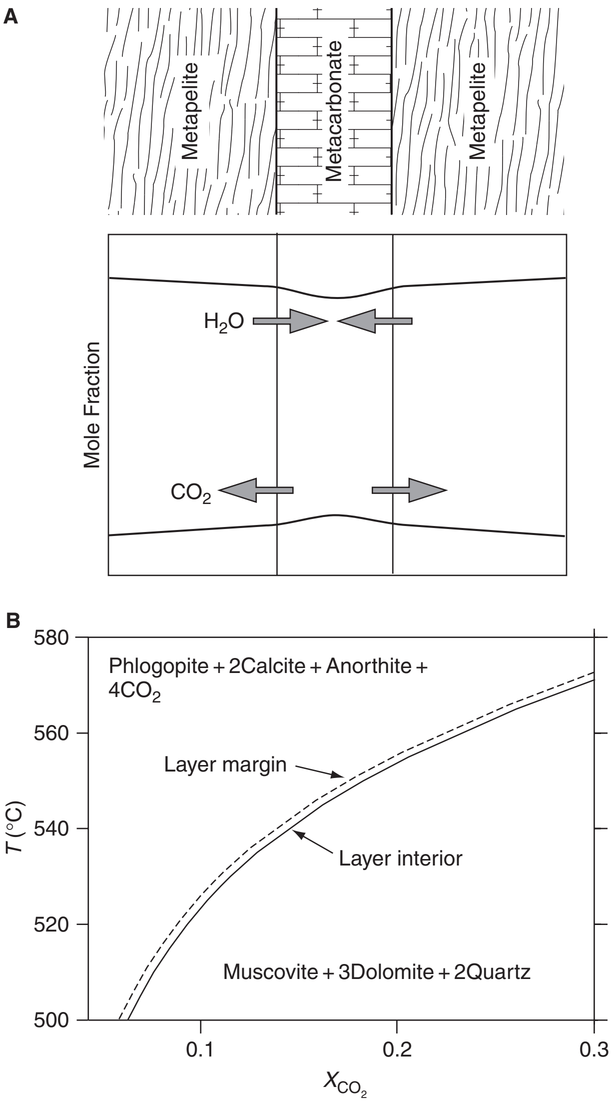


Fig. 21.26 (**A**) Gradients in fluid composition set up by devolatilization of interbedded metacarbonate and metapelitic layers. X_{CO_2} will be somewhat greater in the metacarbonate layer due to reactions that produce CO$_2$, so CO$_2$ will tend to move out of the layer by diffusion and mechanical dispersion down the concentration gradients, whereas H$_2$O will enter (arrows). (**B**) T–X_{CO_2} diagram for reaction (21.85). Equilibrium curve shown with solid line. Dotted line shows that X_{CO_2} values at a given temperature will be somewhat less at the margins of metacarbonate layers than in their interiors, thus driving devolatilization and CO$_2$ loss.

than just at isobaric invariant points. Note that the $T-X_{CO_2}$ path can remain close to the equilibrium curve. This is because the gradients in fluid composition between layers can be quite small and still drive considerable hydrodynamic dispersion and reaction on geologic timescales (Fig. 21.26; Ague, 2002). Because the fluids from the metapelitic surroundings enter at lithologic contacts, reaction begins there and then progresses inward toward the centers of metacarbonate layers, as observed in the classical study of Hewitt (1973) (Fig. 21.27).

Eventually, one of the reactants (dolomite, muscovite, or quartz) is used up in the layer and reaction (21.85) ceases to

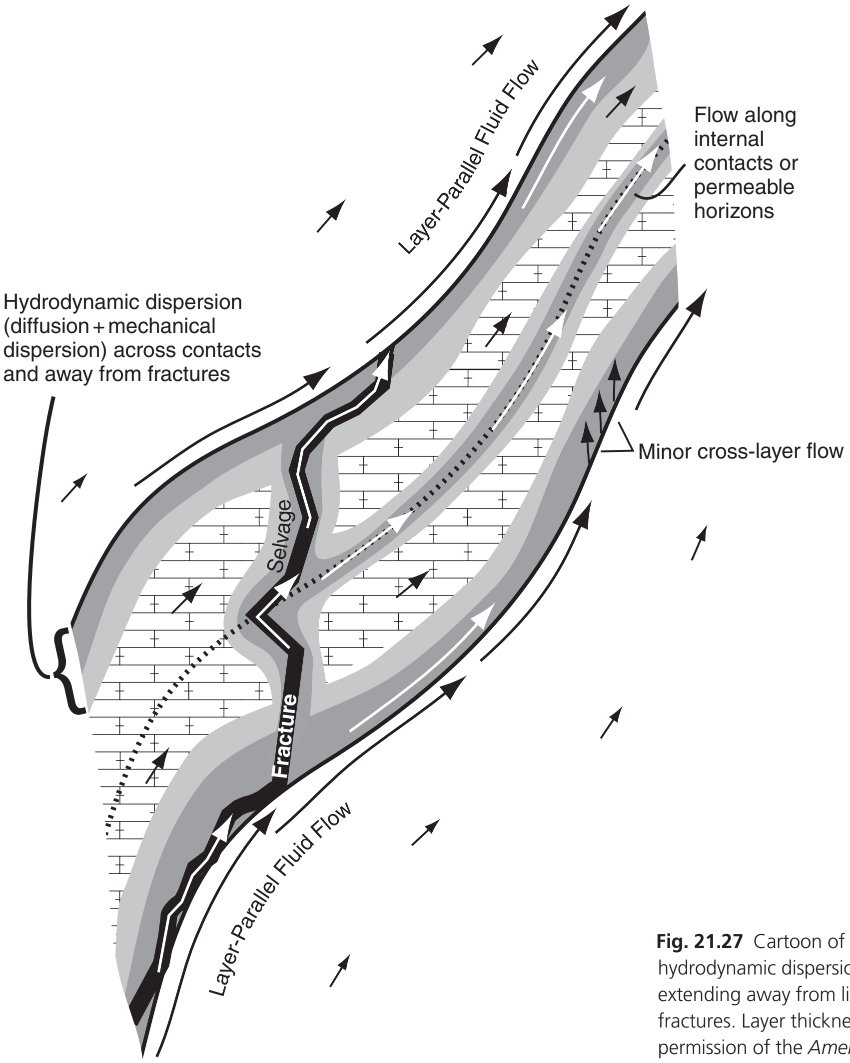


Fig. 21.27 Cartoon of metacarbonate layer illustrating fluid flow and hydrodynamic dispersion. Gray shading denotes devolatilization extending away from lithologic contacts and permeable zones such as fractures. Layer thickness ~1 m. (From Ague, 2003a; published by permission of the *American Journal of Science*).

produce CO_2. Water from the dehydrating surroundings can still be transported into the layer, however, so the X_{CO_2} of the fluid decreases until the next decarbonation reaction (reaction (21.86)) is reached (Fig. 21.25). This general pattern is repeated for each reaction encountered during prograde heating, producing a "zig-zag" reaction path across the T–X_{CO_2} diagram that is quite different from the purely externally or internally buffered cases considered earlier. The path in Figure 21.25 clearly falls between the internal and external buffering regimes.

Other elements were transported in addition to volatiles (Tracy *et al*., 1983; Ague, 2003a). For example, the metapelitic rocks of the Wepawaug Schist contain muscovite but lack K-feldspar. Fluids equilibrated with muscovite-bearing assemblages will in general have lower K concentrations than fluids equilibrated with K-feldspar-bearing ones (Fig. 21.8). Thus, fluids in metapelites had less K than fluids in metacarbonate layers in which biotite was breaking down to produce K-feldspar and tremolite or diopside (reactions (21.87), (21.89)). The K therefore diffused out of and was lost from the metacarbonate layers, to such an extent that K-feldspar is actually quite rare in the metacarbonate rocks of the Wepawaug Schist (Ague, 2002, 2003a).

Hydrodynamic dispersion of H_2O–CO_2 operates on length scales ranging from millimeters to ten meters or more, comparable to the thickness of the metacarbonate layers. Consequently, even though advection is the agent that transports mass at regional scales, hydrodynamic dispersion acting at the scale of the layering can produce the bulk of the devolatilization reaction that occurs in many field settings. Field tests have confirmed the importance of such across-layer transport and reaction during regional metamorphism (Ague, 2002, 2003a; Penniston-Dorland and Ferry, 2006). In addition to hydrodynamic dispersion, external fluids that advectively flow through metacarbonate layers along grain boundaries or through cracks can also drive CO_2 loss (Fig. 21.27; Section 21.12.3).

Dehydration of metapelitic rocks is not the only source for water in regionally metamorphosed sequences. For example, oxygen isotope studies indicate that water-rich fluids derived from or equilibrated with leucocratic igneous rocks were important for driving devolatilization in the Diopside zone of the Wepawaug Schist (Palin, 1992; van Haren *et al*., 1996). These fluids usually have smaller $\delta^{18}O$ values than fluids equilibrated with metasediments, so isotopic studies of reacted rocks are a powerful means to track their flow.

21.12.2 Subduction zone mélange

In subduction zones, water and CO_2 are recycled into the deep Earth by metasediments and hydrothermally altered oceanic crustal rocks that are carried far into the mantle (Peacock, 1990; Kerrick and Connolly, 2001a, b; Schmidt, 2003; Richard *et al*., 2006). The fluids released during prograde metamorphism of these rocks flow along the interface between the subducting slab and the mantle, or ascend into the overlying mantle wedge. The ascending fluids can cause major chemical changes in the mantle, involving, for example, silicification of the mantle wedge and transport of a wide variety of major and trace elements (Sorensen and Grossman, 1989; Bebout and Barton, 1989, 2002; Manning, 1997; King *et al*., 2003). At depths greater than ~100 km, metamorphic fluids ascending into the mantle trigger partial melting, thereby producing arc magmatism. Consequently, metamorphic devolatilization is essential for arc magma generation at convergent margins (Fig. 21.1).

Highly deformed parts of subduction complexes, including the slab–mantle interface, can be occupied by *mélange zones* (Fig. 21.28(A)). These commonly consist of an ultramafic matrix which hosts dismembered blocks of metamorphosed mafic and ultramafic igneous rocks, and in some cases, metasedimentary rocks (Fig. 21.28; Bebout and Barton, 2002; King *et al*., 2003). The blocks range in size from a few centimeters across to hundreds of meters or even kilometer scale. Mounting evidence indicates that large quantities of subduction zone fluids are channelized through mélange zones (e.g. Bebout and Barton, 1989, 2002; Breeding *et al*., 2004). Because fluids equilibrated with the ultramafic matrix are strongly out of chemical equilibrium with the blocks, the fluids mediate chemical reactions on block edges, producing reaction "rinds" ranging in thickness from the centimeter scale to the meter scale or larger (Fig. 21.28(C); Dixon and Ridley, 1987; Bebout and Barton, 2002; King *et al*., 2003).

The reaction rinds have chemical signatures indicative of infiltration metasomatism from the matrix into the blocks as exposed, for example, on the island of Syros (Cyclades, Greece). The ultramafic matrix is rich in elements like Mg and Ni relative to most blocks. Thus, chemical profiles reveal strong increases in Mg and Ni content at block margins (Fig. 21.29). Much of the mass transfer was by diffusion into the blocks, but flow along block contacts and through fractures cutting the blocks was also important in many cases. The principles of metasomatic zonation given in Section 21.8 apply, with the rinds consisting of essentially monomineralic glaucophane or omphacite. Note, however, that the permeability of the blocks on Syros was limited, for most metasomatic reactions were restricted to block margins, leaving the interiors significantly less altered (Ague, 2007).

Mica (phengite) will be present in subducted metasediments whose protoliths contained clays. In addition, mafic rocks that underwent hydrothermal alteration on the seafloor can contain elevated levels of potassium. Consequently, phengite will also crystallize in these rocks during metamorphism. Fluids equilibrated with ultramafic mélange matrix will be undersaturated with respect to mica and be poor in large ion lithophile elements (LILE) like K, Rb, and Ba (similar fluids would be generated by dehydration of ultramafic rocks in the slab as well). Consequently, when these fluids infiltrate mica-bearing rocks, reactions will occur that destroy the micas. The LILE are therefore stripped from the mica-bearing lithologies and are carried away in the fluids (Breeding *et al*., 2004), while other elements, particularly Na, are added (Fig. 21.30). Direct prograde dehydration of mica may also liberate LILE (e.g. Catlos and Sorensen, 2003). Furthermore, U and Pb can be mobilized in addition to LILE during fluid–rock interaction (e.g. Breeding *et al*., 2004). If the fluids encounter other reactive rock types, such as mafic mélange blocks that lack micas, then new micas can precipitate in reaction rinds. However, if the fluids ascend into the mantle wedge overlying the subduction zone, they will carry LILE and other elements with them (Fig. 21.28(A)). Arc magmas are characteristically enriched in a suite of elements that includes LILE, U, and Pb (e.g. Johnson and Plank, 1999; Kelemen *et al*., 2003). Thus, the fluid-driven transfer of these elements from the subduction zone into magmatic source regions in the mantle is one potential mechanism for producing the distinctive geochemical signature of arcs. Partial melting of metasediments is another important mechanism (Johnson and Plank, 1999; Spandler *et al*., 2007). Determining how arc magmas acquire their chemical and isotopic signatures will be an active area of research for many years to come.

21.12.3 Fracturing, vein formation, and fluid flow

The discussion in previous sections illustrates how elevated fluid pressures evolve during prograde devolatilization, leading to hydrofracturing. Moreover, far-field tectonic stresses produce rock fracturing and faulting; if large enough, the resulting rock failure will release seismic energy. Veins are fractures in which mineral precipitation has taken place (Fig. 21.31), and they can host large volumes of fluid flow because of their elevated permeability. Given the cyclic nature of fluid generation and deformation, veins are normally the sites of repeated cracking and mineral precipitation, resulting in the formation of crack–seal textures (Ramsay, 1980). The incorporation of inclusions of rock ripped from vein margins attests to brittle deformation behavior (Fig. 21.31(C)). As we discussed in Section 21.6, quartz is normally the most common vein mineral, but a wide array of other phases can also be found, including calcite, feldspars, sheet silicates, Al_2SiO_5 polymorphs, and sulfide minerals. In subduction zone veins, albite, lawsonite, glaucophane, and omphacite pyroxene are not uncommon.

Fracturing and brittle deformation are often considered to be of limited importance in the middle and lower crust, where

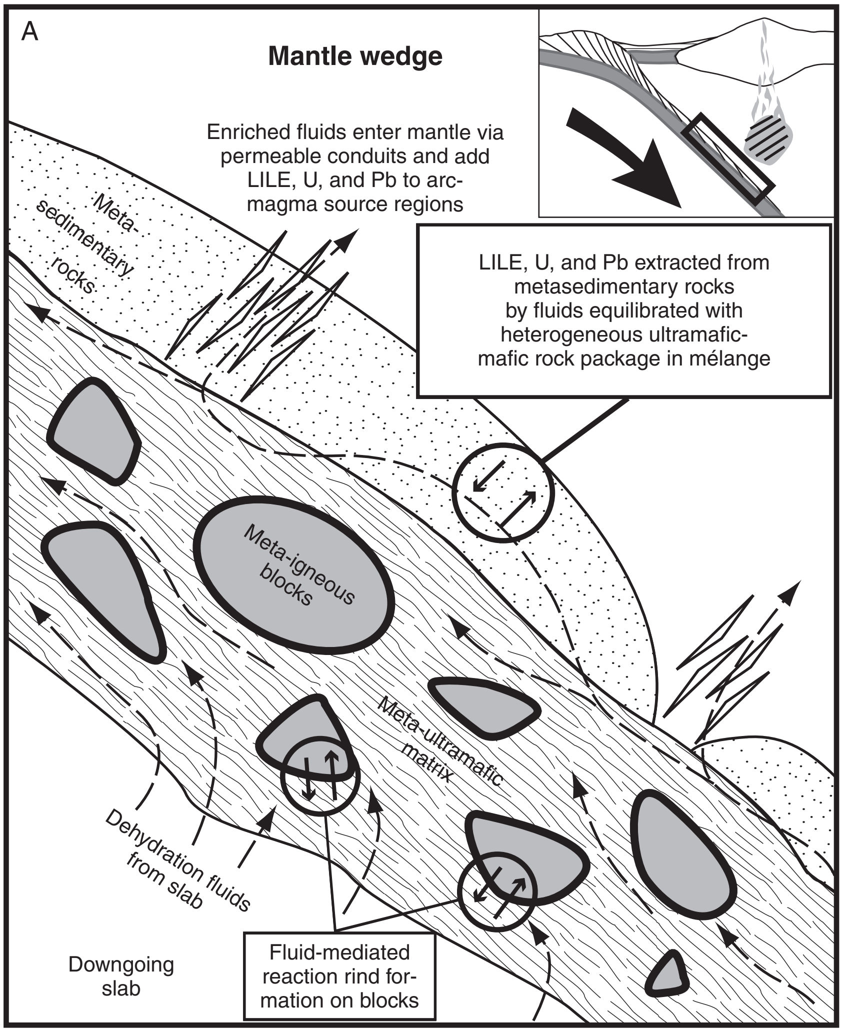


Fig. 21.28 Mélange zone, Syros, Greece. (**A**) Cartoon of fluid flow and chemical reaction modeled on geologic relations between mélange zone and metasedimentary rocks on Syros, Greece. See text for discussion. (After Breeding *et al.*, 2004; published by permission of the Geological Society of America.) (**B**) Metamorphosed metabasalt block (above hammer) in serpentinite matrx. (**C**) Metabasaltic block with glaucophane-rich reaction rind (dark). Rind has broken away from lighter-colored area in front of block by pocket knife, exposing underlying omphacite-rich eclogite.

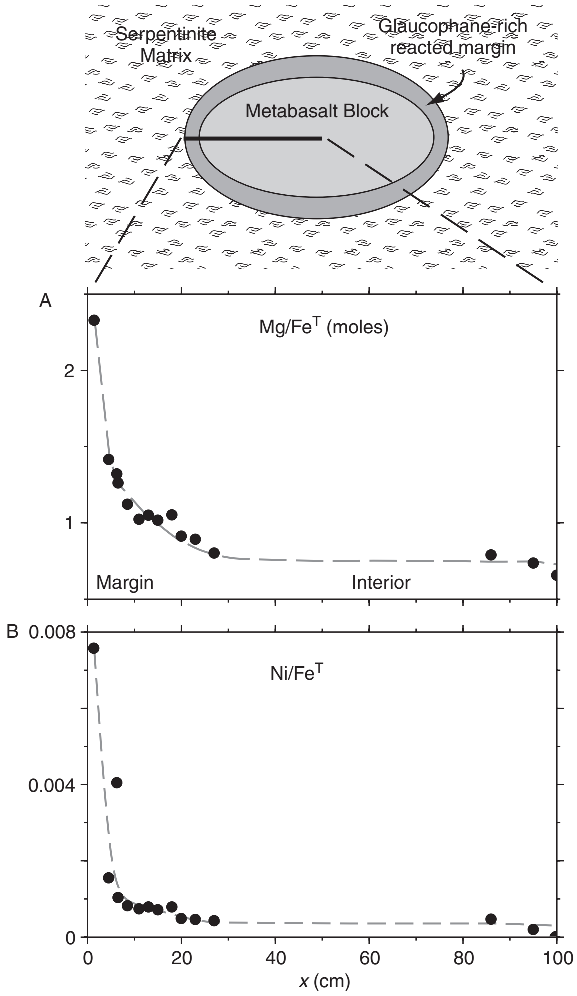


Fig. 21.29 Chemical alteration of metabasaltic mélange block due to infiltration of fluid from surrounding serpentinite matrix. Infiltration and reaction produced a nearly monomineralic glaucophane-rich margin or "rind." Each dot represents an analyzed sample of rock. Fe^T = total iron; Syros, Greece (Ague, unpublished data). (**Top**) Increase in Mg/Fe^T ratio (molar) in reaction rind. (**Bottom**) Increase in Ni content of rind relative to Fe^T (mass ratio).

ductile deformation is thought to predominate. However, in the Los Angeles, California (USA) area alone, nearly 1100 small to moderately large seismic events were detected at depths of 20 to 35 km between 1980 and 1994 (Ague, 1995). Some 110 000 magnitude 1 to 4 earthquakes were recorded in the Taiwan orogenic belt since 1991, many extending to depths of 35–60 km along major fault zones (Fig. 21.32; Carena *et al.*, 2002). In addition to small seismic events, many larger, damaging earthquakes and their aftershocks occur in the continental crust over the 15 to 25 km depth range (e.g. Ague *et al.*, 1998), and major seismicity continues to even greater depths in subduction zones (e.g. Hacker *et al.*, 2003). These observations demonstrate that brittle behavior can extend well into the realm of regional metamorphism and fluid flow.

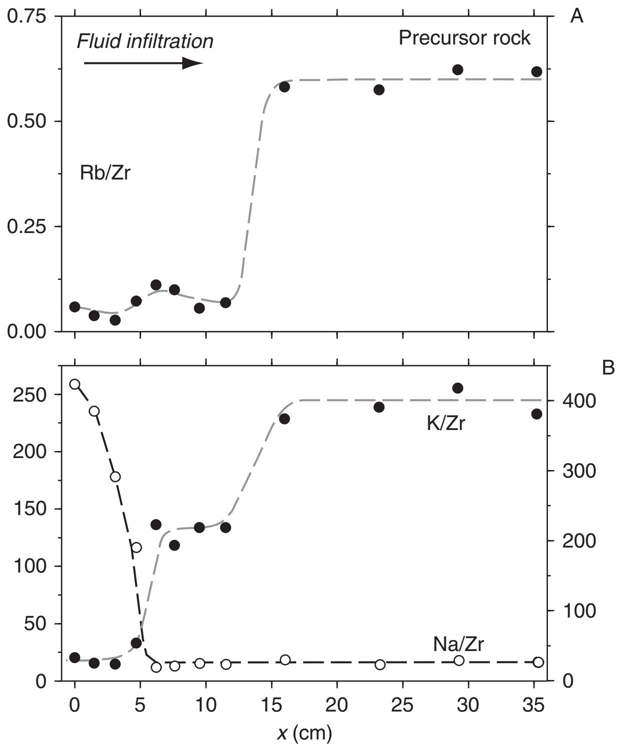


Fig. 21.30 Loss of rubidium (**Top**), and loss of potassium and gain of Na (**Bottom**) for micaceous metacarbonate rock at margin of mélange zone, Syros, Greece (Ague, unpublished data). Fluids from serpentinite infiltrated from the left and stripped K, Rb, and other elements including Ba and Cs (not shown), whereas Na was added. Elements plotted as mass ratios relative to Zr, which was nearly immobile in these fluids (Section 21.13). K/Zr on left axis of part (B), Na/Zr on right.

Veins are common in all types of metamorphic settings. For example, they frequently constitute 10 to 30 volume percent of the rock mass in Barrovian sequences (Ague, 1994b; Masters and Ague, 2005). Despite their wide occurrence, it is somewhat surprising that they are usually absent from rock and thin section sets studied in petrology laboratory classes. Part of the reason for this absence is that veins present sampling challenges. Samples commonly break along veins, making it tempting to discard the vein part. The problem here is that this loss of spatial context makes it impossible to evaluate any effects the vein may have had on the adjacent rock sample's evolution. Furthermore, veins and their alteration effects can be tens of centimeter to meter scale or even larger, and such large samples are difficult to obtain and carry out of the field.

Veins are often found occupying zones of weakness in rocks, such as along lithologic contacts or in pull-aparts between "boudins" (French for "sausage") (Fig. 21.33(A), (B)), but they are not restricted to these settings. The *boudinage structure* is formed when a competent layer embedded in less competent rock necks down as a result of being stretched parallel to its length. The flow of the less competent rock may actually segment the competent layer into separate boudins.

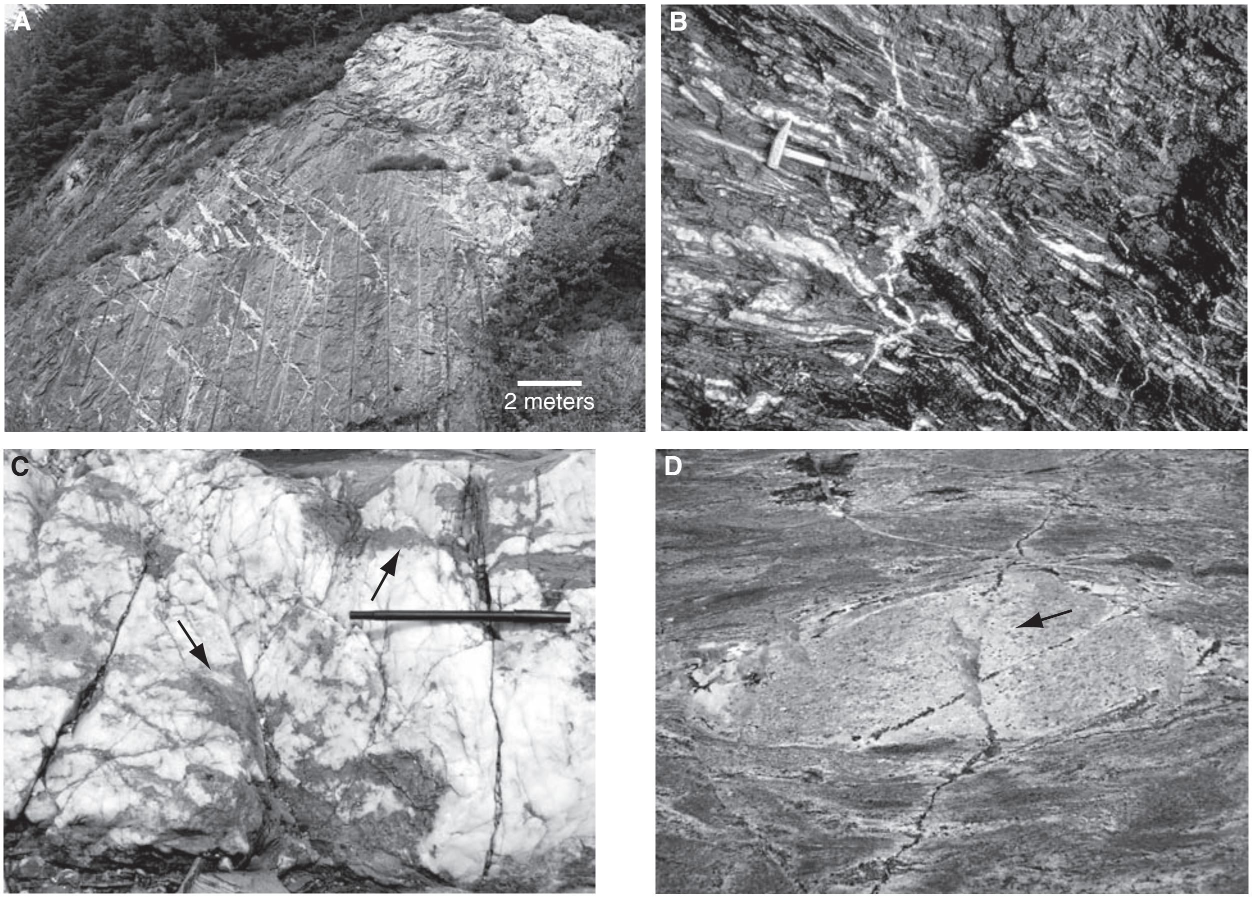


Fig. 21.31 Veins and vein textures. (**A**) Photo of widespread veining in greenschist facies metasediments of the Barrovian sequence, Scotland. Area in top right is massive quartz vein, and contains inclusions of wall rock. (**B**) Photo of veins in greenschist facies metapelitic rocks, Otago Schist, New Zealand. (**C**) Close-up of quartz vein, Barrovian zones, southeastern Scotland. Vein margins run along top and bottom of photo. Inclusions of wall rock dislodged from vein margins denoted with arrows. These inclusions indicate that the vein did not open all at once. Rather, it opened successively in stages, many of which dislodged wall rock material that was then incorporated into the widening vein. Pen for scale is 15 cm long. (**D**) Ductile and brittle deformation, cut slab of metaconglomerate, Brazil. The granitic clast was flattened by ductile deformation processes. The clast was then cut by veins. Note the light-colored, feldspar-rich alteration selvage areas on the margin of the vein cutting up through the central part of the clast (arrow). This chemical alteration was caused by fluids moving through the vein. In the schist matrix, the vein is slightly folded; thus, it cut through the rock after a significant amount of cleavage development and clast deformation had taken place, but before the deformation was completely finished. Field of view is 0.56 m wide.

To get an idea of the impact of fractures on permeability, consider an initially unfractured rock mass having a very low permeability, *k*, of 10^{-22} m^2. Suppose that parallel, 1 μm wide fractures form, and that they comprise a total interconnected fracture porosity of 0.001. Using reasoning analogous to that employed to derive Eq. (21.4), we compare the law for flow through a planar fracture (Eq. 3.16) with Darcy's law (Eq. 21.1) to find that the permeability in a fractured rock mass can be estimated using $k = (2w)^2\,\phi_F/12$, where $2w$ is the fracture width and ϕ_F is the porosity due to fractures. The permeability in our case after fracturing is thus $\sim 8\times10^{-17}$ m^2, an enormous increase over the unfractured state. This example demonstrates that even tiny cracks can have a major impact on permeability. Obviously, larger fractures will have an even greater effect. These simple results illustrate that fluid flow will tend to concentrate in fractures because they are high permeability conduits.

Advective fluxes through fractures can be large due to elevated permeabilities, but fluids also exchange chemically and isotopically with adjacent rocks (*wall rocks*) by diffusion and mechanical dispersion transverse to the flow. Such mass transfer produces chemically and isotopically altered *selvages* (or *selvedges*) along the vein margins that range in thickness from fractions of a millimeter to several meters (Fig. 21.33(C); Evans and Trommsdorff, 1974; Tracy *et al*., 1983; Abart, 1995). An example of fluid infiltration and selvage formation along cracks can be seen in Figure 4.45, and other examples are provided in Figures 21.31(D) and 21.34. The mass transfer can reflect the processes we have described earlier, including fluid flow along regional temperature and pressure gradients, and the propagation of geochemical signatures from one rock layer to another. In this way, veins and their alteration selvages can provide important insights into the directions of fluid motion, time-integrated fluid fluxes, and the sources of infiltrating fluids.

Note that because of the permeability contrasts between fractured and unfractured rock, fluid fluxes will in general be much larger in cracks than in the rock matrix (Ague, 2007).

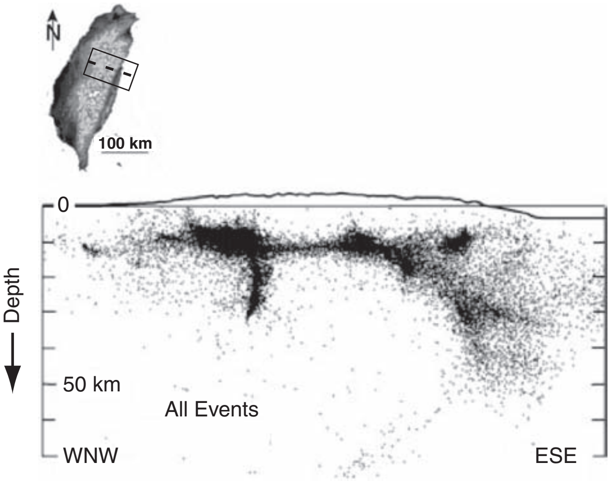


Fig. 21.32 Locations of over 110 000 magnitude 1–4 earthquakes beneath Taiwan (modified after Carena *et al.*, 2002). Earthquakes occurred under the box shown on inset map, and their locations were projected onto a plane parallel to the dashed line in box. Seismic events associated with the main detachment fault between the Taiwan accretionary prism and the Eurasian plate, and with convergence of the Philippine Sea plate. No vertical exaggeration. (Published by permission of the Geological Society of America.)

As a result of this spatial localization, chemical and isotopic alteration associated with fracture flow will be localized around veins, and surrounding parts of the rock mass distal to the veins may be little affected.

Veins can form before, during, or after peak metamorphism. Early veins are commonly overprinted by peak metamorphic mineral growth. Synmetamorphic veins can truncate and fracture metamorphic porphyroblasts, but the porphyroblasts themselves are not retrograded. Furthermore, the growth of large crystals of metamorphic index minerals in and around veins suggests the presence of fluids during crystallization. Later, retrograde veins can cut porphyroblasts and are commonly surrounded by zones of retrograde hydration, often characterized by the widespread growth of greenschist facies minerals like chlorite. Of course, age dating of rocks and minerals (Chapter 13) in and around veins can illuminate the absolute timing of fracturing and fluid flow. The timescales of flow through fractures probably range widely depending on such factors as fluid flux and tectonic setting. Mounting evidence based mainly on the partial resetting of oxygen isotope systematics suggests that the duration of flow in some veins was as little as ~10^4 years, or even less

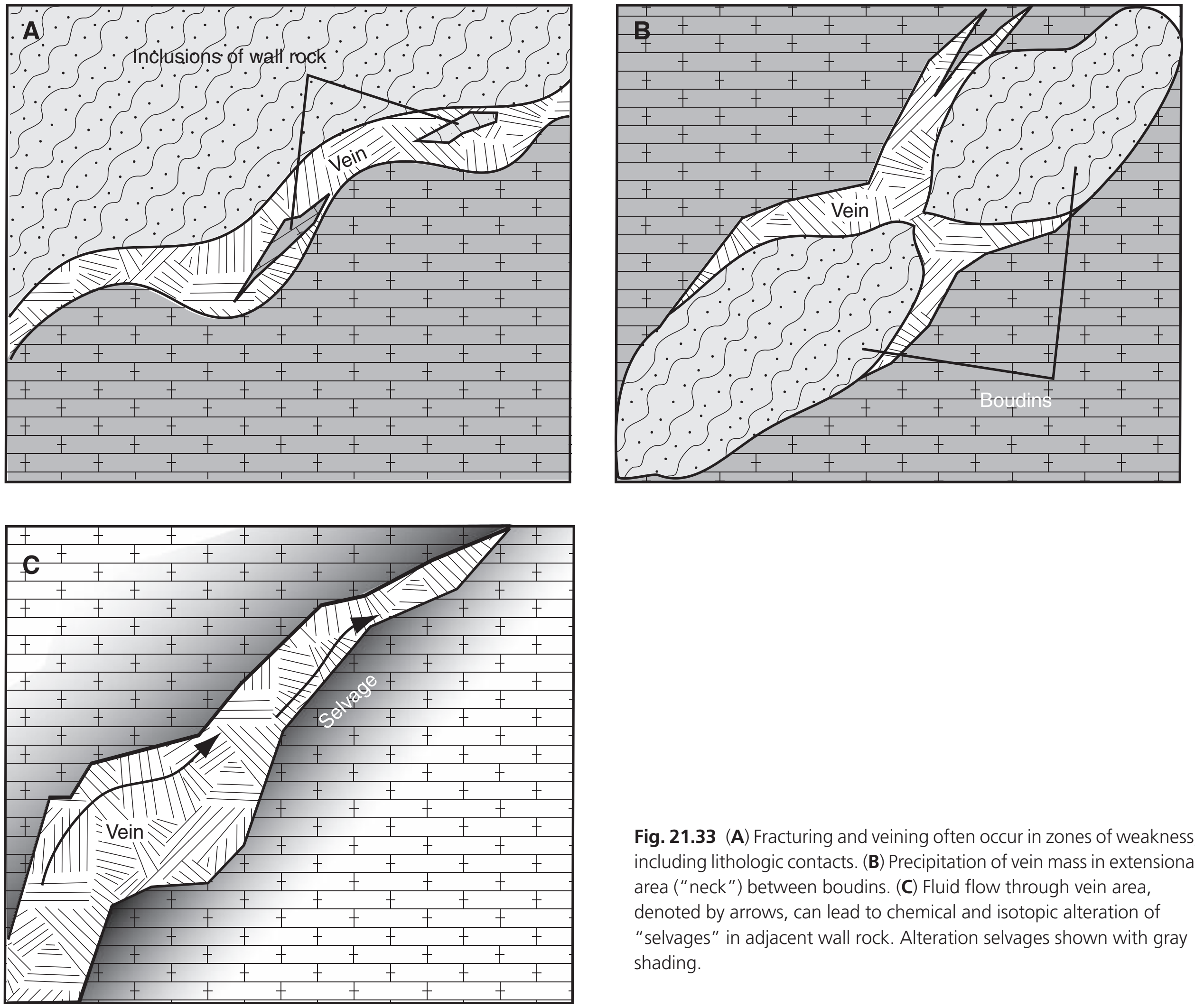


Fig. 21.33 (**A**) Fracturing and veining often occur in zones of weakness, including lithologic contacts. (**B**) Precipitation of vein mass in extensional area ("neck") between boudins. (**C**) Fluid flow through vein area, denoted by arrows, can lead to chemical and isotopic alteration of "selvages" in adjacent wall rock. Alteration selvages shown with gray shading.

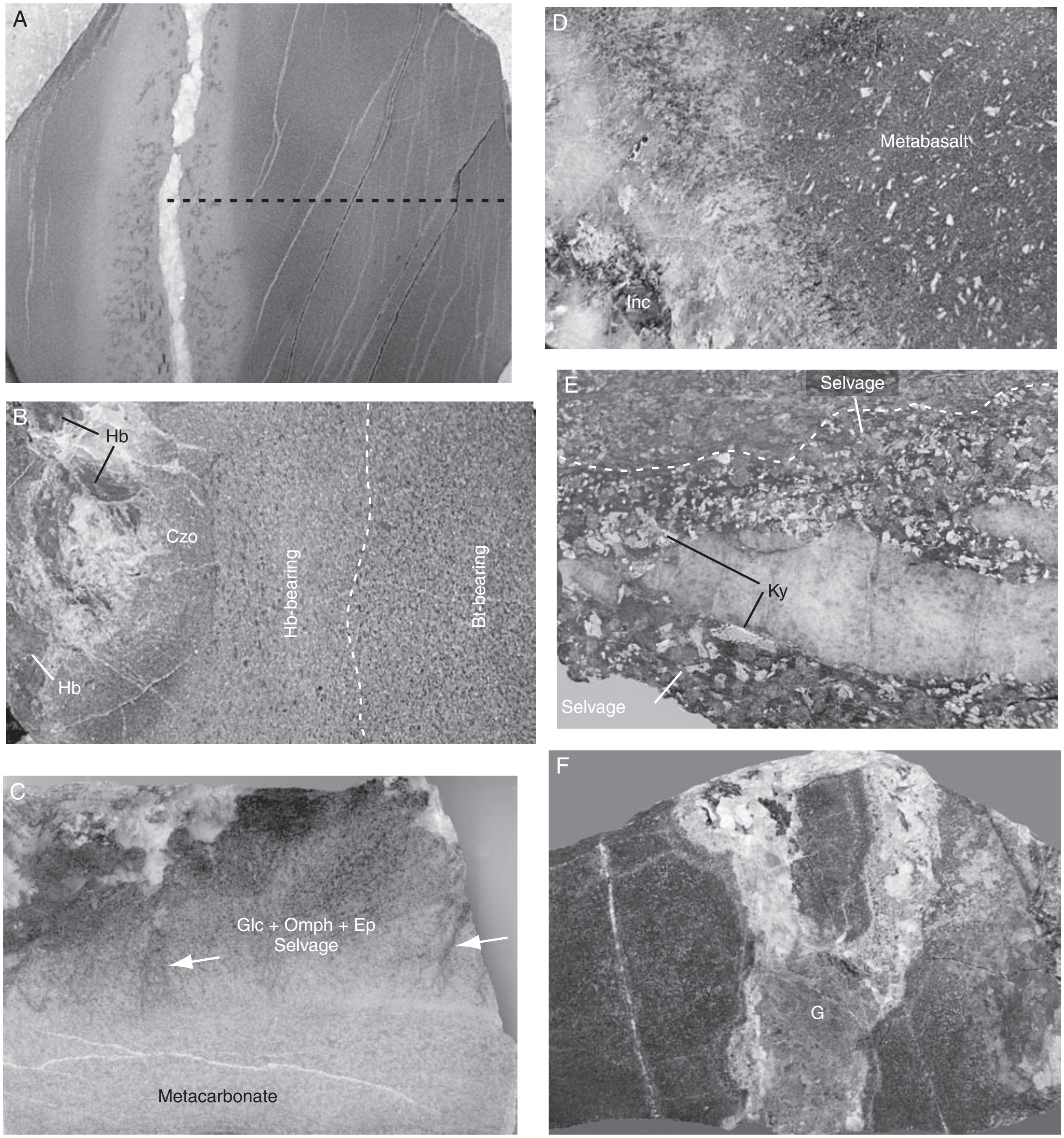


Fig. 21.34 Photographs of cut rock slabs illustrating vein–selvage relations. (**A**) Vein and alteration selvage cutting Ankerite–albite zone metacarbonate rock, south-central Connecticut. Dashed line denotes geochemical profiles shown in Figure 21.35 and discussed in Section 21.13. Width of field is 12.5 cm. (**B**) Vein and selvage cutting Amphibole zone metacarbonate rock, south-central Connecticut. Vein, at left, contains coarse hornblende (Hb) and clinozoisite. Moving to the right, there is a clinozoisite-rich zone directly adjacent to the vein (Czo), then a zone where biotite has reacted to form amphibole, then finally the less reacted biotite-bearing rock (Bt). Infiltrating fluid had lower X_{CO_2} than that in equilibrium with biotite-bearing assemblage, and therefore drove decarbonation producing amphibole (reaction (21.87)). Width of field is 9.5 cm. (**C**) Glaucophane (Glc) and omphacite (Omph) in vein and selvage along top half of photo. Note "fingers" of glaucophane-rich alteration extending away from vein zone (arrows). Metacarbonate rock, Cycladic subduction complex, Tinos, Greece. Width of field is 17 cm. (**D**) Greenschist facies metabasalt which still contains relict igneous plagioclase phenocrysts at right, cut by vein and plagioclase + chlorite selvage at left. Vein contains inclusion of selvage (Inc) dislodged from the vein margin. Width of field is 12 cm. (**E**) Amphibolite facies vein and kyanite + staurolite + biotite-rich selvage cutting metapelitic rock, south-central Connecticut. Kyanite (Ky) appears as light gray crystals in photo. Note large grain size of kyanite along vein margin. Width of field is 11 cm. (**F**) Amphibolite facies quartz–garnet–kyanite–biotite–carbonate vein cutting granulite facies plagioclase + pyroxene mafic gneiss, northeastern Connecticut. Note large garnet mass growing in vein (G). Width of field is 15 cm.

(e.g. Palin, 1992; Young and Rumble, 1993; Abart, 1995; van Haren *et al*., 1996). In these cases, the *total* timescale of vein activity could have been longer if the fluid flow occurred in multiple pulses separated by quiescent periods.

The mass precipitated in veins is derived from through-going fluids, the local wall rocks, or some combination. Study of amphibolite facies veins in southern Connecticut (Ague, 1994b) indicates that about 30 volume percent of the mass was derived from external fluids, and about 70 volume percent from the local wall rocks. The volume of externally derived quartz would have required an average time-integrated fluid flux, q_{TI}, of about $3 \times 10^5\,\mathrm{m}^3\,\mathrm{m}^{-2}$ (Eq. (21.67)). This result was later tested by examination of geochemical fronts (Ague, 2003a). It turns out that where large amounts of fluid flowed along cracks from metapelitic rocks into metacarbonate layers, Al was transported from metapelite to metacarbonate (Fig. 21.34(B)). K_v can be estimated using fluid and rock concentration values for metapelite. Taking an Al concentration of 10^{-3} molal (Manning, 2001, 2007), and a typical Al_2O_3 concentration in the rock of 25 wt % (Ague, 2003a), the resulting K_v is about 7×10^{-5}. The distance of Al metasomatism along the cracks varies but is generally in the range of 1 to 10 m. The corresponding q_{TI} values are 10^4 to $10^5\,\mathrm{m}^3\,\mathrm{m}^{-2}$ (Eq. (21.60)). These values are compatible with the estimate obtained independently from quartz vein precipitation, and confirm that some veins were conduits for very large fluid fluxes.

The sequence of vein-forming events in these Connecticut rocks is envisioned as follows. First, fracturing opened up channel-ways through which regionally migrating fluids moved. Some quartz was precipitated as the fluids ascended and cooled. In addition, the migrating fluids interacted chemically and isotopically with the surrounding wall rocks to produce alteration selvages (Fig. 21.34(A), (B), (E)). The exchange was probably dominated by diffusion, but some flow through the wall rock margins may have occurred as well. Silica also diffused down pressure gradients to the fractures in order to help seal them. In this way, both externally derived and internally derived silica was deposited in the fractures.

This "crack–flow–seal" sequence of events was repeated many times in order to produce a typical amphibolite facies vein. Selvages in metapelitic rocks are depleted in silica because it was repeatedly transferred to the adjacent veins in order to seal them. Mass transfer of alkalis in selvages helped the aluminous index minerals staurolite and kyanite to grow at the expense of alkali-rich minerals like muscovite (Fig. 21.34(E)). The infiltration of fluids derived from metapelitic or igneous rocks drove decarbonation reactions around veins cutting metacarbonate layers (Figs. 21.27, 21.34(B)). The presence of fluids also facilitated diffusion and the growth of large crystals in and around the veins (Fig. 21.34(E)). Some vein kyanite crystals reached ~5 cm in length, as opposed to the grain sizes of all minerals outside selvage margins which are usually <0.5 cm. We can conclude that fluid flow through veins may exert considerable controls on both rock composition and grain size in metamorphic belts.

21.13 DETERMINING CHANGES IN COMPOSITION AND VOLUME USING MASS BALANCE

This chapter has been concerned with mass transfer and changes in rock composition. The expression for the time-integrated fluid flux based on Na metasomatism, for example, requires that we know how much Na the rock gained or lost (Eq. (21.71)). But we have not yet addressed the question of how the magnitudes of such changes can be determined in real rocks. The equations needed to assess metasomatic changes are straightforward to derive and are firmly based on mass balance principles. Mass balance, in turn, follows directly from the first law of thermodynamics. The following treatment is based largely on the work of Gresens (1967), Grant (1986), Brimhall *et al*. (1988), Ague (1994a), and Ague and van Haren (1996); the reader is referred to these papers for further details.

In order to determine compositional changes, one must first have a chemical analysis of the altered rock, as well as an analysis of the precursor rock from which the altered rock was derived. In practice, identification of the precursor can be a difficult task, particularly if metasomatism has been extensive. Misidentification of precursors renders a mass balance analysis invalid, so the utmost care must be exercised at this stage. One setting in which precursor uncertainties are minimized is alteration selvages adjacent to veins that cut across a metasedimentary layer. Relatively little original sedimentary variation would be expected within a single layer, although graded bedding and other sedimentary features can introduce compositional heterogeneity and must be identified prior to geochemical sampling. Sedimentary variations tend to be most pronounced across layers (top to bottom) rather than along them, so profiles that stay within a single horizon parallel to layering generally have the smallest precursor uncertainties. For such profiles, the amount of metasomatism can be monitored progressively from little-altered wall rock precursor outside the selvage margins, to more strongly altered rock in vein selvages.

Figure 21.34(A) shows a synmetamorphic quartz–albite–calcite–ankerite vein cutting across a greenschist facies metacarbonate layer in the Wepawaug Schist, Connecticut (Ague, 2003a). In the alteration selvage adjacent to the vein, muscovite was destroyed and albite precipitated. Finely disseminated organic matter and pyrite in the matrix organized into larger "clumps" in the selvages, lending a bleached and spotted appearance to the selvage areas. Xenotime, a phosphate mineral that is normally rich in rare earth elements (particularly the heavy rare earths) precipitated in and around the vein. The vein represents a channelway for fluids that migrated through the metamorphic sequence, and the selvage records chemical and isotopic alteration that occurred due to reactions between these fluids and their wall rocks.

Geochemical profiles across the wall rock and vein are shown in Figure 21.35. These data were obtained using X-ray fluorescence and inductively coupled mass spectrometry techniques carried out on aliquots of rock cut from the sample at regular intervals. Take, for example, the plot illustrating the

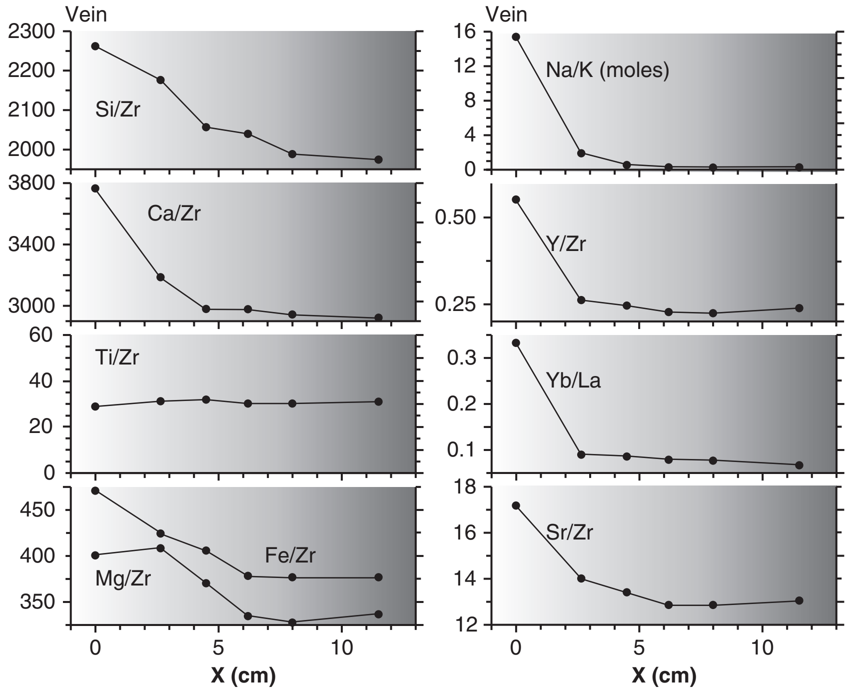


Fig. 21.35 Geochemical profiles across traverse shown in Figure 21.34(A). See text for discussion. (Modified from Ague, 2003a; published by permission of the *American Journal of Science.*)

increase in the Na/K ratio in the alteration selvage and vein area. While this may seem like a simple relationship, there are actually six different ways to interpret it! First, the precursor rock that the altered selvages were derived from could have been different from the wall rock outside the selvages. This kind of precursor uncertainty can be ruled out given that we are examining a profile across a vein that cuts a single, homogeneous metasedimentary layer. Consequently, we can be confident that the Na/K ratio did in fact increase due to fluid infiltration and chemical reaction. This, in and of itself, is very valuable information. But we can go further and investigate what kind of mass changes caused the ratio to increase. A second possibility, then, for interpreting the profile is that Na was added to the rock, and K was lost. At first glance, this would seem to be the primary possibility, but there are several more to consider. Thus, a third possibility is that Na was added, but K was immobile and did not leave the rock. Fourth, Na may have been immobile, but K was lost. Fifth, Na and K may both have been added to the rock, but more Na was added than K. Finally, Na and K may both have been lost from the rock, but more K was lost than Na.

To assess these possibilities, we need to define a *reference frame*. One might conclude that all one has to do is plot up the weight percents of Na and K by themselves to determine what mass changes occurred. Unfortunately, it turns out that weight percents, in and of themselves, are not suitable for assessing such mass changes because they are, by definition, normalized to sum to 100 weight percent. Consider the following hypothetical example from Ague and van Haren (1996). A quartzite contains quartz, rutile, and zircon, and has a bulk composition of 99.2 wt% SiO_2, 0.72 wt% TiO_2, and 0.08 wt% ZrO_2. Suppose that metasomatism caused by fluid flow in a direction of increasing temperature (Eq. (21.67)) dissolves a substantial amount of quartz from the rock, such that 100 kg of precursor loses 20 kg of SiO_2 and the mass of SiO_2 is now 79.2 kg. Ti and Zr are assumed, for this example, to have remained inert. Normalization of the altered rock chemistry to weight percents yields: 99.0 wt% SiO_2, 0.90 wt% TiO_2, and 0.10 wt% ZrO_2. So even though 20 kg silica per 100 kg rock were lost, the weight percentage of SiO_2 has barely changed! Comparison of SiO_2 weight percents alone, therefore, would not reveal any substantial mass transfer.

Three different kinds of reference frame can be used to better assess mass changes. One is a constant volume reference frame, which assumes that the volume of the altered rock is the same as the precursor. A second is a constant rock mass reference frame, which stipulates that even though the chemical composition of the altered rock may have changed, the bulk rock mass didn't. Thus, 100 kg of precursor would yield 100 kg of altered rock. Mass transfer produces changes in rock composition and, hence, volume, so both of these reference frames are rather special cases. Pseudomorphs that preserve recognizable shapes of the replaced grains are sometimes taken as evidence for constant volume replacement. However, a potential pitfall here is that the pseudomorphs could have retained the general shape of the original crystals, but not the size (and thus volume) of the originals.

The third reference frame can treat all mass transfer scenarios, including constant mass or volume. It is based on identification of an immobile element or suite of elements. In nature, of course, no element is truly immobile. However, some have far lower concentrations in solution than others and can be used as good approximations for the geochemical reference frame. Elements most commonly used as reference species in this regard are the high field strength elements

Zr, Ti, Th, and the rare earth elements (REE). However, even these elements can be mobile under appropriate pressure, temperature, and fluid composition conditions (Bröcker and Enders, 2001). Particular caution should be exercised when working with rocks which have undergone partial melting, because most elements can be mobilized to some degree in silicate melts.

Once the precursor and reference frame are identified, it is straightforward to derive the necessary mass balance equations. For a reference species *i*, the basic statement of mass balance is

$$V^0 \rho^0 c_i^0 = V' \rho' c_i' \tag{21.90}$$

where V is rock volume, ρ is bulk rock density, c_i is the concentration of the reference species *i*, and the superscripts ^{0} and $'$ refer to precursor and altered rock, respectively. Equation (21.90) simply states that the mass of *i* in the altered rock and in the precursor are the same (because *i* is inert). The total rock mass is given by the product of volume and density. Thus, the change in total rock mass, T_{mass}, can be written as

$$T_{\text{mass}} = \frac{V'\rho' - V^0\rho^0}{V^0\rho^0} \tag{21.91}$$

From Eq. (21.90), we can write

$$V^0\rho^0 = V'\rho' \frac{c_i'}{c_i^0}$$

Substitution of this expression into (21.91) and rearranging yields (Ague, 1994a)

$$T_{\text{mass},i} = \frac{c_i^0}{c_i'} - 1 \tag{21.92}$$

where $T_{\text{mass},i}$ is the total rock mass change calculated based on reference species *i*. This value is a fraction; the percentage mass change can be found by multiplying by 100. The units of concentration, such as weight percent or parts per million, don't matter as long as they are the same for both numerator and denominator. If rock mass has been lost, the concentration of *i* in the altered rock will be greater than that in the precursor, and vice versa. If no rock mass change occurred, then the concentration ratio will be unity. Experience has shown that total rock mass changes in excess of ±35% are unusual in metamorphic rocks, unless they have undergone partial melting.

Volume changes can also be assessed based on Equation (21.90). The volume change (strain, ε) can be defined as

$$\varepsilon = \frac{V' - V^0}{V^0} \tag{21.93}$$

Substitution of (21.91) into this expression yields, after some rearrangement

$$\varepsilon_i = \left(\frac{c_i^0}{c_i'}\right)\left(\frac{\rho^0}{\rho'}\right) - 1 \tag{21.94}$$

The bulk density of the rock, ρ, is reduced as the volume of void space increases. Thus, bulk density and porosity are related by

$$\phi = 1 - \frac{\rho}{\rho_{\text{g}}} \tag{21.95}$$

where ρ_{g} is the density of the mineral grains themselves, and porosity is understood to be, in this case, the volume of void space (not fluid) per unit volume of rock. Substitution for the density terms in (21.94) gives

$$\varepsilon_i = \left(\frac{c_i^0}{c_i'}\right)\left(\frac{\rho_{\text{g}}^0}{\rho_{\text{g}}'}\right)\left(\frac{1-\phi^0}{1-\phi'}\right) - 1 \tag{21.96}$$

This expression reveals that volume change is the product of three factors. The concentration ratio term accounts for volume losses or gains resulting from mass losses or gains. The second ratio accounts for changes in volume due to changes in density. For example, the density of calcite is 2720 kg m^{-3}, and that of aragonite is 2950 kg m^{-3}. Therefore, transformation of calcite to aragonite produces a nearly 8% volume decrease. The third term describes volume changes due to porosity changes. If the precursor and altered rocks had small porosities, or if they had similar porosities, then this term is approximately unity and can be ignored.

In addition to the total change in rock mass, we can also define the change in mass for any individual mobile species *j*

$$\tau^j = \frac{V'\rho' c_j' - V^0\rho^0 c_j^0}{V^0\rho^0 c_j^0} \tag{21.96}$$

where τ^j is the fractional change in mass for species *j*. To eliminate the volume terms, we make use of the fact that

$$\frac{V'}{V^0} = \varepsilon_i + 1 \tag{21.97}$$

so as to obtain

$$\tau_i^j = \left(\frac{c_j'}{c_j^0}\right)\left(\frac{\rho'}{\rho^0}\right)(\varepsilon_i + 1) - 1 \tag{21.98}$$

Equation (21.98) is very useful because it relates changes in composition, density, and volume. It can be simplified, however, by substituting (21.94) for the volume strain to yield

$$\tau_i^j = \left(\frac{c_i^0}{c_i'}\right)\left(\frac{c_j'}{c_j^0}\right) - 1 \tag{21.99}$$

Taken together, Equations (21.92), (21.96), and (21.99) provide quantitative expressions for evaluation of rock mass change, volume strain, and changes in the masses of individual elements, provided that the precursor and geochemical reference frame are identified correctly. Note that all these expressions are cast in terms of concentration ratios. The use of ratios cancels out the 100% normalization inherent in all weight percent data. If we return to our quartzite example, using Zr or Ti as the reference frame we obtain the correct

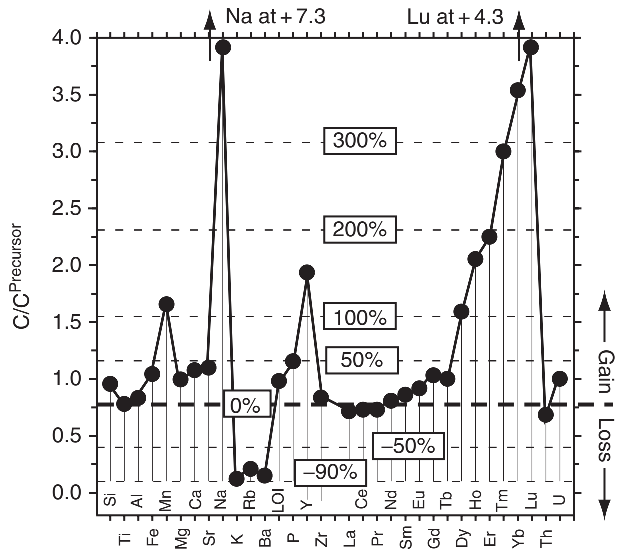


Fig. 21.36 Concentration ratio (CR) diagram for most altered rock along profile in Figure 21.34(A) relative to least altered rock. Geochemical reference frame denoted by thick dashed line. Mass changes include addition of Na, P, Y, and heavy REE like Lu, and loss of K, Rb, and Ba. See text for discussion. (From Ague, 2003a; published by permission of the *American Journal of Science*.)

results that the rock lost 20% total mass, and 20.2% SiO_2 mass. It is impossible to calculate the correct values without using the appropriate concentration ratios.

A convenient graphical technique for evaluating mass changes is simply to plot altered/precursor concentration ratios on the y-axis for the elements of interest arranged in any convenient order on the x-axis (Fig. 21.36). If some of the mass changes are very large, it is better to plot the logarithms of the mass ratios. The concentration ratio (or CR) diagram in Figure 21.36 depicts veined and altered selvage rock from Figure 21.34(A), relative to little altered precursor 11.5 cm from the vein (Ague, 2003a). The first thing to do is inspect the behavior of the relatively "inert" elements in order to determine the geochemical reference frame. It is clear that the concentration ratios for Al, Ti, Zr, Th, and the light REE are all about the same, averaging 0.83. Note that this ratio is the inverse of that in the total mass change expression (Eq. (21.92)). A value less than one on a CR plot indicates that the concentrations of the low solubility elements were diluted due to the addition of other, mobile elements. Conversely, a value greater than one indicates bulk mass loss. The value will be equal to one only for the special case of no bulk mass loss or gain. We can draw a line through 0.83 on the CR graph to establish the reference frame. From Eq. (21.92), this translates into a volume gain of about 20%. This estimate makes sense qualitatively, as mass was precipitated in cracks in the selvages and the main vein itself during extension of the rock.

Elements added to the rock plot above the reference line on Figure 21.36, whereas elements that were lost plot below it. Contours of constant mass gain or loss, calculated using Eq. (21.99), appear as horizontal lines. The gains of Na, heavy REE, Y, and Mn are immediately obvious, as are the losses of K, Rb, and Ba. Gains of Si, Mg, Ca, Sr, P, and volatiles are also indicated. The addition of Na and losses of K, Rb, and Ba correspond to the observed growth of albite at the expense of muscovite. The gains of Y, P, and heavy REE are primarily the result of xenotime precipitation during alteration. Some calcite and ankerite precipitated in the vein and selvage, accounting for the gains in Ca, Mn, Mg, Fe, Sr and potentially some of the heavy REE. The gain of silica corresponds to the precipitation of quartz in and around the vein. A critical point to emphasize is that the mass balance analysis is consistent with observed modal changes in the rock.

It is useful to return to the geochemical profiles of Figure 21.35. Most elements are plotted as ratios with respect to Zr, because Zr was relatively immobile and thus provides a geochemical reference frame. The ratios with Zr in the denominator clearly confirm the mass changes deduced from the CR diagram. Furthermore, the CR diagram tells us that the increases in Na/K and Yb/La along the profiles were due to gain of Na and loss of K, and to gain of Yb but no gain of La, respectively. The Ti/Zr ratio is unchanged across the profile. This means that Ti and Zr behaved the same way geochemically during alteration. From the CR diagram, in fact, it is reasonable to conclude that they were essentially immobile. Note that even if the geochemical reference frame cannot be identified, profile plots like Figure 21.35 still reveal the fact that the rock composition was altered, and show quantitatively how specific element ratios changed as a result of metasomatism.

In some cases, spatial profiles like those of Figure 21.35 are unavailable, but a suite of altered rocks that can be compared to a suite of precursors is. One may wish to compare the "average" altered rock with the "average" precursor, in order to determine the mass changes. It turns out, however, that averaging and statistical analysis of compositional data has many pitfalls; the reader is referred to Aitchison (1986), Ague (1994a), and Ague and van Haren (1996) for detailed discussion. One approach for assessing an average τ_i^j value is based on the geometric (logarithmic) mean concentration ratios (Ague, 1994a). Taking logarithms helps correct for the fact that ratio statistics are commonly non-Gaussian. First recast Eq. (21.99) as

$$\tau_i^j = \left(\frac{c_i^0}{c_j^0}\right)\left(\frac{c_j'}{c_i'}\right) - 1$$

and, by taking logarithms

$$\ln\left(\tau_i^j + 1\right) = \ln\left(\frac{c_i^0}{c_j^0}\right) + \ln\left(\frac{c_j'}{c_i'}\right)$$

Then take the mean of the *logarithms* of each concentration ratio

$$\text{Mean precursor log ratio} = \text{ML}^0 = \frac{1}{N^0}\sum_{n=1}^{N^0}\ln\left(\frac{c_{i,n}^0}{c_{j,n}^0}\right)$$

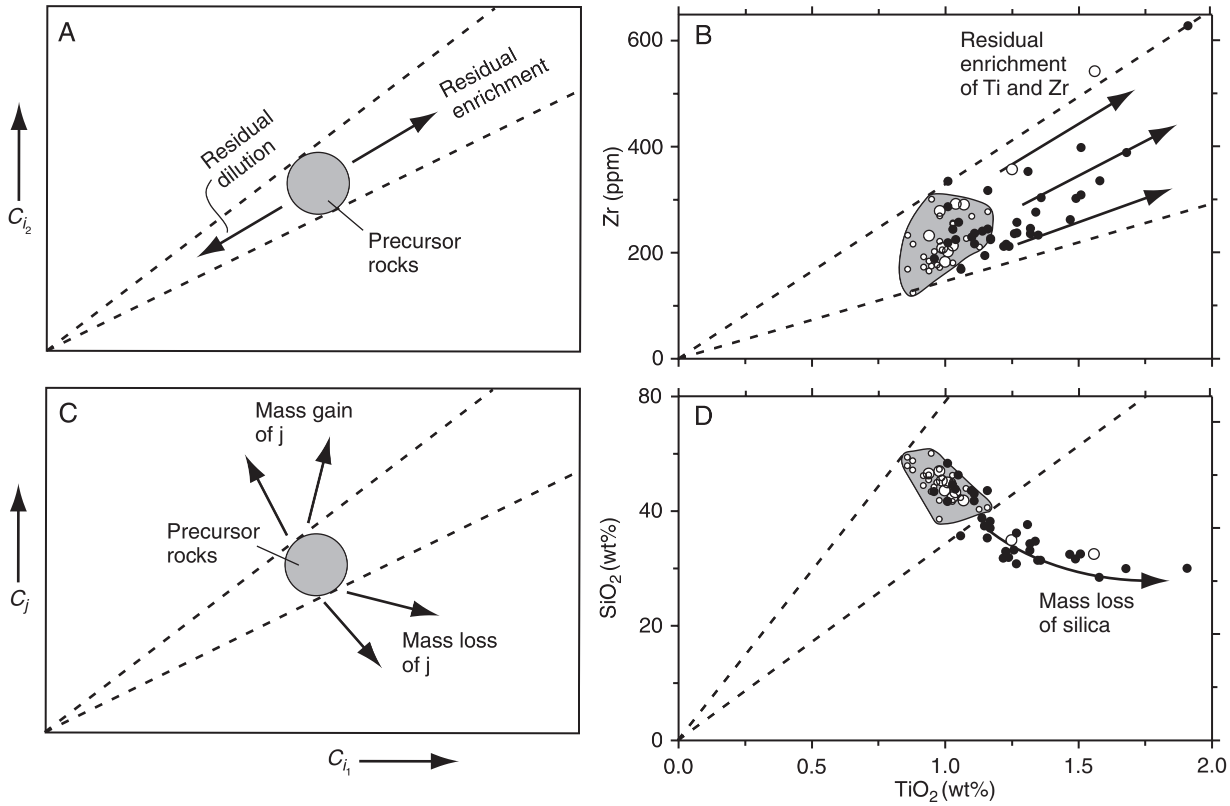


Fig. 21.37 Wedge diagrams for evaluation of compositional changes for suites of rock samples (Ague, 1994a). (**A**) Illustration of concentration systematics for two reference species i_1 and i_2. Rock mass loss produces residual enrichment, whereas rock mass gain produces residual dilution. (**B**) Residual enrichment of Ti and Zr for upper greenschist and amphibolite facies rocks of the Wepawaug Schist (Ague, 1994a). Small open circles defining gray-shaded field are low-grade greenschist facies precursor rocks. Upper greenschist facies rocks with macroscopic garnets shown with large open circles, and amphibolite facies rocks (Staurolite–kyanite zone) shown with filled circles. Rock mass loss due mostly to local silica loss and, to a lesser degree, loss of volatiles and other elements. See part (D) and text for further discussion. (**C**) Mass loss or gain of mobile element *j* will cause concentrations to plot outside the residual wedge area. (**D**) Silica loss from rocks plotted in part (B). The silica moved locally from the schists into widespread upper greenschist and amphibolite facies veins (Ague, 1994a, b).

$$\text{Mean altered log ratio} = \text{ML}' = \frac{1}{N'}\sum_{n=1}^{N'} \ln\left(\frac{c'_{j,n}}{c'_{i,n}}\right)$$

in which n denotes a particular sample and N^0 and N' are the total numbers of precursor and altered samples, respectively. The average, or "most probable" mass change value and the associated 2σ uncertainty range are computed from

$$\hat{\tau}_i^j = \exp\left(\text{ML}^0 + \text{ML}' \pm 2\sigma\right) - 1 \qquad (21.100)$$

where $\hat{\tau}_i^j$ denotes the mean mass change value and σ is a measure of uncertainty (the standard error). Note that setting $\sigma = 0$ yields the average mass change value. An estimate for σ can be made by summing the standard errors on ML^0 and ML$'$ in quadrature

$$\sigma = \left[\left(\sigma_{\text{ML}^0}\right)^2 + \left(\sigma_{\text{ML}'}\right)^2\right]^{1/2}$$

where the standard errors are simply the sample standard deviations on ML^0 and ML$'$ divided by $\sqrt{N^0}$ and $\sqrt{N'}$, respectively. Results for each element can be plotted on a diagram similar to Figure 21.36 that has mass changes and their uncertainties on the y-axis, instead of concentration ratios. Importantly, statistical analysis with only a few samples is subject to large uncertainties and often provides very unreliable results.

A simple graphical technique can help identify reference species and mobile elements when compositional variability exists (Ague, 1994a). Consider two reference species i_1 and i_2. A plot of a suite of inferred precursor rock compositions forms a data "cloud" on a graph of c_{i_1} versus c_{i_2}. If the concentrations of i_1 and i_2 increase due to overall rock mass loss (*residual enrichment*) or decrease due to rock mass gain (*residual dilution*), then the compositions of the altered rocks will lie in a wedge-shaped region that converges on the origin (Fig. 21.37(A), (B)). If a similar graph is made for two mobile elements, the trend in concentration for the altered rocks would fall within the wedge only in the highly improbable case that both elements underwent exactly the same amounts of mass gain or loss. Plotting the concentration of a mobile element *j* versus a reference element like i_1 for precursors will also yield a cloud of points (Fig. 21.37(C), (D)). If the concentrations of *j* and i_1 for the altered rocks are then plotted, they will fall outside the reference frame wedge, facilitating identification of mass loss or gain of *j* (Fig. 21.37(C), (D)). These methods are most powerful when the mass changes are large, and are less sensitive to more subtle changes.

Application of such “wedge diagrams” to Barrovian metapelitic rocks of south-central Connecticut is shown in Figure 21.37(B) and (D). The abundance of quartz veins in these rocks increases systematically with metamorphic grade (Ague, 1994b). The lowest-grade greenschist facies rocks have relatively few veins and form the precursor group. Overall mass loss for many upper greenschist and amphibolite facies samples is indicated by the residual enrichment trend for the reference species Ti and Zr (Fig. 21.37(B)). The mass loss was caused by loss of silica (Fig. 21.37(D)) and, to a lesser degree, volatiles and other elements. The silica moved locally at centimeter to decimeter scales from the rocks into adjacent veins, forming silica-depleted selvages. At the outcrop scale and larger, however, the rocks actually gained silica because it was also precipitated in veins by through-going fluids (Section 21.12.3; Ague, 1994b).

21.14 REGIONAL FLUID TRANSPORT THROUGH THE CRUST

Simple calculations indicate that devolatilization of typical metasedimentary sequences can produce large time-integrated fluid fluxes approaching 10^3 m^3 m^{-2} (Problem 16.2). Application of equations for estimation of time-integrated fluxes to real rocks, such as Eqs. (21.60) and (21.73), confirms that fluxes of this magnitude are achieved in nature (Fig. 21.38). These estimates are primarily for *pervasive* flow, which occurs at the grain scale through the porosity.

Even greater fluxes require that the flow is *focused* or *channelized*, that flow recirculates through a metamorphic terrain multiple times, or that there is a large source of fluid at depth within or below the metamorphic sequence. For example, fluxes channelized through individual veins or shear zones can easily reach magnitudes of ~10^4 m^3 m^{-2} or more (Fig. 21.38),

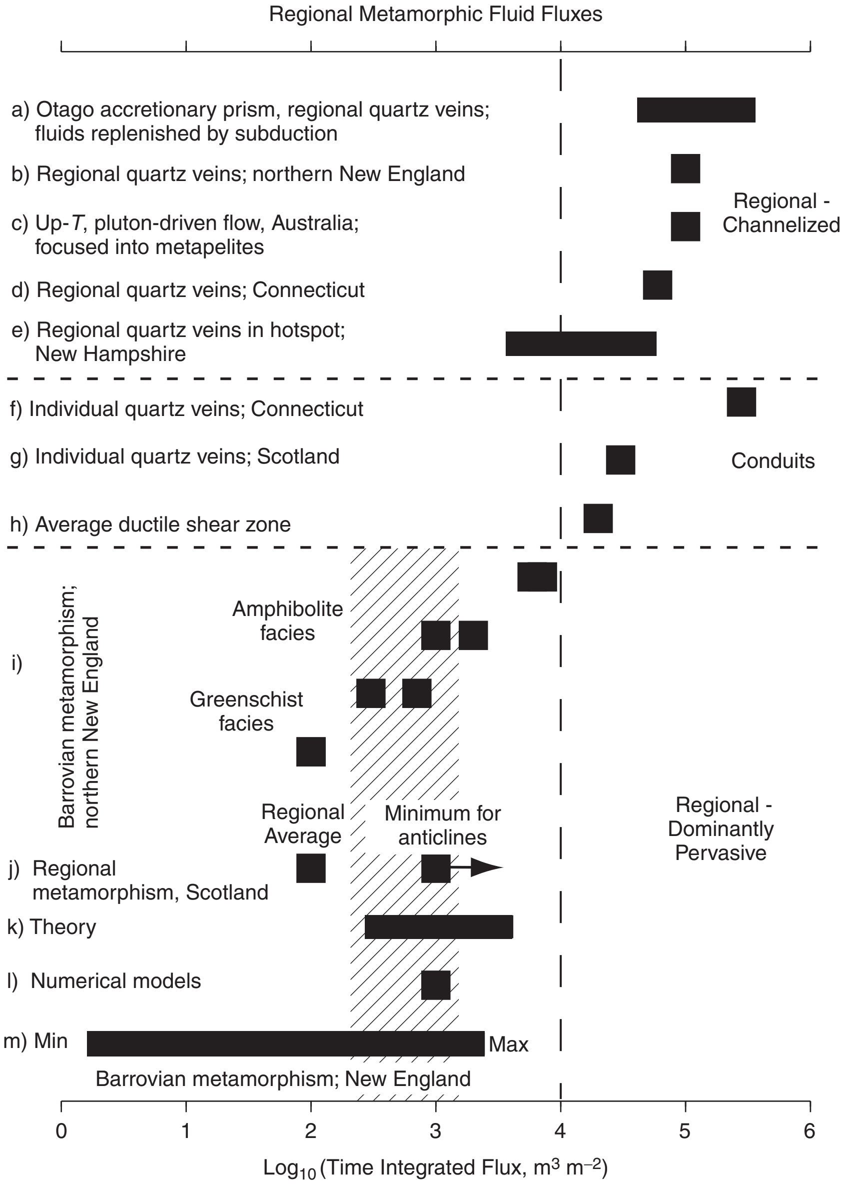


Fig. 21.38 Estimates of time-integrated fluid fluxes. The average regional, pervasive flow-dominated flux of $10^{2.7} \pm 0.5$ $m^3_{(fluid)}$ $m^{-2}_{(rock)}$ (2σ) denoted with diagonal ruled bar (computed using geometric mean; e.g. Ague, 1994a). References: (a) Breeding and Ague (2002); (b) Ferry (1992); (c) Ague (1994b); (d) Oliver *et al*. (1998); (e) Chamberlain and Rumble (1989). Flux range computed using average flux of 1.5×10 m^3 m^{-2} s^{-1} for 10^5 and 10^6 years. (f) Ague (1994b); (g) Ague (1997); (h) Dipple and Ferry (1992a); (i) Ferry (1992) and Léger and Ferry (1993); (j) Skelton *et al*. (1995); (k) Walther and Orville (1982) and Walther (1990). Range computed using total timescales of fluid flow of 10^6 and 10^7 years. (l) Hanson (1997); (m) Evans and Bickle (1999). (From Ague, 2003b; published by permission of Elsevier.)

large enough to change the major and trace element compositions of rocks (Eqs. (21.67), (21.71)). Focusing can also occur at the regional scale through fracture (vein) networks, as observed, for example, in the Otago Schist accretionary prism, New Zealand, or in highly veined rocks from New England, USA (Fig. 21.38).

The large regional fluxes in excess of ~10^4 m^3 m^{-2} probably resulted from the focusing of flow from large source regions into fracture networks or high permeability zones. Potential sources include thick sequences of devolatilizing metasediments, degassing arc magmas, or, in the case of the Otago Schist, fluids derived from prograde metamorphism of the downgoing slab and/or underplated sediments that were continually replenished by subduction (Breeding and Ague, 2002). The examples in Figure 21.38 are all for middle or lower crustal regional metamorphism, making convective recirculation unlikely. Convection could, however, play a major role at shallow crustal depths.

It is important to remember that H_2O- or CO_2-rich fluids are not abundant in all metamorphic environments. For example, Valley *et al*. (1990) found that the activities of water and CO_2 were very low during peak upper amphibolite and granulite facies metamorphism in the Adirondack Mountains, New York, USA. Many of these rocks were dehydrated during partial melting, as fluids dissolved directly into the silicate melt phase. Fluid flow was more extensive, however, during early, shallow level contact metamorphism around anorthosite bodies. In addition, small, retrograde veins are widespread and host CO_2-rich fluid inclusions.

We may conclude that metamorphism is far from being the passive reaction of rocks to a raised geothermal gradient. Because large volumes of fluid are produced during prograde heating, metamorphism must involve a complex interaction between the dissipation of thermal, chemical, and mechanical energy. Indeed, fluid fluxes may be large enough to even influence the thermal evolution of mountain belts, as we will explore in Chapter 22. Working out these physical–chemical relations and finding features in rocks that provide a record of the metamorphic history will keep petrologists busy for many years to come.

21.15 PROBLEMS

21.1 **(a)** What is the average flow velocity of fluid buoyantly rising through a 500-nm-wide planar fracture if the fluid's viscosity is 10^{-4} Pa s and its density is 1 Mg m^{-3}, and the density of the surrounding rock is 2.8 Mg m^{-3} (express your answer in m a^{-1})? Is flow laminar or turbulent?

(b) Repeat part (a) for fracture widths of 1 micrometer and 1 millimeter. Is flow laminar or turbulent for these cases?

21.2 Because gold is such a minor constituent of the Earth's crust, its concentration in ore deposits is evidence of enormous enrichment. The solubility of gold in hydrothermal waters is low (about 1 to 100 ppb), so that large volumes of fluid are needed to transport the quantities of gold in ore bodies. One way in which this may occur is through deep convection of meteoric water in the upper crust (Nesbitt, 1988). For convection to occur, however, rocks must be sufficiently permeable (Eq. (21.6)). In the upper crust, where rocks are brittle, this permeability might be provided by fractures. If the upper temperature limit for brittle behavior is 400 °C and the temperature gradient is 30 °C km^{-1}, calculate the minimum permeability necessary for convection. The water can be assumed to have a density of 1 Mg m^{-3}, a viscosity of 10^{-3} Pa s, a coefficient of thermal expansion of 10^{-3} K^{-1}, and a thermal diffusivity of 10^{-6} m^2 s^{-1}.

21.3 Derive the permeability relationship for idealized fractured rock discussed in Section 21.12.3: $k = (2w)^2\,\phi_F / 12$.

21.4 Figure 21.8 indicates the stability fields of muscovite, K-feldspar, and albite with respect to the activity ratios of K^+, Na^+, and H^+ ions in coexisting fluid at 300 °C. At high activities of H^+, the mineral kaolinite [$Al_2Si_2O_5(OH)_4$] becomes stable. Kaolinite is related to albite and muscovite by the following ion-exchange reactions:

(1) $2Ab + H_2O + 2H^+ = \text{kaolinite} + 4SiO_2 + 2Na^+$

$$K = 6.31 \times 10^8$$

(2) $2Musc + 3H_2O + 2H^+ = 3\ \text{kaolinite} + 2K^+$

$$K = 5 \times 10^4$$

Assuming unit activities for the minerals and H_2O, and using the equilibrium constants given, determine the position of the kaolinite stability field in Figure 21.8, and determine the $\log(a_{Na^+}/a_{H^+})$ and $\log(a_{K^+}/a_{H^+})$ in the fluid in equilibrium with albite, muscovite, and kaolinite at this temperature.

21.5 If periclase (MgO) and quartz are heated together, two reaction layers form between them, one of forsterite in contact with periclase, and the other of enstatite in contact with quartz. The rate of thickening of these layers depends on the rate of diffusion of Si and Mg through the layers, which in turn depends on the chemical potential gradients across the layers. First, construct a saturation surface for these minerals in a μ_{MgO} versus μ_{SiO_2} diagram for 700 °C and 35 MPa using the data in Table 7.1.

(a) What is the chemical potential decrease for SiO_2 and MgO across the forsterite layer and across the enstatite layer?

(b) Assuming that diffusion coefficients in forsterite and enstatite are similar, would you expect the forsterite or enstatite layer to thicken faster?

21.6 One could argue that sillimanite grew adjacent to the vein in Figure 21.18 because the vein fluids were very hot. This would require a steep thermal gradient from the vein selvage to the cooler surroundings where relict kyanite was preserved.

- **(a)** If the rocks had a typical thermal diffusivity of 10^{-6} m^2 s^{-1}, how long would it take before significant heat conduction would flatten such a steep thermal gradient? The distance from the vein to the area where relict kyanite is preserved is 2 cm.
- **(b)** Given this timescale, do you think it likely that the spatial distribution of sillimanite and kyanite is related to steep local thermal gradients?

21.7 Calculate diffusion–reaction profiles using Eq. (21.48). Use values of 10^{-8} m^2 s^{-1}, 0.5, and 10^{-3} for the diffusion coefficient, tortuosity factor, and porosity, respectively. The pore velocity will be zero for this case of pure diffusion. For simplicity, take $c_s^{\mathrm{o}} = 0$. If your program has trouble solving (21.48), see Problem 21.8 for some hints.

- **(a)** Plot results in terms of the normalized concentration $c_s^{x,t}/c_s^{\mathrm{B}}$ for $K_v = 0.1$ after 10^2, 10^4, and 10^6 years. How do the shapes of these profiles differ from those for advection-dominated mass transport (e.g. Fig. 21.15)?
- **(b)** Examine the measured geochemical profiles extending into the mélange block in Figure 21.29. Are these profiles qualitatively more consistent with diffusion-dominated or advection-dominated transport?

21.8 Use Eq. (21.48) to compute the propagation of an oxygen isotope front from metapelitic schist into a metacarbonate rock. The fluid in the schist is in isotopic equilibrium with calcite which has δ^{18}O = 15‰, whereas the fluid in the metacarbonate is initially in equilibrium with calcite having δ^{18}O = 19‰.

To solve the equation, use high (15 digit) precision for your computer calculations. Note that the complimentary error function, erfc, is related to the error function by: $\mathrm{erfc}(y) = 1 - \mathrm{erf}(y)$. The $\exp[\bar{v}_x x/D_{\mathrm{HD}}]$ term can become very large and exceed the computer program's numerical capacity for some combinations of variables. In this case, the erfc term that the exponential is multiplied by becomes very small, so their product can be set to zero. In addition, some programs have problems when large values or negative values are input into the error function. If this happens, remember that if the value of y in erf(y) is greater than ~4, erf(y) can be taken as one. If the program does not work for negative values of y, the situation is easily handled, as $\mathrm{erf}(-y) = -\mathrm{erf}(y)$ (see Table 5.1). The calculations for this problem should work using distance in meters and time in years.

- **(a)** For a pore velocity of 0.5 m a^{-1}, how far will the isotopic front propagate in 10^6 years if the porosity is 10^{-3}, the diffusion coefficient is 10^{-8} m^2 s^{-1}, the tortuosity is 0.5, and the coefficient of longitudinal dispersivity is 5 m?
- **(b)** From Eq. (21.50), what is the time-integrated fluid flux?
- **(c)** Use Eq. (21.60) to estimate the time-integrated fluid flux from the profile computed in part (a). How does this result compare with the value that you calculated using Eq. (21.50)?

21.9 Using Equation (21.67), calculate the time-integrated fluid fluxes needed to precipitate quartz veins at 600 ºC and 0.2 and 1.0 GPa assuming local fluid–rock equilibrium, silica-saturated H_2O, and a geothermal gradient of −25 ºC km^{-1}. The molar volumes of water at 0.2 and 1.0 GPa are 3.05×10^{-5} m^3 and 1.89×10^{-5} m^3, respectively (Eq. (20.6)). The values of $(\partial c_{\mathrm{SiO_2}}/\partial T)_P$ can be estimated graphically from Figure 21.6. Why do the results for 0.2 and 1.0 GPa differ so greatly?

21.10 This problem examines the time-integrated fluid flux (q_{TI}) needed to drive wollastonite production and decarbonation under conditions of 600 ºC and 0.35 GPa at the west-central New Hampshire locality described by Rumble *et al.* (1982) (see Section 20.6).

- **(a)** Estimate the q_{TI} value in m^3 m^{-2} using Eq. (21.73), assuming a temperature gradient of 25 ºC km^{-1}. Here, 17.5 moles of CO_2 were lost per 1000 cm^3 of rock. The mole fraction CO_2 and $(\partial X_{\mathrm{CO_2}}/\partial T)_P$ can be estimated from Figure 20.11. For this problem, assume that $(\partial X_{\mathrm{CO_2}}/\partial P)_T$ is small and can be neglected. Take the molar volume of the fluid to be 27.5 cm^3; this value can be calculated using Eq. (20.6) and ideal mixing of H_2O–CO_2 volumes.
- **(b)** The fluid flow in part (a) is in a direction of increasing temperature. From the shape of the equilibrium curve for wollastonite production in Figure 20.11, explain why this must be the case for one-dimensional advection and decarbonation proceeding according to this reaction under conditions of local fluid–rock equilibrium.
- **(c)** Compare your result for part (a) to the fluid:rock ratio of 4.5:1 computed in Section 20.6 for the same rock. Why are the values so different?

21.11 A high-grade metapelite cooling at 0.2 GPa contains andalusite + K-feldspar + quartz. The rock is dehydrated and contains no significant water in its pore space. If the rock is infiltrated by water during cooling, however, the retrograde growth of muscovite will occur by the general reaction: H_2O + K-feldspar + andalusite = muscovite + quartz. The purpose of this problem is to investigate how the reaction rate will change depending on the temperature of infiltration.

- **(a)** The ln K for this reaction at 0.2 GPa and 200–700 ºC is well described by: $\ln K = -7.446 + 6669/T$, where T is in K. At a reference temperature T^0 of 600 ºC, κ^0 is 10^{-4} moles m^{-2} a^{-1} (J mole^{-1})$^{-1}$ for a linear kinetic model (Schramke *et al.*, 1987). Calculate and plot the reaction rate for 600, 590, 550, 500, 400, 300, and 200 ºC using Eq. (21.78) for a rate-limiting surface area of 100 m^2 m^{-3}, an activation energy of 80 kJ mole^{-1}, and a porosity of 10^{-3}. Assume that the solids and H_2O are pure.
- **(b)** Discuss why the curve plotted in part (a) has the shape that it does. Do higher temperatures always result in faster rates? What are the implications for retrograde hydration?

Table 21.1 *For problem 21.12*

Major oxides, minor oxides (wt%); Rb, Sr, Ba, Zr, Nb, and Y (ppm)

	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	MnO	CaO	K_2O	Na_2O	P_2O_5	LOI	Total	Rb	Sr	Ba	Zr	Nb	Y
Selvage	41.3	0.33	14.70	4.74	6.38	0.39	25.10	0.04	0.09	0.34	6.10	99.6	4	624	31	57	9	13
Precursor	28.5	0.37	6.55	4.79	6.89	0.42	32.10	0.69	0.13	0.37	18.40	99.3	35	876	209	65	8	13

Rare earth elements, Th, and U (ppm)

	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Th	U
Selvage	17.0	32.3	3.9	15.9	3.3	1.18	3.2	0.4	2.7	0.46	1.5	0.2	1.4	0.19	6.2	1.4
Precursor	18.7	35.0	4.4	17.6	3.4	1.17	3.2	0.4	2.7	0.49	1.6	0.2	1.7	0.21	6.4	1.5

Total Fe is given as Fe_2O_3. LOI denotes loss on ignition, which is determined during the course of X-ray fluorescence analysis and is a proxy for $H_2O + CO_2$ content as long as the rock is not extremely rich in Fe^{2+}, carbonaceous organic matter, or sulfides. Total wt% includes Rb, Sr, Ba, Zr, and Y summed as oxides.
Data from Ague (2003a).

21.12 Metasomatism resulting from fluid flow through a fracture produced a diopside- and zoisite-rich selvage in metacarbonate rock. The less altered precursor rock contains abundant biotite which broke down during the selvage-forming reaction. From the geochemical data given in Table 21.1, construct a concentration ratio (CR) diagram to determine the total mass change associated with selvage formation. Next determine the mass changes for individual elements by drawing mass change contours on the CR diagram, as was done in Figure 21.36. Finally, estimate the volume change, given selvage and precursor densities of 3210 kg m^{-3} and 3040 kg m^{-3}, respectively. Summarize in words the major compositional changes associated with selvage formation. Al is often assumed to be immobile in metamorphic fluids. From your CR diagram, is this always a valid assumption?

21.13 Metacarbonate rocks from south-central Connecticut lost substantial CO_2 during prograde heating. K mass transfer occurred when biotite broke down to form amphibole or diopside. Tables 21.2 and 21.3 contain K_2O, LOI (loss on ignition), and Zr data for representative suites of Ankerite–albite zone precursor rocks, and high-grade Diopside zone rocks in which biotite was destroyed. LOI is a reasonable proxy for $CO_2 + H_2O$ content in these rocks (Ague, 2003a; see Table 21.1).

(a) Construct two wedge diagrams by plotting LOI versus Zr and K_2O versus Zr (see Fig. 21.37). Zr was less mobile than volatiles or K and thus provides a useful geochemical reference frame. Was K lost or gained from the Diopside zone rocks when biotite broke down?

(b) Use Eq. (21.100) to compute the average, or "most probable" mass changes for volatiles and K, and the associated 2σ uncertainties.

Table 21.2 *Ankerite–albite zone data for Problem 21.13*

K_2O (wt%)	LOI (wt%)	Zr (ppm)
1.81	27.9	53
2.15	27.7	68
1.26	27.2	56
1.25	28.9	49
1.51	27.4	61
1.79	28.5	58
1.82	28.3	66
1.76	29.8	49
1.90	27.5	67

Table 21.3 *Diopside zone data for Problem 21.13*

K_2O (wt%)	LOI (wt%)	Zr (ppm)
1.22	14.6	64
0.11	19.8	52
0.04	14.6	63
0.01	12.0	68
0.04	13.5	75
0.15	21.7	54
0.13	13.5	78
0.10	18.3	61
0.12	13.6	69
0.09	23.6	58
0.03	14.9	90
0.06	21.1	50
0.04	12.4	106

21.14 Table 21.4 gives representative data for a geochemical traverse along an altered metasiltstone layer cut by a quartz vein from southeastern Scotland (Ague, 1997). Distances are measured from the center of the quartz vein, with 0.8 cm being the closest wall rock to the

Table 21.4 *Major oxides, minor oxides (wt%); Rb, Sr, Ba, Zr, Nb, and Y (ppm)*

Distance (cm)	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MgO	MnO	CaO	K_2O	Na_2O	P_2O_5	LOI	Total	Rb	Sr	Ba	Zr	Y
0.80	73.3	0.54	11.0	6.15	1.86	0.05	1.48	0.84	1.11	0.07	2.55	98.95	41	137	102	230	16
2.55	73.6	0.54	11.4	6.05	1.68	0.04	1.14	1.54	0.94	0.07	2.75	99.75	67	121	226	218	12
4.20	74.5	0.56	11.1	5.74	1.61	0.04	0.62	1.89	0.59	0.08	2.90	99.63	73	88	369	219	14
5.05	74.3	0.54	11.3	5.79	1.64	0.05	0.54	1.98	0.51	0.08	2.85	99.58	76	77	426	229	13
10.2	76.2	0.51	10.9	5.51	1.50	0.05	0.32	2.12	0.29	0.08	2.50	99.98	72	61	476	193	14
14.2	76.9	0.51	10.6	5.18	1.40	0.05	0.22	2.12	0.26	0.08	2.35	99.67	72	60	498	185	13

vein. The mass transfer stabilized garnet in the alteration selvage adjacent to the vein.

(a) The vein selvage contains the *AFM* assemblage garnet–biotite–chlorite, whereas the less altered rock distal to the vein contains chloritoid–biotite–chlorite. Draw schematic *AFM* diagrams for the selvage and the less altered precursor.

(b) Plot geochemical profiles showing Na/K and Ca/K as molar ratios, and Sr/Rb and Sr/Ba as mass ratios. Next plot Na_2O/K_2O (wt%) and compare with the molar ratio graph. Are your conclusions about the change in the abundance of Na relative to K dependent on whether you plot molar or wt% ratios?

(c) Ti was relatively immobile in these rocks. By plotting ratios with Ti in the denominator, determine if Si, Na, K, Ca, Sr, Rb, and Ba were lost or gained.

(d) The overall chemical reaction in the selvages destroyed muscovite and produced garnet and abundant plagioclase. Are the results from part (c) consistent with these mineralogical changes? Based on the Si/Ti relations, do you think it is possible that silica was removed from the selvage to fill the adjacent vein?

21.15 Equation (21.53) for transport and kinetically controlled reaction can be simplified if there is no advection

$$\frac{\partial c_s}{\partial t} = D\frac{\partial^2 c_s}{\partial x^2} + R_s$$

where D is the diffusion coefficient incorporating the effects of tortuosity. A very simple expression for the rate of an irreversible, first-order reaction is

$$R_s = \kappa^*\left(c_s^{\mathrm{eq}} - c_s\right)$$

where κ^* is a rate "constant," and c_s^{eq} denotes the equilibrium concentration of species s in the fluid. Consequently, we can write

$$\frac{\partial c_s}{\partial t} = D\frac{\partial^2 c_s}{\partial x^2} + \kappa^*\left(c_s^{\mathrm{eq}} - c_s\right)$$

A solution of this partial differential equation for diffusion–reaction away from a boundary at $x = 0$ is (cf. Eq. (14.46) in Crank, 1975)

$$c_s^{x,t} = c_s^{\mathrm{eq}} + \left(c_s^{\mathrm{B}} - c_s^{\mathrm{eq}}\right)\Bigg[0.5\exp\left(-x\sqrt{\kappa^*/D}\right)\operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}} - \sqrt{\kappa^* t}\right) + 0.5\exp\left(x\sqrt{\kappa^*/D}\right)\operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}} + \sqrt{\kappa^* t}\right)\Bigg]$$

where c_s^{B} is the concentration at the boundary, and $c_s^{x,t}$ is the concentration at position x and time t. For this expression, fluid in the interior of the region ($x > 0$) initially has the equilibrium concentration c_s^{eq}. Imagine that a K-enriched fluid flowing through ultramafic mélange encounters a metabasaltic block that contains a small amount of phengitic mica. New mica begins to grow in the reaction rind on the block as a result of K diffusion into the block. This diffusion occurs along the x-direction, perpendicular to the contact between block and mélange matrix. The fluid–rock interaction causes mica to grow in the reaction rind, and this ultimately removes the excess K from the infiltrating fluid so that the concentration in the interior of the block remains at the equilibrium value c_s^{eq}. If there is a large amount of fluid flow in the matrix, the K concentration at the mélange-block contact at $x = 0$ (c_s^{B}) can remain fixed.

(a) Compute and plot the concentration profile in terms of the normalized concentration $\left(c_s^{x,t} - c_s^{\mathrm{eq}}\right)/\left(c_s^{\mathrm{B}} - c_s^{\mathrm{eq}}\right)$ for 0.01, 0.1, 10, 100, and 10 000 years using a diffusion coefficient of 10^{-8} m^2 s^{-1}, a tortuosity factor of 0.5, and $\kappa^*=10$ a^{-1}. Plotting the normalized concentration simplifies the problem because it illustrates the characteristic shape of the profiles without having to plug in specific values for the actual concentrations. Read the instructions for Problem 21.8 if your computer calculation runs into problems. Explain why the profiles look the same for long times.

(b) How do your results change if κ^* is increased to 100 a^{-1} or reduced to 0.1 a^{-1}? Why do these changes occur, and what does this mean regarding the length scale over which fluid and rock are out of equilibrium?

22 Pressure–temperature–time paths and heat transfer during metamorphism

22.1 INTRODUCTION

Most metamorphic rocks have mineral assemblages that can be shown to have crystallized at elevated pressures and temperatures. With certain assemblages, these conditions can be specified quite closely using thermodynamic, kinetic, and experimental data. The important question that we must now address is: to what point in the history of a metamorphic rock do these conditions relate? The protoliths of many metamorphic rocks have their origins on the surface of Earth as sediment or lava. With burial, increasing pressure and temperature lead to the development of metamorphic minerals, which undergo successive changes as pressure and temperature continue to change. Then, through erosion and exhumation, the rocks return to Earth's surface. The rocks consequently undergo changes in pressure and temperature, which are both functions of time. The question, then, is: what part of this pressure–temperature–time (*P–T–t*) path is recorded in the final metamorphic rock?

The approach to equilibrium at one set of elevated temperature and pressure conditions is not always complete in metamorphic rocks, as evidenced by the presence of zoned minerals and overgrowths of one mineral on another (Fig. 22.1). The zoning in many porphyroblasts, in particular of aluminous minerals such as garnet, records ranges of pressure and temperature during the growth of the mineral. This information, in conjunction with *P–T* conditions indicated by the mineral assemblage of the rock, allows partial *P–T–t* paths to be constructed for some metamorphic rocks. As might be expected, metamorphic rocks can follow a variety of different *P–T–t* paths, depending on the specific tectonic conditions. The fact that many of these paths can be identified using analyses of zoned porphyroblasts and other petrologic data means that metamorphic petrologists can make important contributions to the understanding of the development of mountain belts.

Following England and Molnar (1990) and Stüwe (2002), we use the words *exhumation* and *burial* to denote vertical movement of rocks relative to Earth's surface (Fig. 22.2). For example, exhumation is the removal of overburden by erosion or tectonic processes that reduce the distance between the surface and a rock at depth (Problem 22.9). *Surface uplift* and *subsidence* refer to vertical motion of Earth's surface with respect to some reference level. Thus, the total vertical displacement of a rock relative to some reference level is the sum of the amounts of exhumation/burial and surface uplift/subsidence. We can refer to this total displacement as the *uplift of rock* (England and Molnar, 1990; Stüwe, 2002). All of the motions discussed above can also be expressed as rates in units such as m s^{-1}, cm a^{-1}, or km Ma^{-1}. Exhumation and burial rates will have opposite signs, as will surface uplift and subsidence rates. By comparing Figure 22.2(B) and (C), we can see that geobarometry can be used to estimate exhumation, but not surface uplift or the total uplift of a rock.

22.2 PRESERVATION OF METAMORPHIC MINERAL ASSEMBLAGES

The fact that metamorphic rocks on Earth's surface today have mineral assemblages that were stable at elevated pressures and temperatures indicates that they are now out of equilibrium (metastable) under the conditions on Earth's surface. It is remarkable, then, that rocks which so easily become metastable during unloading under retrograde conditions can, during prograde metamorphism, recrystallize so completely. Clearly, the rates of metamorphic reactions must change considerably with changing conditions, and kinetics must play an important role in determining what is preserved in a metamorphic rock.

The fluid phase is extremely important in determining what mineral assemblage is preserved. Most prograde reactions involve devolatilization. The extremely low porosity of metamorphic rocks indicates that these volatiles are almost completely expelled during reaction. Thus, when metamorphic temperatures fall, the volatiles needed for the retrograde reactions are absent. Small amounts of fluid that may remain are consumed rapidly in the formation of small amounts of retrograde minerals, such as chlorite rims on garnet crystals. A cooling metamorphic rock therefore tends to have a very low fluid content.

If a fluid phase is present to facilitate retrogression, the reaction rates will depend on several geologic factors. As we saw in Chapter 21, these include the intrinsic reaction rate constant κ, the reactive surface area, and the free-energy change of reaction (Eq. (21.78)). As a rock cools below the

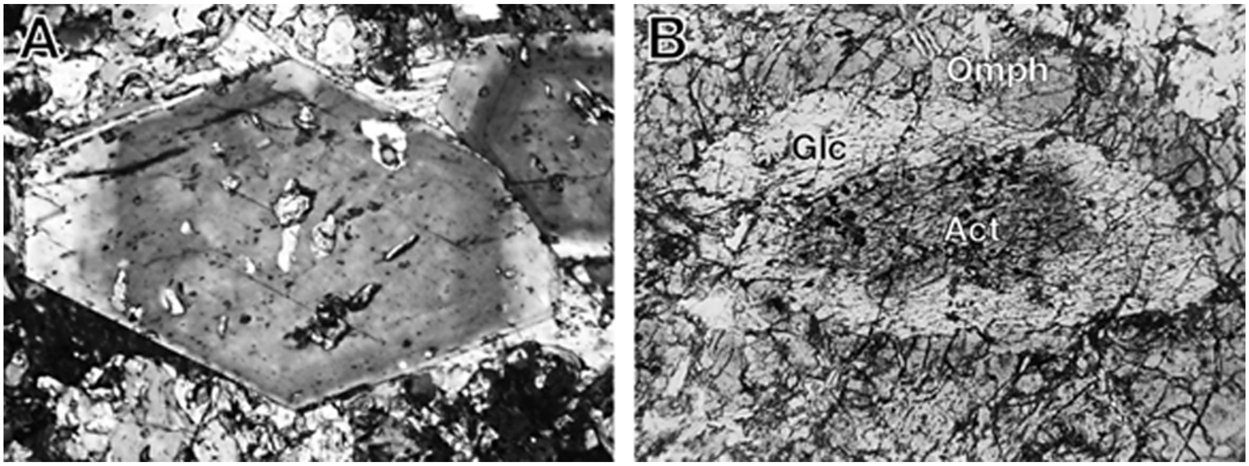


Fig. 22.1 (**A**) Glaucophane crystals that are chemically zoned, Tinos Island, Greece. Crossed-polarized light. Field of view is 0.7 mm. (**B**) Relict actinolite (Act) rimmed by glaucophane (Glc) which is in turn rimmed by omphacite (Omph). Actinolite is probably a relict from hydrothermal alteration on the sea floor, whereas the glaucophane and then the omphacite formed at successively greater depths during subduction and prograde metamorphism. Metagabbro, Syros Island, Greece. Plane-polarized light. Field of view is 4.1 mm.

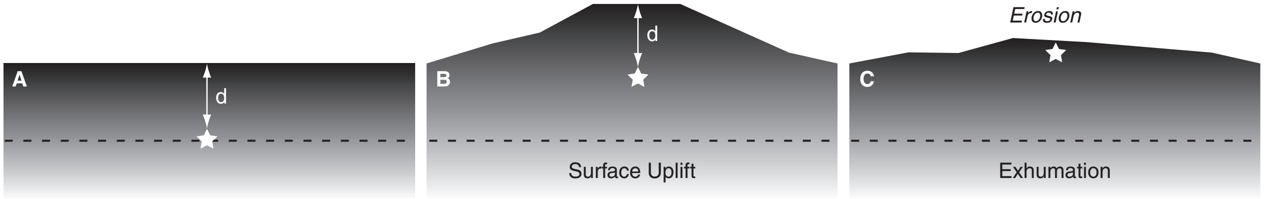


Fig. 22.2 Uplift of rock. Rock initially at depth *d* shown with star in part (**A**). Arbitrary datum shown with dashed line. (**B**) Tectonic processes produce *surface uplift* as well as *uplift of the rock* relative to the datum. No exhumation takes place, however, because the distance *d* has remained the same. (**C**) Exhumation due to erosion. Uplift of the rock relative to the datum has continued.

equilibrium temperature for a retrograde reaction, the value of κ, which is strongly dependent on temperature, will decrease in magnitude and this will act to decrease rates (Fig. 21.17). However, the amount of overstepping of the reaction increases as temperature falls, so the ΔG of reaction becomes more negative, acting to increase rates. These two effects operate together. First, ΔG dominates and rates will tend to increase with cooling away from the equilibrium condition. As temperatures continue to fall, however, the strong temperature dependence of κ becomes dominant, and rates then decrease with further cooling (Problem 21.11). Of course, if there were changes in other key variables, such as reactive mineral surface area, more complex behavior would be possible.

The fluid phase also plays an important kinetic role during metamorphism through its effect on diffusion. Prograde devolatilization reactions cause fluid pressures to rise and, in some cases, slightly exceed lithostatic pressures. This ensures that grains have fluid passing along dilated grain boundary channels and may even cause hydrofracturing. Diffusion rates in intergranular fluid are orders of magnitude greater than in dry rocks. Reaction rates in the presence of these fluids are therefore greater than in rocks that lack fluid.

Retrograde metamorphic rocks certainly do occur, and are usually found in tectonic settings where H_2O and CO_2 have been able to gain access to the rock. Fault zones, for example, are commonly flanked by retrograde metamorphic rocks formed from fluids that were able to enter the rock through the fault zone. Where retrograde metamorphism has developed on a regional scale, higher-grade rocks are normally found to have been transported over lower-grade rocks, either in thrust slices or in nappes. As the lower-grade rocks are heated and undergo prograde reactions, the liberated volatiles rise into the overlying higher-grade rocks, where they cause retrograde reactions to occur.

The burial necessary to bring about regional metamorphism normally takes place during periods of plate convergence. Once convergence stops, exhumation processes bring deeply buried rocks back to the surface. Most prograde metamorphic reactions therefore occur while folding and faulting are still active. The accompanying strain provides an additional source of energy to promote reactions. Grain breakage and recrystallization also provide grain boundaries along which reactants can migrate more easily. By contrast, the small amounts of strain associated with exhumation provide little extra driving force for retrograde reactions.

Because of the marked differences in the processes of prograde and retrograde reactions, most metamorphic rocks retain the mineral assemblages formed at peak metamorphic conditions; that is, they preserve the highest-temperature or

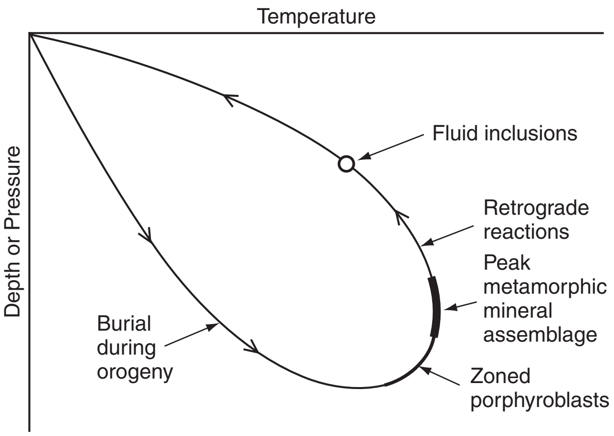


Fig. 22.3 Possible path that a sedimentary rock might follow as a result of burial during orogenesis, followed by metamorphism, and eventually uplift due to erosion. The mineral assemblage that is likely to be preserved is developed near the maximum temperature. Zoned porphyroblasts may preserve a record of conditions just prior to the peak temperatures, and retrograde minerals may develop as rims on the high-*T* minerals during cooling. Fluid inclusions may be trapped in healed fractures during exhumation.

highest-entropy assemblages formed along the *P–T–t* paths. However, zoned porphyroblasts and rims on minerals preserve mineralogical records of conditions before and after the peak conditions (Figs. 12.27, 19.6, 22.1). Because of the Arrhenius relationship discussed in Section 5.6, rates of diffusion will become increasingly sluggish as temperatures fall, particularly in aluminous minerals like garnet. In this way, prograde mineral compositions can be preserved in the cores of grains, although rims may be retrograded to varying degrees (Fig. 19.6(B), (C)). Relict textures, such as inclusions within porphyroblasts, can also indicate the former presence of diagnostic prograde minerals.

Another source of *P–T* information comes from fluid inclusions. During exhumation, mineral grains may fracture, and if the rock is still hot enough and especially if fluids are present, they heal themselves. In doing so, they commonly trap fluid. In thin section, healed fractures are consequently visible as planar arrays of fluid inclusions. As the rock rises to the surface and cools, the fluid may split into a gas and a liquid phase, the gas forming a small bubble in the liquid. By heating such samples on a microscope hot stage until the inclusions become a single phase again, we can determine the minimum temperature of entrapment.

Figure 22.3 shows one possible path that a metasedimentary rock might follow during its history. Initially, burial and compaction leads to the lithification of the sediment. The thickening of the crust during orogenesis results in further burial and increasing pressures. At the same time the influx of heat from the geothermal gradient and from igneous intrusions and the generation of heat by radioactive elements in the sedimentary pile cause the temperature to rise. The rock undergoes a series of progressive metamorphic reactions during this stage, but only zoned porphyroblasts of certain minerals and relict textures preserve any record of these *P T* conditions. Eventually, the rock reaches its maximum temperature, but by this time it has already started to return to the surface of Earth as a result of exhumation occurring concurrently with the erosion of overlying mountains. As the rock gets nearer to the surface, the temperature begins to fall. If little or no fluid enters the rock, the amount of retrogression will be very limited. Moreover, as temperatures continue to fall, the rates of retrograde reactions and diffusion will decrease sharply, further limiting any modification of the high-temperature mineral assemblage. Some fluid inclusions may be trapped, but their time of entrappment is often difficult to determine precisely. Eventually, the high-temperature mineral assemblage arrives at Earth's surface essentially unchanged.

22.3 METAMORPHIC FIELD GRADIENTS

Changes in mineral assemblage across metamorphic terranes indicate systematic changes in metamorphic grade, which can be correlated mainly with changes in temperature and to a lesser extent pressure. The record of these *P–T* conditions has been referred to as a *metamorphic field gradient* (Spear *et al*., 1984), a *PT array* (Thompson and England, 1984), and a *metamorphic geotherm* (England and Richardson, 1977). The latter term is rather close to the word *geotherm*; however, these gradients are not geotherms, despite their early interpretation as such. We will therefore use the term "field gradient" or "*PT* array."

Barrow's (1893) mapping of isograds in the Scottish Highlands clearly established the existence of regional metamorphic field gradients. As more areas were studied, it became evident that the gradients were not everywhere the same, as evidenced by different sequences of metamorphic facies. In 1961, Miyashiro published an important paper in which he showed that three main types of gradient could be recognized. Because these are distinguished by the sequence of metamorphic facies, he named them *metamorphic facies series*. They are the high-temperature, low-pressure (andalusite–sillimanite) type; the moderate-temperature, moderate-pressure (kyanite–sillimanite) type – this was the series mapped by Barrow; and the low-temperature, high-pressure (jadeite–glaucophane) type (Fig. 22.4). To these main facies series Miyashiro added two intermediate ones: an intermediate-temperature, low-pressure type, and an intermediate-temperature, high-pressure type (Fig. 22.4). A complete continuum of *P–T* series probably exists, but the frequency of occurrence of the three main types is so great that tectonic conditions must favor their formation.

Miyashiro (1961) also recognized that around the Pacific plate, metamorphic rocks belonging to the low-*T*, high-*P* series and to the high-*T*, low-*P* series develop penecontemporaneously in *paired metamorphic* belts. The low-*T*, high-*P* series invariably occurs nearest the oceanic trench, whereas the high-*T*, low-*P* series develops in continental crust away from the oceanic plate (Figs. 21.1, 23.10). Although the low-*T*, high-*P* series contains metamorphosed igneous rocks from

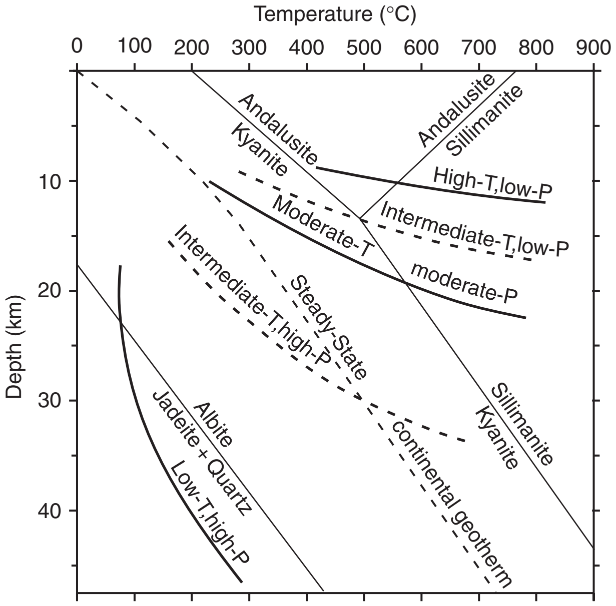


Fig. 22.4 Mineral facies in metamorphic terranes commonly define arrays of *P–T* conditions corresponding to the three solid lines in the diagram. Miyashiro (1961) defined these as metamorphic facies series, each being named for the typical range of *P–T* conditions they cover. Less common intermediate facies series (dashed lines) are also recognized. For reference, the stability fields of the Al_2SiO_5 polymorphs, the albite = jadeite + quartz reaction, and the steady-state continental geotherm (Section 1.6) are included. (Modified from Miyashiro, 1961.)

the ocean floor (ophiolite suite, Section 15.2), they contain almost no intrusive igneous rocks. By contrast, the high-*T*, low-*P* terranes contain an abundance of intrusive and extrusive calc-alkali rocks (Section 15.3). The occurrence of these paired metamorphic belts in specific tectonic settings is clear evidence that tectonic processes play an important role in determining the types of metamorphic field gradient that develop (Chapter 23).

Figure 22.4 shows that rocks belonging to the high-*T*, low-*P* and low-*T*, high-*P* metamorphic facies series are formed, respectively, at temperatures above and below the normal steady-state continental geotherm. The elevated temperatures in the high-*T*, low-*P* series can be accounted for by large-scale intrusion of calcalkali magma, whereas the low temperatures in the other series can be accounted for by rapid burial of cold surface sediments. In all facies series, however, the temperature rise in field gradients is normally much greater (shallower slope in Fig. 22.4) than can be accounted for by the geothermal gradients that must have produced these rocks. Although rapid advective rise of metamorphic fluids through major conduits and the proximity to igneous intrusions may account for steep gradients in temperature, they cannot account for all of the regional gradients observed in nature.

This apparent paradox is resolved once it is recognized that field gradients are based on *P–T* conditions that were not developed simultaneously throughout a given metamorphic terrane. This will be evident if we consider the *P–T–t* paths of three different sedimentary rocks that, during metamorphism, are buried to different depths. These three paths are shown in Figure 22.5 along with lines of constant time and a typical steady-state continental geotherm. When rock 1 reaches its maximum temperature and develops the mineral assemblage that will be preserved and brought to the surface, rock 3 is just reaching the deepest part of its *P–T–t* path. When rock 2 reaches its maximum temperature, rock 3 is still heating. The three rocks therefore reach their maximum temperatures at different times. The metamorphic field gradient mapped on Earth's surface is the line joining the temperature maxima on the individual *P–T–t* paths. The slope of the field gradient is obviously very different from that of the geothermal gradient at any given time. We can conclude, therefore, that the range of *P–T* conditions recorded in the rocks of a metamorphic terrane are probably formed at different times and thus cannot be interpreted as fossilized geotherms.

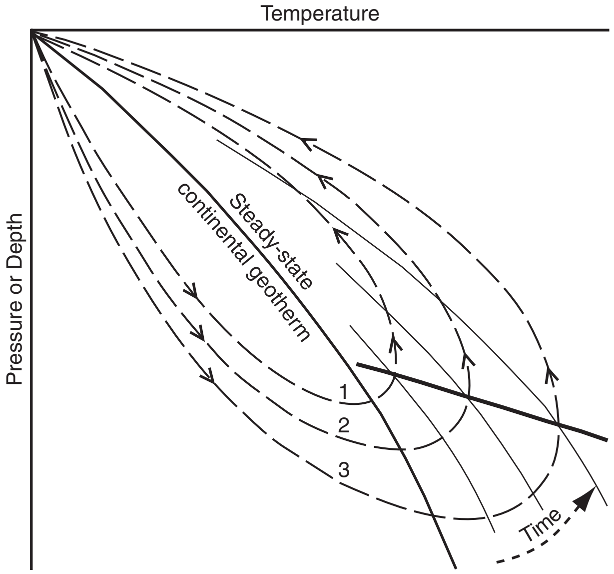


Fig. 22.5 Schematic representation of *P–T–t* paths for three rocks, 1, 2, and 3 (dashed arrowed lines). Fine solid lines join points of equal age on the *P–T–t* paths; they also follow the geothermal gradient at any given time. The solid heavy line joining the temperature maxima on the *P–T–t* paths is the field gradient. Steady-state geothermal gradient also shown. See text for discussion.

22.4 CALCULATION OF PRESSURE–TEMPERATURE–TIME PATHS

Unraveling the history of most metamorphic terranes is complex and in many areas may be impossible to do in detail because pertinent parts of the record may have been eroded away or eliminated by annealing and prograde textures. Despite general similarities imposed by particular tectonic settings, each metamorphic terrane is unique. The particular *P–T–t* paths followed by rocks depend on many factors, such as the style of deformation, number of thrust faults, number of periods of folding, the amount of magma intruded, erosion rates, and duration of the orogenic episode. In view of these complexities, a clearer understanding of the principles governing *P T* paths can be gained from a study of simple models than from a detailed analysis of any one area. For additional perspectives, Evans (2007b) provides a highly readable commentary on the landmark *P–T–t* path study of England and Richardson (1977), as well as on many other classic papers in metamorphism.

Several simple tectonic models have been investigated for which *P–T–t* paths can be calculated numerically (England and Richardson, 1977; England and Thompson, 1984; Karabinos and Ketcham, 1988). In each model, the crust is thickened to give the pressures recorded in metamorphic rocks. Thickening of the crust eventually leads to radiogenic heating (Problem 1.5). At the same time, because of erosion of the overlying mountains formed during crustal thickening, rocks move toward the surface of Earth. The actual temperature of a rock at any given time depends on the rates of these various processes.

To determine the temperature and pressure of a rock at any given time, the equation of heat transfer in a conductive, moving medium with its own sources and sinks of heat must be solved. We have already dealt separately with each aspect of this equation. First, Fourier's equation (Eq. (5.11)) provides the basic relation for heat conduction:

$$\text{conductive heating:}\quad \frac{\partial T}{\partial t} = k_{\mathrm{T}}\frac{\partial^2 T}{\partial z^2}$$

where k_{T} denotes the thermal diffusivity. We use the subscript $_{\mathrm{T}}$ in this chapter to avoid confusion with the symbol k used for permeability in Chapter 21. The heat flux due to conduction is

$$\text{conductive heat flux:}\quad J_{z,\mathrm{Heat,cond}} = K_{\mathrm{T}}\frac{\partial T}{\partial z}$$

where $J_{z,\mathrm{Heat,cond}}$ is the flux and K_{T} is the thermal conductivity (Eq. (5.3)).

The temperature change resulting from heat sources or sinks is some function of t and z; that is, $h(t, z)$. It might include heats of metamorphic reaction, frictional generation of heat on faults, intrusion and crystallization of magma, and radiogenic heat. We will consider only this last source for now, and examine magmas and heats of reaction in the following sections. If radiogenic elements are distributed evenly throughout the body of rock in question, then $h(t, z)$ becomes simply the rate of radiogenic heat production per unit volume of rock, or $A/C_{\mathrm{P}}\rho$ (Eq. (1.8)). The temperature change resulting from radiogenic heating is then

$$\text{radiogenic heating:}\quad \frac{\partial T}{\partial t} = \frac{A}{C_{\mathrm{P}}\rho}$$

If the temperature at a given depth in the crust is to be changed by vertical movement of the rock mass during orogeny, there must exist a temperature gradient, $\partial T/\partial z$, in the rock. Then, as the rock mass moves, the temperature will change. For example, if the temperature at a given point in Figure 22.6 is T_1 at time t_1, its temperature rises to T_2 at time t_2 because of the passage of the temperature gradient. The temperature at any time can be determined if we know the temperature gradient and the velocity at which the rock mass moves.

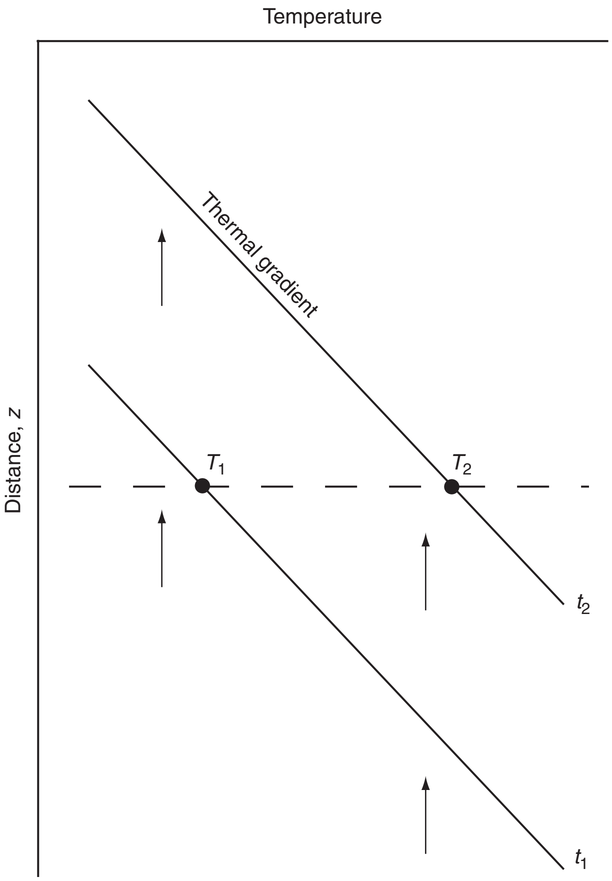


Fig. 22.6 Advection of rock in which there is a temperature gradient causes the temperature at a given point T_1 at time t_1 to rise to T_2 at time t_2 as the rock passes.

The heat flux due to this mass transport is simply equal to the product of the rock mass flux in $\mathrm{m}^3{}_{(\mathrm{rock})}\,\mathrm{m}^{-2}{}_{(\mathrm{rock})}\,\mathrm{s}^{-1}$, the heat content, and the temperature

$$J_{z,\mathrm{Heat,MT}} = UC_{\mathrm{P}}\rho T \tag{22.1}$$

where $J_{z,\mathrm{Heat,MT}}$ is the heat flux and U is the mass flux. The mass flux U can be replaced by velocity (u, in m s^{-1}), which is also a function of t and z. Next we need to formulate the energy conservation equation. It turns out that we can write a general balance equation for mass or energy having the following form (Problem 22.3)

$$\frac{\partial(\Omega)}{\partial t} = -\frac{\partial(J_\Omega)}{\partial z} + A(\Omega) \tag{22.2}$$

where Ω is the amount of the conserved quantity per volume, J_Ω is the flux of Ω per area per time, and $A(\Omega)$ is the local production or consumption of Ω per time (Haase, 1990). For energy balance, Ω must be in terms of joules $\mathrm{m}^{-3}{}_{(\mathrm{rock})}$. As we saw earlier in Chapter 5, this is simply the product of the density, specific heat capacity, and temperature. The energy flux J_Ω is $J_{z,\mathrm{Heat,MT}}$. At this stage, we won't worry about heat production, so $A(\Omega) = 0$. Thus, using Eq. (22.2) as a model, we can write the following expression for energy conservation due to mass transport heating

$$\frac{\partial(C_{\mathrm{P}}\rho T)}{\partial t} = -\frac{\partial\left(J_{z,\mathrm{Heat,MT}}\right)}{\partial z} = -\frac{\partial(uC_{\mathrm{P}}\rho T)}{\partial z} \tag{22.3}$$

The reader will note that this equation is mathematically analogous to Eq. (21.44) written for mass conservation, illustrating the applicability of (22.2) to mass or energy balance problems. Equation (22.3) can be simplified by holding the density and heat capacity constant. In our model, the velocity depends on the rate of erosion, which we will also assume to be constant. With these simplifications, the rate of temperature change at any depth z due to mass transport of rock is then

$$\text{mass transport heating:} \quad \frac{\partial T}{\partial t} = -u\frac{\partial T}{\partial z}$$

where $u = \partial z / \partial t$. The simplified model we consider below does not explicitly account for surface uplift or subsidence (vertical motion of Earth's surface relative to a reference datum). The rocks are brought to the surface by erosional exhumation. For this particular problem, it is convenient to adopt the convention that z increases *downward*, so exhumation gives a negative value of u and, thus, $\partial T/\partial t$ will be positive. Note that the discussion of fluid flow in Chapter 21 uses the convention that z increases *upward*.

The expression for the total change in temperature with time at any depth z needs to consider conduction, mass transport, and radiogenic heating operating together. Making use of the general balance equation (Eq. (22.2)), we can write

$$\frac{\partial(C_{\mathrm{P}}\rho T)}{\partial t} = -\frac{\partial\left(J_{z,\mathrm{Heat,cond}} + J_{z,\mathrm{Heat,MT}}\right)}{\partial z} + A$$

which in turn gives

$$\frac{\partial(C_{\mathrm{P}}\rho T)}{\partial t} = -\frac{\partial\left(-K_{\mathrm{T}}\frac{\partial T}{\partial z} + uC_{\mathrm{P}}\rho T\right)}{\partial z} + A$$

and

$$\frac{\partial(C_{\mathrm{P}}\rho T)}{\partial t} = -\frac{\partial\left(-K_{\mathrm{T}}\frac{\partial T}{\partial z}\right)}{\partial z} - \frac{\partial(uC_{\mathrm{P}}\rho T)}{\partial z} + A$$

For constant K_{T}, u and $C_{\mathrm{P}}\rho$, the expression simplifies to

$$\frac{\partial T}{\partial t} = k_{\mathrm{T}}\frac{\partial^2 T}{\partial z^2} - u\frac{\partial T}{\partial z} + \frac{A}{C_{\mathrm{P}}\rho} \tag{22.4}$$

Thus, the final result is simply the sum of the conductive heating, mass transport heating, and radiogenic heating terms discussed above. This partial differential equation can be solved by the finite-difference method of Crank and Nicolson (1947) (see Section 5.4). Substituting the relations in Eqs. (5.36) and (5.37) into Eq. (22.4), and making use of the same grid as in Figure 5.10, Eq. (22.4) expressed in finite-difference terms is

$$\begin{aligned}\frac{T_{m,n+1} - T_{m,n}}{\delta t} = {} & \frac{k_{\mathrm{T}}}{2(\delta z)^2}\left(T_{m+1,n} + T_{m-1,n} - 2T_{m,n}\right. \\ & \left. + T_{m+1,n+1} + T_{m-1,n+1} - 2T_{m,n+1}\right) \\ & + \frac{A}{C_{\mathrm{p}}\rho} - u\frac{\partial T}{\partial z}\end{aligned}$$

which on rearranging gives

$$\begin{aligned} & T_{m,n+1}\left[1 + \frac{k_{\mathrm{T}}\delta t}{(\delta z)^2}\right] \\ & = \frac{k_{\mathrm{T}}\delta t}{2(\delta z)^2}\left(T_{m+1,n} + T_{m-1,n} + T_{m+1,n+1} + T_{m-1,n+1}\right) \\ & \quad + T_{m,n}\left[1 - k_{\mathrm{T}}\frac{\delta t}{(\delta z)^2}\right] + \delta t\frac{A}{C_{\mathrm{P}}\rho} \\ & \quad - u\frac{\delta t\left[T_{m,n} - T_{m-1,n}\right]}{\delta z} \end{aligned} \tag{22.5}$$

This equation can be simplified by selecting values of δt and δz so that $k_{\mathrm{T}}\ \delta t / (\delta z)^2 = 1$. In addition, values of $C_{\mathrm{P}}\rho$ for crustal and upper mantle rocks are nearly constant at $2.5 \times 10^6\ \mathrm{J\ m^{-3}\ K^{-1}}$. Equation (22.5) thus reduces to

$$\begin{aligned} T_{m,n+1} = {} & \frac{1}{4}\left(T_{m+1,n} + T_{m-1,n} + T_{m+1,n+1} + T_{m-1,n+1}\right) \\ & + 2 \times 10^{-7}\delta t A - u\frac{\delta t\left[T_{m,n} - T_{m-1,n}\right]}{2\delta z} \end{aligned} \tag{22.6}$$

Values of T at any depth m and time $n+1$ can be solved for iteratively using Eq. (22.6) if initial and boundary conditions are specified. These conditions depend on the particular model being investigated. We will consider one particular model in order to illustrate the effects of the various parameters on *P–T–t* paths. The real world is, of course, far more complicated than the simple model dealt with here, and to make matters worse, only small segments of *P–T–t* paths are likely to be preserved in rocks. The model, however, provides a framework with which to make comparisons and test hypotheses.

Most metamorphic rocks found on the surface have mineral assemblages that indicate crystallization at elevated pressure and, consequently, at some depth in the Earth. In ancient, eroded mountain belts these rocks are now underlain by normal thickness continental crust (about 35 km), and thus at the time of metamorphism the crust must have been thicker. Models for deriving *P–T–t* paths must start, therefore, with some mechanism for thickening the crust. In orogenic belts, crustal thickening is a consequence of plate convergence, and in detail is produced by folding, faulting, and simple compression and flattening. The particular mechanism affects the details of the model, but the essential consequence of each of these processes is to produce a thickened crust, which because of its content of heat-producing elements, causes temperatures to rise, and metamorphism ensues. We will consider the simplest of the crustal thickening processes, that caused by thrust faulting. Other processes could be considered, but their overall effect is similar to that caused by thrusting (England and Thompson, 1984).

Consider a stable continental lithosphere with heat-producing radioactive elements concentrated in a surface layer of thickness D (Fig. 22.7(A)). The heat-generating capacity of this rock, A, remains constant throughout the layer. In the lower crust, the heat-generating capacity is smaller owing to a higher

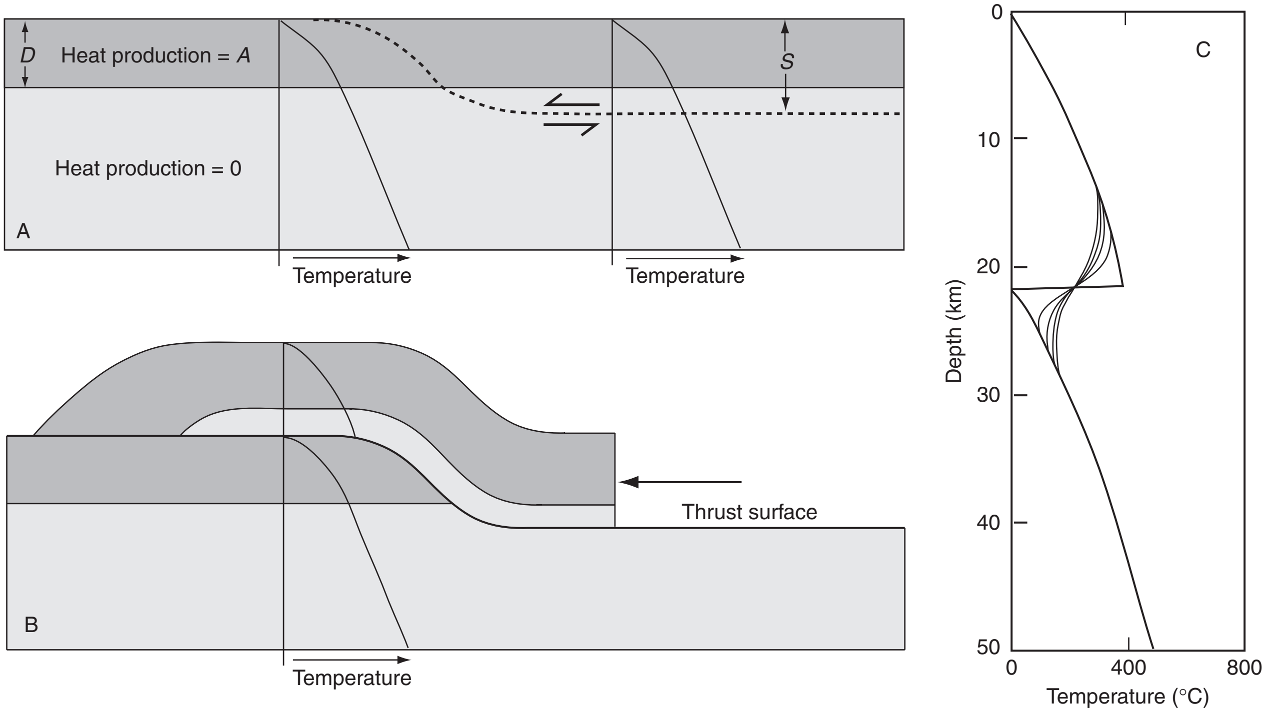


Fig. 22.7 Thickening of continental lithosphere by thrust fault. (**A**) The lithosphere prior to faulting consists of an upper layer of thickness *D*, which contains radioactive elements that have a heat production of *A*; no radioactive elements occur beneath this layer. The dashed line marks the future position of the thrust fault, which penetrates to a depth of *S*. The steady-state continental geotherm is shown in two localities. (**B**) Following thrusting, which takes place rapidly relative to the rate of heat transfer by conduction, the heat-producing layer and its geothermal gradient are repeated by the thrust fault. (**C**) The shape of the geothermal gradient is shown at 50 000-year intervals during the first 200 000 years of relaxation of the perturbation caused by the fault. Later, the temperatures rise as radiogenic heat is produced in the thickened crust and the geotherm continues to readjust (see Fig. 22.8).

proportion of mafic rocks, and processes including extraction of heat-producing elements from felsic rocks by fluids during high-grade regional metamorphism (Rudnick and Fountain, 1995). For simplicity, therefore, we assume that no radiogenic heat is produced below the surface layer at depths greater than *D*. The reduced heat flux into the base of the layer is calculated from the measured heat flux at Earth's surface using Eq. (1.13). The steady-state geothermal gradient is calculated using Eq. (1.16). To do this we must know the thermal conductivity, K_T, of the rocks. Although K_T varies with rock type, no serious error is introduced by assuming that all rocks have a value of 2.25 $\mathrm{W\,m^{-1}\,K^{-1}}$ (England and Thompson, 1984). The resulting geothermal gradient is shown schematically in Figure 22.7(A).

Also shown in Figure 22.7(A) is a dotted line marking the site of a future thrust fault. This line cuts down through the layer containing the heat-producing elements and eventually becomes horizontal at a depth *S*. For illustrative purposes, a second geothermal gradient is drawn so as to show its relation to the future thrust fault. The rates of tectonic processes are orders of magnitude greater than the rates at which heat moves by conduction through rock. Consequently, when the thrust becomes active (Fig. 22.7(B)), the segment of the geothermal gradient above the fault is transported up the fault ramp and along the surface overlying the undisturbed geothermal gradient. The result is that in regions where crust has been thickened by this process, the geothermal gradient initially has a sawtooth shape.

The assumption that a fault is able to move a significant distance before the perturbed geotherms begin relaxing is, of course, a simplification. The sharp thermal gradient across the fault (Fig. 22.7(B) and (C)) causes some heat to transfer almost immediately, with rocks above the fault cooling and those below heating. However, these effects are local and rather small compared with those resulting later from radiogenic heating of the thickened crust. Nonetheless, these early changes can explain some metamorphic reactions near faults (see Karabinos and Ketcham, 1988). Once heat conduction across the fault begins, the sharp thermal perturbation is dissipated almost completely in 5 million years. The relaxation of the thermal gradient across a thrust fault during the first 200 000 years is shown in Figure 22.7(C). We will make the simplifying assumption that the thrusting is complete before relaxation of the faulted geothermal gradient begins (England and Thompson, 1984).

The shape of the geothermal gradient at any time *t* following thrusting can be calculated using the Crank–Nicolson finite-difference technique (Eq. 22.6). The initial conditions are simply the temperatures of the faulted steady-state geotherm. The temperature at Earth's surface remains constant, for example at 0 °C. If the reduced heat flux from the mantle remains constant and equal to that entering the base of the

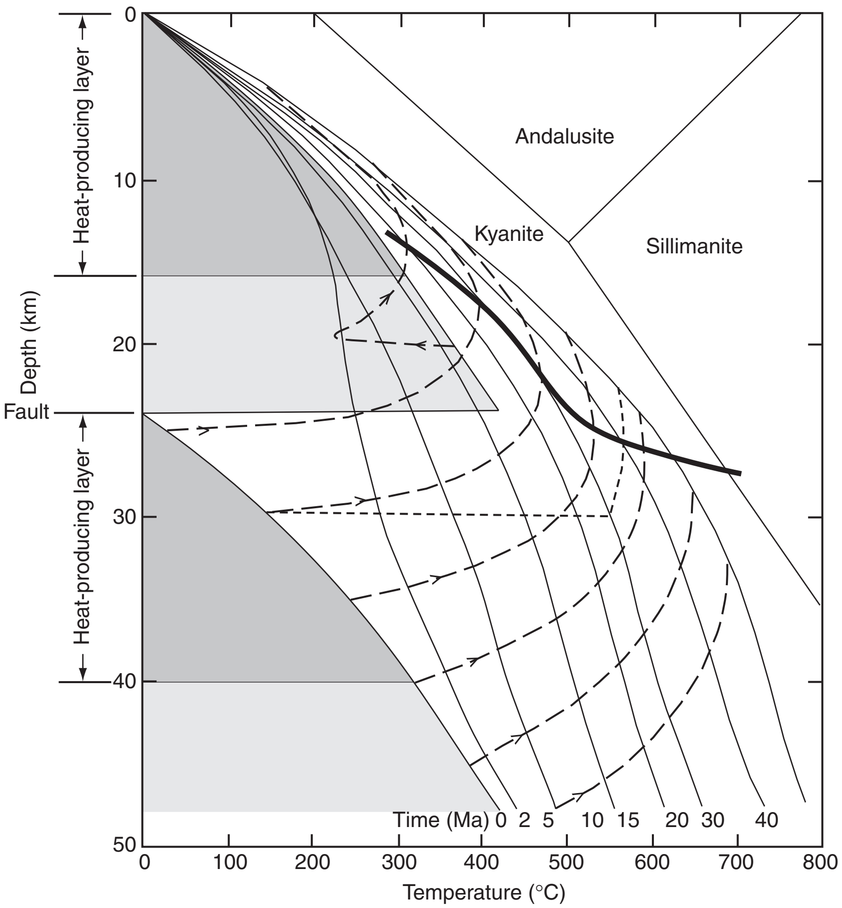


Fig. 22.8 Perturbation of a steady-state geotherm by a thrust fault that repeats a 16-km-thick heat-producing layer. The fault is at a depth of 24 km. Thermal properties of the rocks are specified in text. The erosion rate is 0.4 m per 1000 years. Geotherms are shown for the first 40 Ma after faulting. The *P–T–t* paths of rocks starting at depths of 20, 25, 30, 35, 40, 45, and 50 km are shown with dashed lines. The dotted *P–T–t* path starting at 30 km is for an initial period of 20 Ma in which there is no uplift. The heavy solid line is the field gradient or array of *P–T*$_{max}$ points on *P–T–t* paths; these are the conditions that are likely to be preserved in the rocks returned to the surface. See text for discussion.

heat-producing layer, J_Q^D, then the temperatures in the deepest cells, *p,n* (see Fig. 5.10) are given by

$$T_{p,n} = T_{p-1,n} + J_Q^D \frac{\delta z}{K_\mathrm{T}} \tag{22.7}$$

where δz is the finite difference in depth and K_T is the thermal conductivity. An erosion rate must also be specified, for this is a critical control on rates of exhumation. It is becoming increasingly clear that there are complex feedbacks between tectonics, climate, erosion, and exhumation (Beaumont *et al.*, 2001). England and Richardson (1977) assumed that the erosion rate, d*H*/d*t*, is proportional to the height of land above sea level, *H*; that is,

$$\frac{\mathrm{d}H}{\mathrm{d}t} = -cH \tag{22.8}$$

This equation is similar in form to that for radioactive decay (Eq. 13.2), and on integration [$H = H_0 \exp(-ct)$] gives an exponential erosion rate. In the western Tauern Window of the eastern Alps (Austria and Italy), a well-constrained *P–T* path has been combined with absolute age determinations to show that exhumation rates constantly decreased from a rate of 3.6 mm a^{-1} at the time of the thermal maximum (20 Ma) to 0.1 mm a^{-1}, 13 Ma later (Selverstone, 1985; Blanckenburg *et al.*, 1989). Part of this exhumation history, however, involved tectonic thinning. In contrast, England and Thompson (1984) assumed that as a first approximation, erosion rates can be considered constant; that is,

$$\frac{\mathrm{d}H}{\mathrm{d}t} = E \tag{22.9}$$

where *E* is a constant that possibly has values between 2×10^{-4} and 5×10^{-4} m a^{-1}. Again, for simplicity, we will assume a constant rate of erosional exhumation.

Figure 22.8 shows the solution to Eq. (22.6) for the first 40 Ma following thrusting of a 16-km-thick surface layer that has a heat productivity of 2×10^{-6} W m^{-3}. The thrust fault is at a depth of 24 km. The thermal conductivity of all rocks is 2.25 W m^{-1} K^{-1}, and the thermal diffusivity is 10^{-6} m^2 s^{-1}. The surface heat flow is 0.06 W m^{-2}, so the flux into the base of the heat-producing layer is 0.028 W m^{-2}. The calculation was carried out with time intervals of 0.5 Ma so the depth increments must be 4 km for $k_\mathrm{T}\delta t/(\delta z)^2$ to be equal to unity. The velocity of erosion was taken to be 4×10^{-4} m a^{-1}. Note that this gives a velocity *u* in Eq. (22.6) of -4×10^{-4} m a^{-1}, because *z* is positive downward. Consequently, the last group of terms in Eq. (22.6) becomes positive. To understand fully the following discussion, Problem 22.2 should be worked through slowly and carefully. More will be gained by solving this problem than by reading about it.

First, let us examine the change in the geotherm with time. On the left of Figure 22.8 is the sawtooth pattern of the initial faulted geotherm with the heat-producing layer (darker shading) appearing twice because of faulting. Initially the rocks above the fault cool as heat migrates downward into the cooler rooks below the fault. By 2 Ma, however, the geothermal gradient is everywhere positive; that is, heat is transferred upward. By about 12 Ma, the geotherm is back to where it was prior to thrusting. However, because of crustal thickening and the repetition of the heat-producing layer, the rocks continue heating and the geotherm rises above its initial steady-state position. As the gradient steepens, the heat flux increases (Eq. (5.3)), so the rate of steepening decreases with time. In addition, erosion causes exhumation and gradual thinning of the upper heat-producing layer. Eventually, as the thrust slice is eroded away, the geotherm returns to its initial steady-state form. Figure 22.8 shows geotherms only for the first 40 Ma following faulting, for this is the period during which most significant heating occurs.

The *P–T–t* path of a rock can be followed in Figure 22.8, knowing the rate of exhumation due to erosion (2 km/5 Ma). Consider, for example, the path followed by a rock that starts at a depth of 30 km. Its initial temperature is 150 °C. However, by 10 Ma the rock has risen to a depth of 26 km, and its temperature is over 400 °C, well within the *P–T* range of the greenschist facies (Fig. 16.6). By 20 Ma the rock has risen to 22 km, but the *P–T–t* path has gradually steepened, because of the steady decrease in the rate of temperature rise. Indeed, the *P–T–t* path becomes vertical at this depth, and thus the corresponding temperature of 475 °C is the maximum temperature reached by this rock. With continued exhumation, the rock cools, and the metamorphic assemblage developed at the peak conditions (22 km = about 0.6 GPa and 475 °C) would probably be preserved, indicating upper greenschist facies metamorphism. Note that despite the wide range of *P–T* conditions at the temperature maxima on the various *P–T–t* paths, all rocks pass through a narrow range of *P–T* conditions during cooling.

A rock that starts at a depth of 20 km and 370 °C follows a different type of *P–T–t* path. Because it is above the thrust fault, its temperature decreases for the first 2 Ma, dropping to nearly 200 °C as heat flows downward into the cooler underlying rocks. But then the temperature begins to rise, reaching a maximum after 15 Ma of 310 °C, which is 60 °C less than the initial temperature. During the initial cooling period, retrograde reaction would occur especially if volatiles were liberated by prograde reactions occurring beneath the fault. These retrograde minerals would then undergo prograde reactions as the rock began to heat.

In the discussion above, erosion and exhumation are assumed to begin immediately following thrusting. There is evidence, however, that erosion may be delayed by as much as 20 Ma (England and Thompson, 1984). The increased pressure near the base of the thickened crust may cause high-*P*, high-density mineral assemblages to form that would not induce isostatic rebound. The mountain chain would have low relief at this stage. Following some period of heating, rocks would revert back to the lower-density assemblages, and then erosional exhumation would begin. In this scenario, *P–T–t* paths initially follow a 20-Ma isobaric heating period before the onset of exhumation. Thus a rock starting at a depth of 30 km reaches a temperature of 550 °C before exhumation begins (dotted path in Fig. 22.8). This rock reaches its temperature maximum at 570 °C, approximately 100 °C above the temperature reached by the same rock when exhumation immediately followed thrusting. The effect of an erosional delay, therefore, is to raise the maximum metamorphic temperatures on *P–T–t* paths.

The field gradient recorded in rocks originating at different depths is shown by the heavy line passing through the temperature maxima on the *P–T–t* paths in Figure 22.8. Clearly, this line does not correspond to the geotherm at any given time. Nor are the rocks along the field gradient formed at the same time. The age of the peak metamorphic assemblage becomes progressively younger with increasing grade; that is, the maximum temperature is reached later at successively greater depths. An interesting consequence of this relation is that if rocks at depth reach the minimum melting temperature, magmas would rise into rocks that had already started to cool. This explains the common occurrence of postmetamorphic magmatic bodies in many metamorphic terranes.

Figure 22.8 includes the phase diagram for the Al_2SiO_5 polymorphs. The field gradient produced by our simple model lies almost entirely within the kyanite stability field, except at high temperatures where it passes into the sillimanite field. This field gradient would consequently be characterized by a metamorphic facies series similar to the Barrovian series. Clearly, conditions must be very different if the high-*T*, low-*P*, or low-*T*, high-*P* series are to form. Let us consider how the boundary conditions for Eq. (22.6) might be modified to produce these other facies series.

The high-*T*, low-*P* facies series passes through the andalusite stability field. The geothermal gradient must therefore be increased significantly over that resulting from the boundary conditions for which Figure 22.8 was generated. First, let us consider the effect of erosion rate on the field gradient. The more rapid the erosion, the greater the advective transfer of heat due to uplift of rocks toward the surface (last group of terms in Eq. (22.6)). Geotherms, consequently, rise to higher temperatures at shallower depths. But the more rapid the exhumation, the steeper the *P–T–t* paths become, and the lower is the temperature maximum on these paths. These two effects therefore tend to cancel. Higher temperatures can also be generated by greater thickening of the heat-producing layer. For example, the layer could be repeated by more thrust faults. Problem 22.2 explores the consequences of thickening the layer four times by displacements on three thrust faults. Although this does bring the geotherms closer to the andalusite field, it still does not give high enough temperatures. This leaves the advective transfer of heat by ascending bodies of magma as the most likely way of raising temperatures into the andalusite field (Problem 22.2 shows how this magmatic advection can be taken into account in Eq. (22.6)). Advection of metamorphic fluids could also raise temperatures, as could tectonic processes, such as thrusting and folding, that bodily move hot metamorphic rocks to shallower depths. If advection does not occur, then the only other means of raising temperatures into the andalusite field is to have initial thermal gradients that are much steeper (60 °C km^{-1}) than the steady state geotherm. As far as we know, this requires that the rocks be near a divergent plate boundary, such as an oceanic ridge or continental rift, or a hot spot (Section 1.4). These various possibilities are discussed further in Section 22.7.

Rocks of the low-*T*, high-*P* facies series (LT/HP blueschists and eclogites) form at temperatures below the normal steady-state geotherm. Although the early stages of many *P–T–t* paths pass through this region, continued heating normally destroys these mineral assemblages in favor of those of the other facies series. To prevent this heating, rapid exhumation is necessary. These conditions exist today where oceanic plates are subducted beneath the accretionary wedge that develops at the convergent plate boundary. In such regions, the subduction of cold oceanic lithosphere may, in addition, provide relatively low initial geothermal gradients. Finally, many of the sedimentary rocks forming the accretionary wedge that are converted to blueschists are graywackes derived from nearby island arcs and continental margins, or deep-sea pelagic sediments that lack a large input of material from ancient cratons. They consequently tend to be relatively poor in heat-producing radioactive elements. These factors combine to return rocks to the surface along *P–T–t* paths that do not rise above the normal steady-state geotherm.

22.5 HEAT ADVECTION BY FLUIDS AND MAGMAS

The models of *P–T–t* evolution discussed above do not include the roles that fluid flow or magma intrusion can play in heat transfer. The importance of magmatic heating in contact metamorphism is obvious, but it is less clear in many regional metamorphic terranes. In this section, we will examine some of the thermal consequences of fluid flow and magmatism during metamorphism.

In Chapters 20 and 21 we have seen how fluids affect specific types of metamorphic reactions. We have not, however, considered their effects on the overall metamorphic process. In a regional metamorphic terrane or in a contact aureole, numerous reactions may occur at the same time in different places. Most of these liberate a fluid phase that is likely to migrate upward and affect other reactions occurring above. If fluxes are sufficiently large, the fluid transports heat with it, and in so doing modifies the thermal gradient. The enthalpy of the reactions, which are mostly but not entirely endothermic, may also affect the gradient (Section 22.6). Heat can enter the rock by simple thermal conduction and through advection of a fluid phase.

The treatment of heat transfer by advective fluid flow is similar to that described above in Section 22.4 for mass transport heating. However, for flow, we consider movement of fluid through the rock, instead of movement of the entire rock mass. In order for fluid flow to affect the temperature of the rock mass, there must exist a temperature gradient, $\partial T/\partial z$, in the advective fluid. In a porous medium, the rate of advance of the gradient is directly related to the fluid flux, J_z (Eq. (21.1)). As we saw in Section 21.3, this flux can be written as: $J_z = \bar{v}_z \phi$, where ϕ is porosity and $\bar{v}_z$ is the mean pore fluid velocity in the *z* direction. To obtain the heat flux due solely to advection, we need to multiply the fluid flux by temperature, as well as the density and heat capacity of the fluid

$$J_{z,\mathrm{Heat,adv}} = \bar{v}_z \phi \rho_f C_{\mathrm{P},f} T \tag{22.10}$$

where $J_{z,\mathrm{Heat,adv}}$ is the heat flux due to fluid advection, ρ_f is the density of the fluid, and $C_{\mathrm{P},f}$ is the specific heat capacity of the fluid in J kg^{-1} K^{-1}. Thus the heat flux has units of J m^{-2} s^{-1}. Note that this expression has the same general form as Eq. (22.1). The total heat flux due to both conduction and advective fluid flow through the rock mass, $J_{z,\mathrm{Heat,tot}}$, is given by the sum of Equations (5.3) and (22.10)

$$\begin{aligned} J_{z,\mathrm{Heat,tot}} &= J_{z,\mathrm{Heat,cond}} + J_{z,\mathrm{Heat,adv}} \\ &= -K_{\mathrm{T}} \frac{\partial T}{\partial z} + \bar{v}_z \phi \rho_f C_{\mathrm{P},f} T \end{aligned} \tag{22.11}$$

where K_{T} is the thermal conductivity of the rock (solid + fluid). If the porosity is fairly small, the thermal conductivity can be approximated by that of the solid rock material.

To formulate the energy conservation equation for heat transport incorporating conduction and fluid flow, we can once again make use of Eqs. (22.2) and (22.3)

$$\frac{\partial\left(C_{\mathrm{P,Rock}} \rho_{\mathrm{Rock}} T\right)}{\partial t} = -\frac{\partial\left(J_{z,\mathrm{Heat,tot}}\right)}{\partial z} \tag{22.12}$$

where the subscript Rock explicitly states that the heat capacity and density terms are for the bulk rock (solid + fluid). The numerator on the left-hand side can be expanded in terms of both solid and fluid as: $C_{\mathrm{P,Rock}} \rho_{\mathrm{Rock}} T = (1-\phi) C_{\mathrm{P,Sol}} \rho_{\mathrm{Sol}} T + \phi C_{\mathrm{P},f} \rho_f T$, where the subscript Sol denotes the solid. Note that if the porosity is small, then $\phi C_{\mathrm{P},f} \rho_f T$ is negligible, and $(1-\phi) \approx 1$. In this case, the heat content of the bulk rock is closely approximated by $C_{\mathrm{P,Sol}} \rho_{\mathrm{Sol}}$.

Substitution of (22.11) into (22.12) yields the following expression for energy conservation due to conduction and advective fluid flow

$$\frac{\partial\left(C_{\mathrm{P,Rock}} \rho_{\mathrm{Rock}} T\right)}{\partial t} = -\frac{\partial\left(-K_{\mathrm{T}} \frac{\partial T}{\partial z} + \bar{v}_z \phi \rho_f C_{\mathrm{P},f} T\right)}{\partial z} \tag{22.13}$$

Holding the densities, heat capacities, thermal conductivity, and fluid flux constant gives

$$\frac{\partial T}{\partial t} = k_T \frac{\partial^2 T}{\partial z^2} - \bar{v}_z \phi \Gamma \frac{\partial T}{\partial z} \tag{22.14}$$

in which k_{T} is the thermal diffusivity of the rock mass and Γ is given by

$$\Gamma = \frac{C_{\mathrm{P},f}\rho_f}{C_{\mathrm{P,Rock}}\rho_{\mathrm{Rock}}} \tag{22.15}$$

The Γ ratio will be fairly close to unity for many fluid–rock systems, and is thus not considered in some treatments (e.g. Bickle and McKenzie, 1987). For example, for typical values of $C_{\mathrm{P},f}\rho_f = 3.5\times10^6\ \mathrm{J\ m^{-3}\ K^{-1}}$ (Brady, 1988) and $C_{\mathrm{P,Rock}}\rho_{\mathrm{Rock}} = 2.5\times10^6\ \mathrm{J\ m^{-3}\ K^{-1}}$, the value of Γ is 1.4.

It is useful to point out that Eq. (22.14) for heat transfer has the same form as Eqs. (21.47) and (21.58) for mass transfer (Bickle and McKenzie, 1987). Thus, solutions to problems of mass transfer, such as Eq. (21.48), can also be used for heat transfer (Problem 22.5(c)). For example, for Eq. (21.47), one needs to substitute temperature for concentration, the thermal diffusivity for the coefficient of hydrodynamic dispersion, and the product $\bar{v}_x\phi\Gamma$ for the average pore fluid velocity.

Nondimensional solutions to Eq. (21.14) have been derived by Bickle and McKenzie (1987) for several situations involving thermal gradients through a layer. The terms in the equation are made dimensionless as follows. Depth z is divided by the thickness h of the layer through which heat and matter are transferred ($z' = z/h$); temperature is divided by the temperature at the base of the layer ($T' = T/T_0$); and time is expressed as $t' = tk_{\mathrm{T}}/h^2$. Use is also made of the nondimensional thermal *Péclet number*, which is defined as $\mathrm{Pe_T} \equiv \bar{v}_z\phi h\Gamma/k_{\mathrm{T}}$ (Bickle and McKenzie, 1987; Brady, 1988). This number expresses the relative contributions of advection and thermal diffusion and is thus an important result of the nondimensional analysis. Equation (21.14) can then be written in nondimensional form as

$$\frac{\partial T'}{\partial t'} = \frac{\partial^2 T'}{\partial z'^2} - \mathrm{Pe_T}\frac{\partial T'}{\partial z'} \tag{22.16}$$

We can illustrate the combined effects of thermal conduction and advection on the temperature of metamorphic rocks by considering the steady-state gradient developed in rocks overlying a convecting magma chamber whose roof rocks are kept at a constant temperature T_0 (Fig. 22.9). The roof of the magma chamber is at a depth h below the surface of Earth. If the magma releases no volatiles and no devolatilization reactions take place in the overlying rocks, heat will be transferred through the overlying rocks only by thermal conduction. Consequently, a linear steady-state temperature gradient is established, which in Figure 22.9 is labeled with a zero to indicate that the Péclet number is zero.

If advective fluids emanate from the magma or are evolved by metamorphic reactions, their upward migration deflects the steady-state temperature gradient toward Earth's surface. A solution to Eq. (22.16) for this condition is given by Bickle and McKenzie (1987). The greater the flux of advective fluid, the greater is the Péclet number. Steady-state thermal gradients are shown in Figure 22.9 for Péclet numbers of 2, 4 and 16.

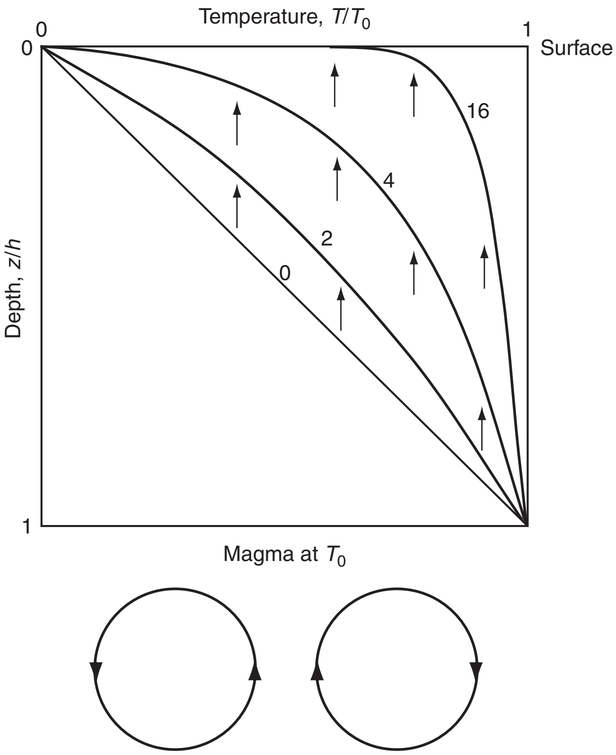


Fig. 22.9 If the roof of a magma chamber is kept at a temperature T_0 by convection, a linear temperature gradient is established in the overlying rocks if there is no vertical advection of fluids. Advection of fluids transports heat toward the surface, with the result that the temperature remains high through much of the thickness of the roof rocks except near the surface where temperatures drop rapidly. The numbers on the curves are Péclet numbers, which express the ratio of advective heat flux to conductive heat flux (see text). Temperatures are expressed in dimensionless values of T/T_0, and depth in terms of z/h, where h is the depth to the roof of the magma chamber. (After Bickle and McKenzie, 1987.)

The effect of advection is to transport heat farther than it would go by conduction alone. This maintains higher temperatures over greater distances. Thus high-temperature metamorphic reactions would be expected at greater distances from the contact for cases where advection has occurred. This would explain why contact aureoles around fluid-rich granitic intrusions are commonly better developed than around hotter, less-fluid-rich mafic intrusions. Although Figure 22.9 is for steady-state gradients above a convecting magma chamber, the form of the curves can be thought of as being generally applicable to other fluid flow settings, regardless of whether or not magmas are present. Interestingly, in an advection-dominated system, the magnitude of the geothermal *gradient* is large near the surface relative to the conductive case, but is actually smaller at depth (Fig. 22.9).

We can further investigate the effects of fluid advection by examining in more detail the thermal Péclet number introduced previously

$$\mathrm{Pe_T} = \Gamma\frac{\bar{v}_z\phi h}{k_{\mathrm{T}}} \tag{22.17}$$

where h is the characteristic length scale of the problem. Note that (22.17) requires knowledge of the fluid flux $\bar{v}\ \phi$. As we saw in Chapter 21, rocks record the time-integrated history of flow and reaction and, as a result, the time-integrated fluid flux (q_{TI}) is normally calculated instead of $\bar{v}_z\phi$. However, if we know the timescale over which flow occurred, then we can estimate the average fluid flux by dividing q_{TI} by the timescale. Thus, (22.17) can be rewritten as

$$\mathrm{Pe_T} = \Gamma \frac{q_{TI} h}{k_T t} \tag{22.18}$$

in which t is the timescale of fluid flow.

Equation (22.18) shows that advective heat transfer will become more important as the fluid flux or length scale of transport increase, or the timescale of flow decreases. Consider a typical regional metamorphic q_{TI} of 1000 $m^3\ m^{-2}$ (Fig. 21.38), a regional length scale h of 10 km, $\Gamma = 1.4$, and $k_T = 10^{-6}\ m^2\ s^{-1}$. If the timescale of flow is 10^7 years, then $\mathrm{Pe_T}$ is only ~ 0.04, which indicates that conduction dominates the heat transfer. As we saw earlier, $\mathrm{Pe_T}$ needs to exceed ~ 2 before the effect of fluid flow on the thermal gradient becomes significant (Fig. 22.9; Brady, 1988). Thus, even though fluxes can be large, fluid advection may play a minor role in transporting heat in many regional metamorphic terranes.

However, if the fluid flux was regionally channelized and attained very large values, then advection could be more important. For example, fluid flow through veins in the amphibolite facies rocks of the Wepawaug Schist, Connecticut, produced an average regional q_{TI} of about $6 \times 10^4\ m^3\ m^{-2}$ (Ague, 1994b). For this case, $\mathrm{Pe_T}$ is ~2.7 for a 10^7-year timescale of flow, which suggests that fluid advection could have transported significant heat. If the timescale of flow was shorter, say, 10^6 years, then $\mathrm{Pe_T}$ would be ~27, indicating an even greater role for advection. The main result would be that temperatures at any given depth would have been higher than normal due to the advective transfer of heat from below (Fig. 22.9; see, for example, Peacock, 1987).

Models such as that shown in Figure 22.7 make several testable predictions about *P–T–t* evolution. One is that rocks of differing metamorphic grade will reach their peak temperatures at different times (Fig. 22.7). Another is that rocks will spend a long time at or near peak temperature conditions. Observations that diverge from these predictions can therefore identify thermal regimes that differ from the general model of conductive relaxation of tectonically overthickened crust during exhumation.

The upper amphibolite to granulite facies "hot spots" in New Hampshire, USA, are one such regime. Ten of these features occur in a belt ~50 km wide and ~150 km long extending north–south across the State. Using the garnet–biotite geothermometer (Eq. (19.5)), Chamberlain and Rumble (1988) found that metamorphic temperatures increase from ~500 °C to > 700 °C over distances of a few kilometers at the hot spot near Bristol, New Hampshire (Fig. 22.10(A)). The GASP geobarometer involving garnet–sillimanite–plagioclase–quartz (Eq. (19.9)) indicates pressures of around 0.4 GPa.

Chamberlain and Rumble (1988) found that the $\delta^{18}O$ values of quartz in rocks in the hot spot area are smaller than those outside of it (Fig. 22.10(B)), and that the hot spot is centered on a network of quartz–graphite veins that contain quartz characterized by similarly small $\delta^{18}O$ values. They concluded that the hot spot thermal anomalies were caused

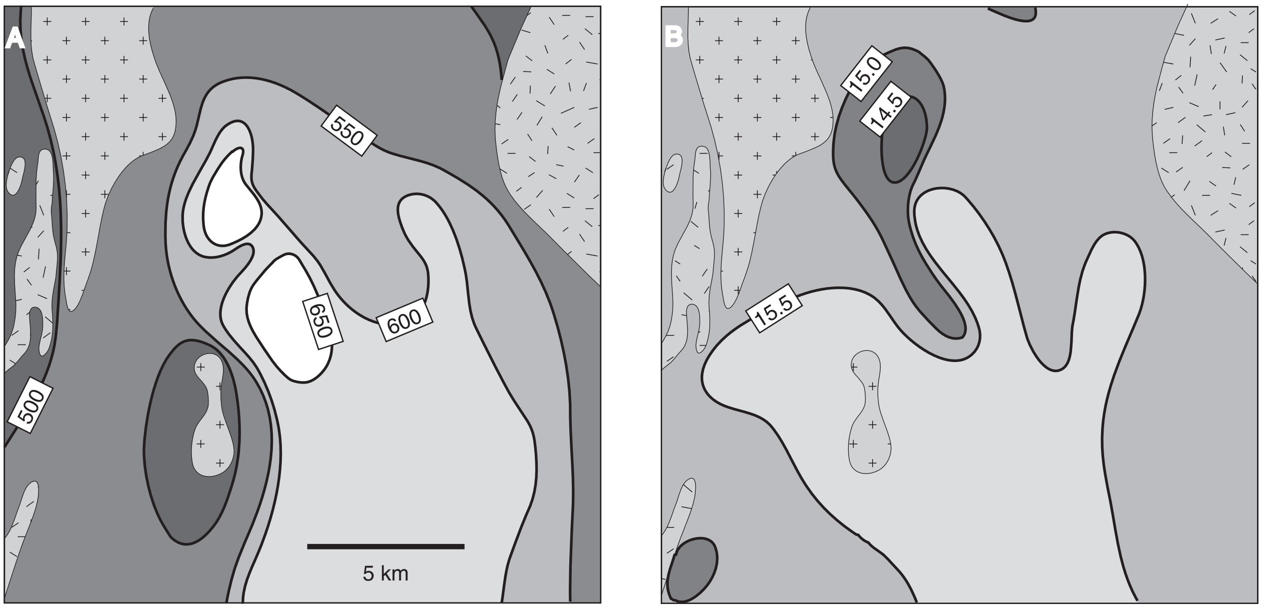


Fig. 22.10 Metamorphic hot spot associated with synmetamorphic quartz–graphite veins near Bristol, New Hampshire. (**A**) Temperature contours based on geothermometry. Maximum temperatures recorded in the core of the hot spot (white areas) can exceed 750 °C. (**B**) $\delta^{18}O$ values for quartz separated from rocks. Note that the thermal high in (A) coincides spatially with a zone of anomalously small $\delta^{18}O$ values. Patterned areas represent igneous intrusions. (After Chamberlain and Rumble, 1988; published by permission of Oxford University Press.)

by the flow of large amounts of hot, isotopically light (small δ^{18}O) fluids through the quartz vein network. The fluids transported heat, precipitated vein quartz, and caused the isotopic alteration in the surrounding wall rocks. In order for such thermal anomalies to be preserved in the field, they would need to be rather widely spaced and short lived, otherwise conduction would operate to wipe out the steep thermal gradients (Brady, 1988). The timescales would have to have been less than ~10^6 years, and Chamberlain and Rumble (1989) calculate that they could have been as short as ~10^5 years (Problem 22.6).

Advective heat transfer can occur by flow of fluid or of magma. In fact, some have argued that the hot spot thermal anomalies represent fossil conduits for the passage of magma, rather than fluid. For either case, however, the timescales must have been short. Furthermore, the δ^{18}O anomaly demonstrates that a substantial amount of external fluid must have flowed through the hot spot area, regardless of whether or not magmas also passed through.

Recent precise Sm/Nd garnet–whole-rock dating in the Barrovian type area, Scotland (Fig. 16.1(B)), shows that peak Barrovian and Buchan metamorphism occurred penecontemporaneously over a geologically short time interval (Oliver *et al*., 2000; Baxter *et al*., 2002). The difference in age between peak Garnet and Sillimanite zone conditions was 2.8 ± 3.7 Ma (2σ), statistically indistinguishable from zero and indicating that the duration of peak conditions was a few million years at most (Baxter *et al*., 2002). Near-simultaneous peak *T* attainment across several metamorphic zones is inconsistent with conduction-dominated models and indicates that heat advection was important. This heat was probably supplied by synmetamorphic intrusions present in the area; circulating metamorphic fluids may have also played a role (Baxter *et al*., 2002).

It is important to be clear about how magmatic heat is transferred. Magmas ascend by advection and are emplaced as dikes, sills, plutons, and other types of intrusive bodies. Although magma flow occurs by advection, the heat would be transferred to the surrounding metamorphic rocks mostly by conduction away from the intrusions and, in fluid-rich environments, by fluid circulation through pores and cracks. In this way, heat transfer by advection of magma, conduction through the rock mass, and fluid flow are linked.

As we saw in Chapter 5, diffusion profiles within minerals can also provide valuable constraints on the timescales of thermal events (see Fig. 5.15). A number of elemental and isotopic systems have diffusion relationships that can be applied to metamorphic rocks (e.g. Wijbrans and McDougall, 1986, 1988; Faryad and Chakraborty, 2005; Carlson, 2006). At the Barrovian type locality, Ague and Baxter (2007) examined Sr elemental diffusion in apatite and Fe–Mg–Ca–Mn diffusion in garnet to quantify peak thermal timescales. The diffusion of Sr in apatite has been calibrated by Watson *et al*. (1985) and Cherniak and Ryerson (1993), and diffusion in garnet has been investigated by many workers, including Ganguly *et al*. (1998), Faryad and Chakraborty (2005), and Carlson (2006).

In the Scottish metasediments, many of the apatites have detrital cores surrounded by metamorphic overgrowths that grew prior to the thermal peak of metamorphism (Fig. 22.11 (A)). The Sr contents of the cores and rims differ (Fig. 22.11 (B)). Thus, over time, diffusion acted to smooth out these compositional heterogeneities (e.g. Fig. 5.15). If the duration of the thermal peak was about the same regardless of the metamorphic grade, then the degree of smoothing will increase with increasing peak temperature because diffusion rates increase markedly with temperature (Fig. 22.11(B)–(D)). In Scotland, temperatures were high enough and timescales were long enough to wipe out Sr compositional variations in apatite from the sillimanite zone (Fig. 22.11(D)). On the other hand, Sr diffusion is very slow below about 450–500 °C, so no significant diffusion is apparent in apatites from very low-grade rocks.

If the diffusion rates, initial elemental Sr concentration profiles, and peak temperatures are known, then the timescale of peak heating can be estimated on the basis of the amount of smoothing that has occurred (e.g. Fig. 5.15). A potential pitfall is that apatite can recrystallize fairly readily (Harlov *et al*., 2005), so care must be exercised to avoid postpeak alteration associated with retrogression. Another problem is that sillimanite zone apatite crystals are homogeneous, so they only place bounds on the minimum timescale of diffusion. However, diffusion rates in garnet are much slower, and can be used to quantify timescales in many high-temperature rocks (Faryad and Chakraborty, 2005; Carlson, 2006).

Using Sr diffusion profiles in apatite for the garnet and staurolite zones, and Fe–Mg–Ca–Mn diffusion in garnet for the sillimanite zone, total thermal peak timescales are estimated to have been only of the order of a few hundred thousand years (Ague and Baxter, 2007; Problem 22.1). The entire orogeny probably involved substantial crustal thickening, and the existing radiometric dates (e.g. Oliver *et al*., 2000; Baxter *et al*., 2002) indicate that it lasted some 10–15 million years. The thermal peak (or peaks) that produced the familiar succession of index mineral zones in southeastern Scotland occurred near the end of this period, and was, according to the diffusion profiles and Sm/Nd dating, very brief.

As discussed above, the brief peak thermal timescales and simultaneous attainment of peak *T* across multiple metamorphic zones are consistent with geologically rapid advective heat transfer by magmas and/or fluids. Exhumation rates were rapid as well. Oliver *et al*. (2000) found detrital Dalradian garnet that has the peak metamorphic age (~ 465 Ma) in 465 ±2.5 Ma sediments eroded from the Highlands. Because the metamorphic age and the sedimentary age are close together, exhumation rates must have been rapid – Oliver *et al*. (2000) estimate rates of ~0.33 cm per year or greater. These transport rates are similar to those of plate tectonic movements; the rapid exhumation would have acted to cool the rocks quickly, thus preserving the diffusion profiles in the minerals.

Evidence for brief thermal pulses has also been found in extensional tectonic environments. For example, ^{40}Ar/^{39}Ar

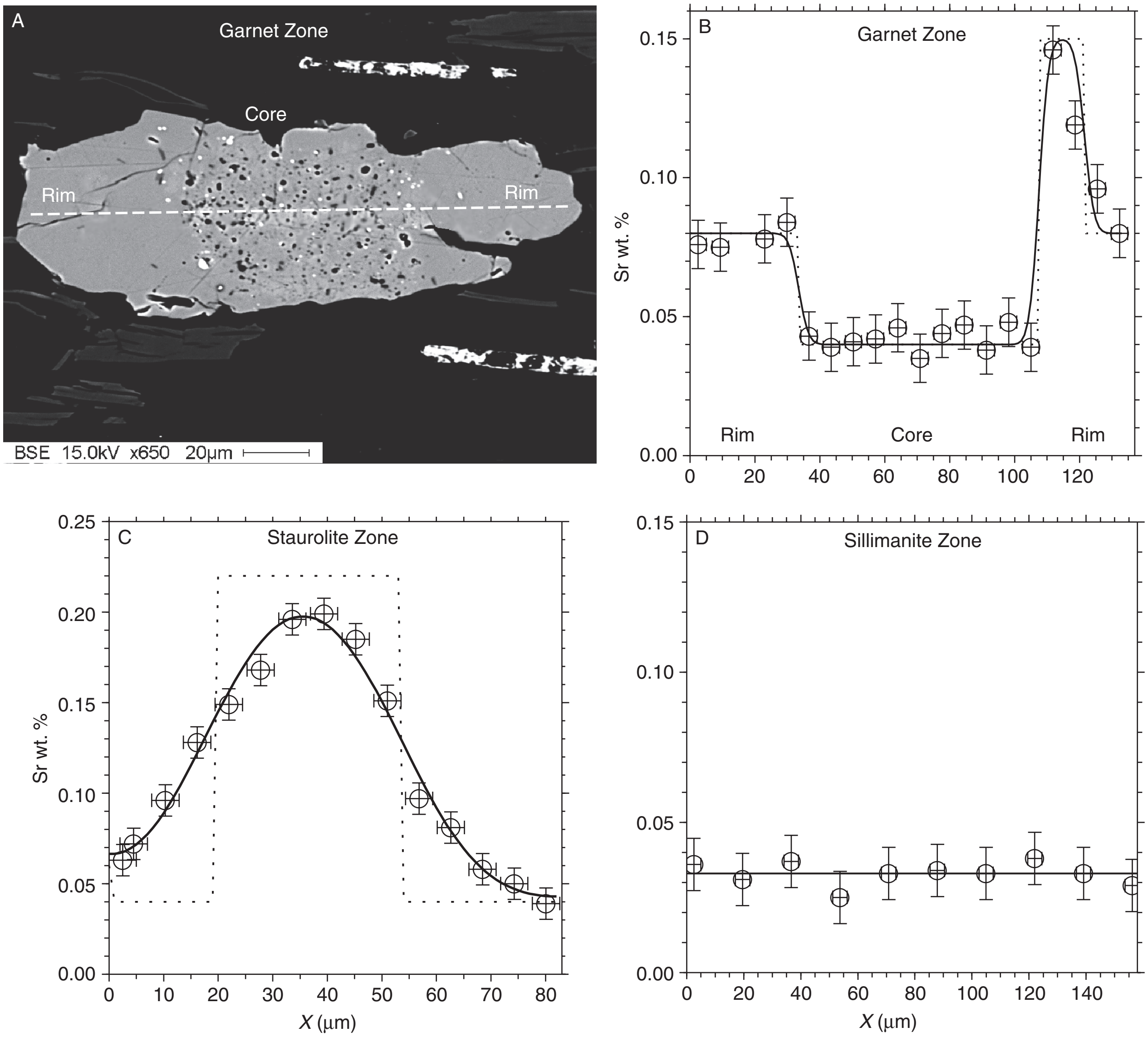


Fig. 22.11 Apatite crystals from the Barrovian zones, Scotland. (**A**) Backscattered electron image of apatite from the Garnet zone. Detrital core is overgrown by metamorphic rims elongate parallel to the foliation. Core contains inclusions of fluid and quartz (dark), as well as ilmenite (light). (**B**) Profile of Sr wt% for the grain shown in part (A). Dotted line is inferred initial, prepeak metamorphic concentration profile, and solid line is the best-fit diffusion model result. Position of traverse is shown with dashed line in part (A). (**C**) Profile of Sr wt% for apatite from the Staurolite zone. Solid and dotted lines as described in part (B). Note increased smoothing of the initial profile as a result of faster diffusion at higher temperatures. (**D**) Sr concentration profile for Sillimanite zone apatite. Rapid diffusion at high temperatures has eliminated any compositional zoning that may have initially been present. (Modified from Ague and Baxter, 2007; published by permission of Elsevier.)

geochronology indicates that some of the metamorphism on the Greek islands of Naxos and Ios was triggered by very brief thermal pulses (Wijbrans and McDougall, 1986, 1988; Lister and Baldwin, 1993). From their studies on Naxos, Wijbrans and McDougall (1988) suggest that the necessary crustal heat input was related to the migration of the Hellenic volcanic arc across the area during the Miocene. Lister and Baldwin (1993) link episodes of mylonitic deformation and brief thermal pulses due to magma intrusion in their analysis of metamorphic core complex formation in Greece and elsewhere. *Metamorphic Core Complexes* are regional-scale, dome-like structures that expose an underlying metamorphic "basement" overlain by less metamorphosed cover rocks. The basement and cover are separated by mylonitic fault zones. Many hypotheses exist to explain core complex formation; most involve significant exhumation of the basement during large-scale crustal extension. Episodic magmatism associated with extension could provide the heat for thermal pulses.

Recent studies suggest that the timescales of fluid flow and heat transport can be extremely short in zones of active deformation. Austrheim (1987) demonstrated that fluid migration along deep, 0.2–5-m-thick shear zones in the Bergen Arcs, Norway, triggered conversion of cold, dry continental crustal rocks to eclogites. The crustal section includes widespread anorthosites and related rocks metamorphosed under granulite facies conditions of 800–900 °C and ~1 GPa that cooled and then were subjected to eclogite facies reheating and metamorphism along shear zones at about 700 °C and 1.8–2.1 GPa (Austrheim, 1987; Jamtveit *et al.*, 1989). The ^{40}Ar/^{39}Ar isotopic results of Camacho *et al.* (2005) are consistent with the hypothesis that conversion of this subducted crust to eclogite occurred due to heat input by multiple fluid pulses along the shear zones over a total timespan of only ~20 ka. Moreover, individual fluid pulses may have lasted just ~10 years! It is also possible that shear deformation contributed to the heating (Section 17.5). Camacho *et al.* argue that if the pulses lasted much longer than 10 years, then the distance of diffusion of excess ^{40}Ar into the rims of K-bearing minerals like phlogopite would be much greater than observed. The authors infer that these extremely short timescales of pulsed fluid release and flow were triggered by earthquakes.

Many of the above studies are relatively recent. Much further work is needed to test hypotheses of brief heat pulses and advective heating, and ideas may change considerably as more is learned. It is likely that regional metamorphic heat transfer spans a large range of timescales from slow conduction-dominated thermal relaxation of overthickened crust, to rapid advective transfer by magmas and/or fluids. It is interesting to note that Barrow (1893) envisioned that the sequence of metamorphic zones he discovered in the Scottish Highlands developed in response to the intrusion of magmas.

22.6 EFFECTS OF REACTION

Thus far we have ignored the heat effects of metamorphic reactions, but now we need to get a better idea of their impact. Basically, the reactions will act as sources or sinks of heat, and so the equation describing temperature evolution with time due to reaction will have the same general form as the equation for radiogenic heating described in Section 22.4. To start, we need a heat production term for reactions in units of W m^{-3} that is analogous to the *A* term for radiogenic heating. Multiplying the enthalpy change of reaction (J mole^{-1}) by the reaction rate (mole m^{-3} s^{-1}) gives the desired quantity. The other factor to consider is that the heat effects of all reactions proceeding in the rock have to be accounted for. With this information in mind, then, we can write

$$\frac{\partial T}{\partial t} = -\sum_l \frac{\Delta H_l R_l^*}{C_{\mathrm{P,Rock}}\rho_{\mathrm{Rock}}} \tag{22.19}$$

where ΔH_l is the enthalpy change for reaction *l*, R_l^* is the reaction rate of *l*, and the sum is over all reactions *l* operating in the rock. The similarity to the radiogenic heat equation should be evident. The minus sign takes into account the thermodynamic convention that endothermic reactions have positive ΔH (heat must flow to the reaction), and exothermic reactions have negative ΔH. The rate term is a bit different than that expressed earlier in Eq. (21.78). In (22.19) we are interested in the rate *per mole of reaction* per unit volume *rock*, and this is given by: $R^* = (\quad)\ \ A\ \ |\Delta G\,|^{N_l}$, where is the rate constant, $A_{p,l}$ is the rate-limiting reactive surface area of mineral *p* per volume rock in reaction *l*, $|\Delta G_l|$ is the absolute value of the Gibbs free energy change for reaction *l*, N_l is the reaction order, and (sgn) is, by convention, +1 if ΔG_l is negative and −1 if ΔG_l is positive.

It is clear that the effect on temperature will be greatest for reactions with large ΔH that proceed rapidly, and intuitively this makes sense. But in order to use Eq. (22.19), we need values for the variables involved. For simplicity, we can study the case in which only one reaction is operating ($l = 1$). As before, we can take $C_{\mathrm{P,Rock}}\rho_{\mathrm{Rock}}$ to be 2.5×10^6 J m^{-3} K^{-1}. For a typical dehydration reaction, representative values for κ_l and ΔH_l are 10^{-4} mole m^{-2} a^{-1} $(\mathrm{J}\cdot\mathrm{mole})^{-N_l}$ (Problem 21.11) and 50 000 J mole^{-1}, respectively. The rate-limiting surface area will typically be given by a mineral in low abundance with the slowest dissolution or precipitation kinetics. If we assume the rate-limiting mineral forms cubes 2 mm on each edge that occupy 10 volume % of the rock, then $A_{p,l}$ is 300 m^2 m^{-3}. Note that these types of surface area calculations have large uncertainties because the simple geometries do not take account of the complexities of real mineral surfaces. We can assume the reaction is proceeding close to equilibrium and use a small value for the ΔG_l of overstepping equal to −1 J mole^{-1}. If the kinetics are linear ($N_l = 1$), then the rate R_l^* works out to be 3×10^{-2} mole$_{\mathrm{(reaction)}}$ m$^{-3}{}_{\mathrm{(rock)}}$ a^{-1}. From Eq. (22.19), then, we calculate that the change in temperature due to reaction is -6×10^{-4} K a^{-1}. In real rocks, heat would be transported into the reacting area by processes like conduction, so the actual temperature change would be smaller (Problem 22.8).

The relative effects of reaction and radiogenic heating can be examined by forming the ratio: $\left(\Delta H_l R_l^*\right)/A$. Taking $A = 2 \times 10^{-6}$ W m^{-3} = 63 J m^{-3} a^{-1} and, based on the above discussion, $\Delta H_l R_l^*$ = 1500 J m^{-3} a^{-1}, the absolute value of the ratio turns out to be about 24. Thus, the results of these simple calculations indicate that heat effects due to reaction can be considerably larger than radiogenic heating. Of course, the value of κ_l decreases exponentially with temperature, so the impact of reaction would be lessened at lower metamorphic grades. For example, the value of κ_l used above is for high temperatures (~ 600 °C). For an activation energy of 80 kJ mole^{-1}, κ_l would decrease to 5×10^{-6} mole m^{-2} a^{-1} $(\mathrm{J}\cdot\mathrm{mole})^{-N_l}$ at 400 °C (see Eq. (21.77)). Under these circumstances, $\left(\Delta H_l R_l^*\right)/A$ decreases to ~ 1.2. In addition, if the reaction was proceeding very close to equilibrium, ΔG_l could be much smaller and the rate correspondingly slower. On the other hand, the amount of heat consumption would increase if the reactive surface area was larger, or if the degree of overstepping increased due to, for example, fluid pressure loss associated with hydrofracturing (Fig. 21.20).

The impact of reactions is perhaps best illustrated at invariant points, such as the various isobaric invariant points examined in Chapter 20 (Fig. 20.9). If the system is fairly close to chemical equilibrium, temperature will stay nearly constant as heat input into the system is consumed to drive the overall reaction progress. To see this, it is necessary to consider heat transport as well as Eq. (22.19). Up until now, we have examined various combinations of heat transport and reaction terms, but here we can go ahead and combine them into a single expression, using Eq. (22.2), that describes conduction, mass transport heating, fluid advection, radiogenic heating, and chemical reaction. The resulting expression is not for the faint of heart

$$\frac{\partial T}{\partial t} = k_{\mathrm{T}}\frac{\partial^2 T}{\partial z^2} - u\frac{\partial T}{\partial z} - \bar{v}_z\phi\Gamma\frac{\partial T}{\partial z} + \frac{A}{C_{\mathrm{P,Rock}}\rho_{\mathrm{Rock}}} - \sum_l \frac{\Delta H_l R_l^*}{C_{\mathrm{P,Rock}}\rho_{\mathrm{Rock}}} \tag{22.20}$$

Equation (22.20) can be enjoyed even further by writing out the analogous expression for the full three-dimensional case.

As in previous examples, however, we don't need to include all of the terms to gain a conceptual understanding. Let us consider invariant point reaction for the case where the radiogenic heat production is negligible and heat is transported by conduction and fluid advection. In the idealized invariant point case, the time derivative of temperature will be zero. Thus, (22.20) reduces to a steady-state problem

$$k_{\mathrm{T}}\frac{\partial^2 T}{\partial z^2} - \bar{v}_z\phi\Gamma\frac{\partial T}{\partial z} = \sum_l \frac{\Delta H_l R_l^*}{C_{\mathrm{P,Rock}}\rho_{\mathrm{Rock}}} \tag{22.21}$$

This expression simply states that the amount of heat transported by conduction and fluid advection into the reacting zone is balanced by the amount of heat absorbed in the overall reaction progress. The steady-state temperature will be maintained until a reactant is used up.

An inventory of the total heat taken in by a rock as a result of temperature change and reaction can also be made. We need to consider the heat needed to change the temperature of the rock, as well as the heat consumed by all the reactions. The first term is simply the product of the temperature change and the heat capacity and density of the rock (cf. Eq. (7.12)).

For the reaction term, we need to relate the reaction rate to the overall reaction progress during the metamorphic episode. To do this, we make use of the fact that $\mathrm{d}\xi_l = R_l^*\,\mathrm{d}t$, where ξ_l is the reaction progress for reaction *l*. The total reaction progress can be found by integrating this expression to obtain ξ_l, but we can also measure it from natural samples. As noted in the discussion of Eq. (21.73), $m_{s,l} = n_{s,l}\xi_l$, where *s* is a fluid species of interest. In Chapter 21 we were concerned with fluids, but here we can generalize *s* to represent solids or fluids and say that and $n_{s,l}$ is the stoichiometric coefficient of *s* in reaction *l* (positive for products, negative for reactants), and $m_{s,l}$ is the total number of moles of *s* produced or consumed by reaction *l*. For example, for the wollastonite-producing reaction (20.32) discussed in Section 20.6, the reacted rock contains 70 modal% wollastonite, which corresponds to 700 cm^3 per liter of rock, or 17.53 mole liter^{-1} (= 17 530 mole m^{-3}). Because the stoichiometric coefficient for wollastonite in reaction (20.32) is one, the reaction progress is simply: $\xi_{20.32} = 1.753\times10^4/1 = 1.753\times10^4\ \mathrm{mole}_{(\mathrm{reaction})}\ \mathrm{m}^{-3}{}_{(\mathrm{rock})}$.

So to obtain the heat taken in or given out by a reaction, we multiply the enthalpy change of the reaction by the total reaction progress. Furthermore, as discussed previously, the reaction heat budget needs to consider all the reactions proceeding in the rock. Thus, for the total amount of heat taken in by a rock during metamorphism incorporating the effects of temperature change and reaction we write

$$Q_{\mathrm{Total}} = Q_{\mathrm{In}} - Q_{\mathrm{Out}} = C_{\mathrm{P}}\rho\Delta T + \sum_l \Delta H_l\xi_l \tag{22.22}$$

where the sum is over all reactions *l*. For the example wollastonite-producing reaction, the enthalpy change is 7.9×10^4 J mole^{-1}, so the reaction term in Eq. (22.22) is very large and equals 1.38×10^6 kJ m^{-3}.

Ferry (1983) studied changes in the modal abundances of minerals in impure carbonate rocks resulting from prograde metamorphism from the Biotite through the Diopside zones in south-central Maine. The total amount of heat added to the rock calculated using Eq. (22.22) was about 1.25×10^6 kJ m^{-3}; of this, the amount consumed by reactions was 9.4×10^5 kJ m^{-3} (rightmost term in Eq. (22.22)). This 9.4×10^5 kJ m^{-3} would have been sufficient to raise the temperature of the rock about 300 degrees had there been no reaction. In fact, Ferry (1983) calculated that most of the heat input to the rocks drove reaction, rather than the prograde increase in temperature. This example illustrates the impact reactions can have on metamorphic conditions.

To conclude, it is likely that the geothermal gradient in an active metamorphic terrane will be more complex than those in Figures 22.8 and 22.9. The heat absorbed by devolatilization reactions may in some cases outweigh that produced by radiogenic heating, at least locally in reacting layers. The geothermal gradient could actually be quite irregular, particularly if thermal buffering of isobaric invariant points produces thermal plateaus.

22.7 OBSERVED *P–T–t* PATHS

A major ongoing challenge in petrology is to determine the *P–T–t* paths of metamorphic rocks and mountain belts. The records are constructed from a variety of data, including mineral assemblages of known *P–T* range, geothermometers and geobarometers based on exchange reactions, zoned porphyroblasts, retrograde reactions, geochronology, and fluid inclusions (see Spear, 1993). The record commonly has large uncertainties associated with it, and at elevated temperatures high reaction rates may erase the record down to *blocking temperatures*, where slowness of diffusion prevents further reaction. In this section we survey briefly the range of *P–T–t* paths that have been identified and discuss their tectonic significance. This is done with reference to Figure 22.12, which, in addition to showing *P–T–t* paths, includes the *P–T* ranges of the various metamorphic facies, the steady-state continental geotherm (Section 1.6), and the solidus for a muscovite granite under conditions of both excess H_2O and no additional H_2O (Huang and Wyllie, 1973).

The thickening of the crust that accompanies the collision of lithospheric plates first causes pressures to rise without large amounts of heating. During this period, many rocks

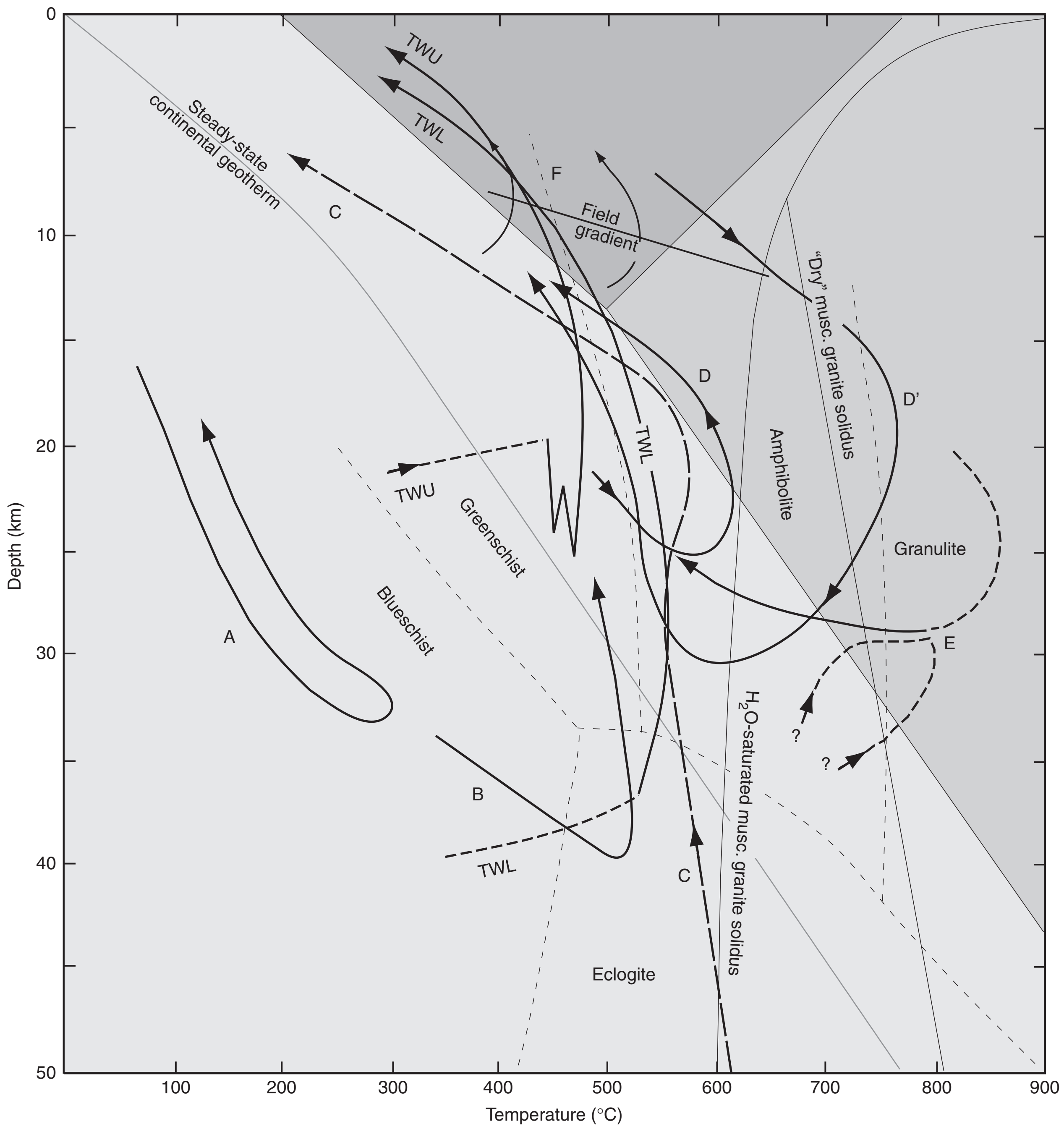


Fig. 22.12 *P–T–t* paths of rocks belonging to the three main metamorphic facies series. Also shown is a steady-state continental geotherm (see Section 1.6), general *P–T* range of metamorphic facies, the Al_2SiO_5 phase diagram (Holdaway, 1971), and the H_2O-saturated and "dry" solidus of muscovite granite (Huang and Wyllie, 1973). Stability fields of the Al_2SiO_5 polymorphs shown with gray shading. Low-temperature/high-pressure (LT/HP) rocks of the Franciscan Complex, California (*A*), and the western Alps (*B*) (Ernst, 1988); *C*, Dora Maira ultrahigh pressure rocks (dashed for clarity; Rubatto and Hermann, 2001; see Fig. 22.13); *D* and *D'* from central Massachusetts (Tracy and Robinson, 1980); *E*, granulites from the Adirondacks, New York State (Bohlen *et al*., 1985); *F*, Trois Seigneurs Massif, Pyrenees (Wickham and Oxburgh, 1985). TWL and TWU, lower and upper stratigraphic units in the Tauern Window, Eastern Alps (Selverstone *et al*., 1984; Selverstone and Spear, 1985). See text for discussion.

pass through the low-temperature/high-pressure (LT/HP) blueschist and eclogite facies stability fields, but only where one of the plates is oceanic are these rocks likely to be preserved. Ernst (1988) recognizes two general types of *P–T–t* path for these rocks, depending on whether the rock remains at LT/HP conditions during exhumation (path *A* in Fig. 22.12) or passes into the greenschist or amphibolite facies fields (path *B*, Fig. 22.12), in which case retrograde greenschist facies minerals rim the LT/HP minerals.

Continuous subduction of cold oceanic lithosphere produces steady cooling of the rocks immediately above the subduction zone. Consequently, progressively higher *P/T*

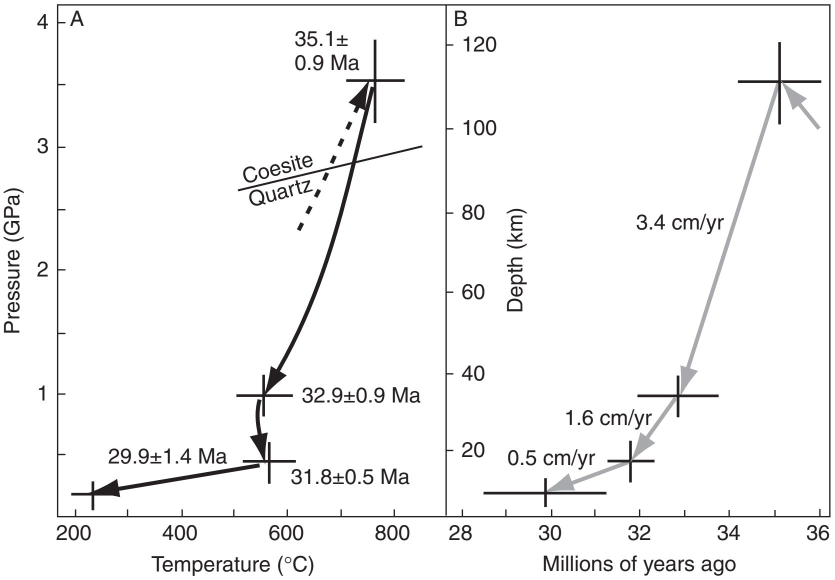


Fig. 22.13 Exhumation of rocks from the Dora Maira Massif, Western Alps. (**A**) *P–T–t* path. Crosses indicate uncertainties on *P–T* estimates. Three oldest ages determined using U/Pb dating of titanite by Rubatto and Hermann (2001); youngest age is a fission-track age from Gebauer *et al.* (1997). (**B**) Depth–time plot showing exhumation rate estimates. (After Rubatto and Hermann, 2001; published by permission of the Geological Society of America.)

ratios are produced, with eclogites forming at the greatest depths. These rocks must be returned to Earth's surface if we are to examine them. Although composed of high-pressure mineral assemblages, LT/HP rocks have lower densities than mantle rocks into which they are subducted. The LT/HP rocks may therefore decouple from the descending slab and buoyantly rise to the surface, almost retracing their path of descent. Platt (1986) has suggested that this happens when the accretionary wedge that builds at a plate boundary becomes overthickened by underplating of material onto the upper plate. To maintain equilibrium between the gravitational forces acting on the surface slope and the drag exerted by the subducting plate, the overthickened wedge develops listric normal faults that dip toward the oceanic plate and penetrate to depths where they intersect the LT/HP rocks. Slices of these rocks are then transported upward and oceanward on the toes of these faults. The process is analogous to the way in which snow, building up in front of a snowplow, will slump forward when it becomes too deep. Regardless of the actual mechanism by which LT/HP rocks rise, their proximity to the cold descending plate in which endothermic dehydration reactions are occurring keeps the rocks "refrigerated" during ascent and preserves the LT/HP mineral assemblages. The resulting *P–T–t* path, which is exemplified by the rocks of the Franciscan Complex of California, is shown by path *A* in Figure 22.12.

If during subduction continental crust or an island arc with silicic rocks descends the subduction zone, the rapid, buoyant rise of such material would occur at almost constant temperature, or perhaps even initially slightly increasing temperatures if the silicic material contains significant amounts of radioactive elements. In this case, LT/HP rocks would rise into the stability fields of the greenschist or amphibolite facies and develop retrograde minerals characteristic of those facies. The resulting *P–T–t* paths (*B* and *C*, Fig. 22.12), for example, are found in LT/HP rocks of the Western Alps.

The combination of geochronology with *P–T* estimates made using thermobarometry provides powerful new constraints on the *P–T–t* paths taken by natural samples. This approach is particularly well illustrated by recent studies of the exhumation of subducted high-pressure (HP) and ultra-high-pressure (UHP) rocks (Fig. 22.13). For example, Baldwin *et al.* (2004) used U/Pb methods to date zircons in HP/UHP eclogites from eastern Papua New Guinea. They found that the rocks contained very young zircons, which grew during eclogite facies metamorphism at 4.3 ± 0.4 Ma. Thermobarometry indicates that the rocks underwent peak metamorphism at ~900 °C and 2.0–2.4 GPa. Thus, the rocks must have traveled some 75 kilometers from the depths of peak metamorphic burial to the surface. Consequently, exhumation rates were rapid – at least 1.7 cm a^{-1} (Baldwin *et al.*, 2004). For another example, Rubatto and Hermann (2001) used U/Pb geochronology to date multiple generations of titanite in UHP rocks from the Dora Maira Massif, Western Alps. This information, combined with *P–T* estimates from thermobarometry, yields rapid exhumation rate estimates of 1.6 to 3.4 cm a^{-1} (Fig. 22.13).

These studies indicate that the HP/UHP rocks were exhumed at "plate tectonic" rates. In other words, the rates are comparable to those of large-scale plate movements on Earth. The processes that drive such exhumation are topics of active research. Many models hypothesize that buoyancy contrasts between the deeply subducted continental material and the surrounding mantle play an important role. In detail, however, the exhumation processes may be very complex; for example, Baldwin *et al.* (2004) concluded that extension associated with very rapid oblique plate convergence was necessary to bring the rocks they studied to Earth's surface.

When convergent lithospheric plates involve continental collisions, the thickened crust may initially pass through the blueschist facies field, but with the production of radiogenic heat the path soon enters the greenschist and amphibolite facies fields. A Barrovian sequence of rocks therefore develops and, if heat transfer is dominated by conduction and exhumation, the greenschists form along shallow, and the amphibolites along deeper, *P–T–t* paths. Isograds plot as points along the resulting field gradient. It is interesting to note that although reactions can be written relating the mineral assemblages on either side of an isograd, the *P–T–t* paths followed by the rocks are very different from the *P–T* array of the field gradient, and thus the rocks may never have undergone the deduced reactions.

A well-documented *P–T–t* path for a region involving continent–continent collision is that for the southwestern Tauern Window in the eastern Alps (Selverstone *et al.*, 1984; Selverstone and Spear, 1985; Blanckenburg *et al.*, 1989). Final metamorphic equilibration temperatures are indicated by the garnet–biotite geothermometer (Eq. (19.5)) (Ferry and Spear, 1978; Hodges and Spear, 1982), and metamorphic pressures are determined from the compositions of garnet and plagioclase coexisting with kyanite and quartz (Eq. (19.9) (Ghent *et al.*, 1979) or biotite and muscovite (Ghent and Stout, 1981; Hodges and Crowley, 1985). *P–T* conditions prior to peak metamorphic temperatures are obtained from zoned garnets according to the method of Spear and Selverstone (1983; see Section 19.7). In addition, epidote–plagioclase–chlorite pseudomorphs after lawsonite provide textural evidence of an early stage of metamorphism in the lawsonite–albite–chlorite subfacies of the blueschist facies. Finally, fluid inclusion data constrain the *P–T–t* path during exhumation.

Two *P–T–t* paths are shown in Figure 22.12 for the Tauern Window rocks, one for the lowest stratigraphic unit exposed (TWL) and the other for the uppermost unit (TWU). Both paths begin in the blueschist facies, the lower unit being at a pressure of >1 GPa (35 to 40 km) and the upper one at a pressure of > 0.7 GPa (25 to 30 km). The conditions in these two units at the peak of metamorphism were 550 °C and 0.7 GPa (25 km) and 475 °C and 0.55 GPa (20 km) respectively. Garnet zoning patterns in the upper unit reveal four distinct phases of tectonic activity (see sawtooth nature of *P–T–t* path in Fig. 22.12). Although the lower and upper units were initially separated by a vertical distance of 10 km, fluid inclusion data reveal that this distance had been reduced to less than 2 km by the time the temperature had decreased to 375 °C. This indicates that tectonic thinning operated, as well as erosion.

Following the collision and subduction of the European plate beneath a continental fragment to the south, pressures in the rocks increased, first as a result of imbrication of subducted units, and then in response to the development of the Austroalpine nappes. Uplift and arching of the lower unit of the Tauern Window resulted in ductile deformation and thinning of the overlying rocks at the same time as the rocks were being exhumed by erosion (Selverstone, 1985).

Differences in the *P–T–t* paths of rocks involved in continental collisions can be due to differences in the types of deformation and the position of rocks relative to major structures, such as thrust faults and nappes. Tracy and Robinson (1980), for example, have shown from careful analysis of zoned garnet crystals (Tracy *et al.*, 1976) that in central Massachusetts rocks within a few kilometers of each other followed very different *P–T–t* paths because of their structural positions. These rocks were deformed during the Devonian Acadian orogeny, first, into a series of nappes overturned to the west, followed by a period of backfolding to the east, and finally, a doming stage, which saw the formation of the Bronson Hill anticlinorium and, immediately to its east, the Merrimack synclinorium (Figs. 19.11, 19.12). The early nappes are highly deformed and are found in both the anticlinorium and synclinorium. During nappe building, cold near-surface rocks were buried beneath large recumbent folds, which resulted in high pressures (kyanite field). Toward the top of these same nappes, relatively hot rocks, transported from depth, were exposed to low pressures (andalusite field). Subsequently, during the backfolding and doming stages, the rocks that had been deeply buried rose as they became hotter due to radiogenic heat and relaxation of the geotherm, producing a counterclockwise *P–T–t* path (*D*, Fig. 22.12). At the same time, the high-*T*, low-*P* rocks sank into the Merrimack synclinorium, tracing a clockwise *P–T–t* path (*D*′, Fig. 22.12). Thus, in this region the *P–T–t* paths were determined largely by the deformation paths.

Note that care must be exercised when describing the direction of a *P–T–t* path. For example, a clockwise path on Figure 22.12 would switch to counterclockwise if plotted on a graph with pressure and depth increasing *upward* along the *y*-axis (cf. Fig. 19.11, 22.13). In fact, the latter convention, illustrated in Fig. 19.11, is the one most commonly used in the petrology literature.

In many regionally metamorphosed terranes, granulite facies rocks record the highest metamorphic conditions attained. Temperatures rise well above 700 °C, high enough for significant diffusion to take place in garnet. Porphyroblasts, consequently, are often homogeneous, and no record of the early part of *P–T–t* paths is preserved (Fig. 19.6(C)). Retrograde reactions, however, provide evidence of conditions following the peak of metamorphism (Fig. 19.6(C)).

Much work remains to determine granulite facies and ultrahigh-temperature *P–T–t* paths. However, two kinds of path are commonly recognized: near isothermal decompression and near isobaric cooling (Harley, 1989). Most examples of the former are consistent with thermal evolution of overthickened crust during exhumation by erosion and, in some cases tectonic extension. Heat input from mantle magmas is important in some areas as well. Near isobaric cooling paths are more difficult to explain; many alternatives have been advanced (e.g. Harley, 1989), and a few are reviewed below.

Many granulite terranes, such as the Grenville age Adirondacks of New York State (Bohlen *et al.*, 1985), exhibit a 200 to 300 °C interval of almost isobaric cooling from peak metamorphic conditions before exhumation begins; this results in *P–T–t* paths that are concave toward the temperature axis (*E*, Fig. 22.12). Such paths may be spurious indicators of *P T* conditions because of the ease with which granulite facies mineral compositions readjust themselves. If they are valid, however, how could cooling occur without simultaneous unloading? A common theme postulated by many models is that, at the time of cooling, the crust was already near normal thickness so that isostatic forces were unable to drive uplift of rocks and exhumation. Additional possibilities for cooling without the necessity of major exhumation include the removal of partial melts or the relaxation of an advective thermal pulse or pulses (Problem 22.4, 22.5).

P–T conditions in the granulite facies exceed the minimum melting temperatures of muscovite-bearing quartzofeldspathic rocks. Figure 22.12 includes the solidus of H_2O-saturated muscovite granite and of muscovite granite containing no additional H_2O (Huang and Wyllie, 1973). Because of the low porosity of high-grade metamorphic rocks, the amounts of H_2O in granulites are not expected to be large. Thus when *P–T–t* paths cross the H_2O-saturated solidus, only small amounts of melt are formed, and not until temperatures rise above the "dry" solidus do amounts become substantial. Nonetheless, the importance of these first-formed melts lies in their strong affinity for the radioactive heat-producing elements. The buoyant rise of the melts depletes the granulite zone in these elements. Thus, following the initial thickening of the crust, temperatures would be expected to rise steadily until sufficient melt was formed to produce buoyantly rising bodies of magma that would take with them the heat-producing elements; at this point the rocks would begin to cool, but not as a result of vertical motion.

The nearly isobaric cooling of some granulite rocks could be a consequence of the influx of hot material; that is, the cooling path reflects the relaxation of a thermal pulse or pulses, which raised the temperature gradient above that which could be sustained by the production of radiogenic heat and the background mantle heat flux. Pulses of magma from the mantle could certainly do this, although many granulite terranes are devoid of large mafic bodies. Such bodies could, however, underplate the continental crust (Bohlen, 1987; Bohlen and Mezger, 1989). If magmas were intruded or underplated early enough in the tectonic cycle prior to or during loading and burial, they could result in clockwise *P–T–t* paths on Figure 22.12 (Bohlen, 1987; paths would be counterclockwise on a graph with pressure and depth increasing upward along the *y*-axis). Peak temperatures would be reached while pressures were increasing, in contrast to overthickening scenarios that lack magmatism (Fig. 22.12). It has also been suggested that some granulite metamorphism results from the flushing of rocks with CO_2-rich fluids from the mantle (Newton *et al*., 1980; Frost and Frost, 1987). The general anhydrous nature of granulites is then explained as much by fluid composition as it is by high temperature. Fluids rising from the mantle could advectively raise the temperature of granulites. Then, when the flux of fluid decreases or stops, the rocks would start cooling as they relax back to the gradient that existed prior to the flushing. The cooling would be enhanced if the fluids stripped heat-producing elements from the rocks. This then provides another means of cooling the rocks without the necessity of unloading.

The high-*T*, low-*P* facies series undoubtedly owes its elevated temperatures to the emplacement of magma (Miyashiro, 1961) and can be thought of as contact metamorphism on a regional scale. In northern New England, for example, metamorphic rocks, which belong to this facies series, increase in metamorphic grade toward the ubiquitous granitic intrusions, and metamorphic isograds are subparallel to igneous contacts (Lux *et al*., 1986). In some places, isograds are overprinted by others that parallel contacts with later intrusions. Lux *et al*. (1986) have shown that the *P–T–t* paths of these rocks can be successfully modeled by thermal pulses emanating from gently dipping sill-like bodies of granitoid rocks.

Although the high temperature in these rocks can be attributed directly to the proximity of igneous intrusions, it does not account for the ultimate source of the heat causing melting. What tectonic conditions give rise to the high-*T*, low-*P* facies series? Oxburgh and Turcotte (1971) have modeled the thermal budget associated with convergent plate boundaries and shown that the paired metamorphic belts of Miyashiro (1961), that is, low-*T*, high-*P* and high-*T*, low *P*, are a natural consequence of this type of plate motion (see Fig. 21.1 and Chapter 23). The low-*T*, high-*P* series forms nearest the oceanic trench where temperatures are below normal because of the subduction of cold oceanic lithosphere. With continued subduction and heating, melting takes place, and ascending magma bodies transfer heat into the upper crust to produce the high-*T*, low-*P* facies series. The ascent of magma also causes crustal extension and may be responsible for the spreading that is common behind island arcs.

Oxburgh and Turcotte's model calls for crustal thickening, but in some areas high-*T*, low-*P* metamorphic rocks are formed in very thin crust. The Chugach metamorphic complex of southern Alaska is one such region in which seismic refraction studies show the subducted oceanic plate to be no deeper than 10 km. In this area, melting of the accreted sedimentary wedge may have resulted from the subduction of very young, hot oceanic crust. The upward transfer of heat was caused by both ascending magmas and fluids liberated by the subducted sediments and ocean floor rocks (Sisson and Hollister, 1988).

In some regions, high-*T*, low-*P* metamorphism is not closely associated with igneous rocks. In the Big Maria Mountains of southeastern California, for example, massive wollastonite is developed in siliceous limestone on a regional scale (Hoisch, 1987). The fluids were focused into the limestones, and the minimum volume ratio of fluid to rock is estimated to have been 17:1 from Eq. (20.32), according to the method of Rice and Ferry (1982) (Section 20.6). Hoisch concludes, therefore, that the high temperatures recorded by these rocks resulted from heat introduced by large fluid fluxes.

One of the largest developments of high-*T*, low-*P* rocks is in the Pyrenees. The metamorphic field gradient and schematic *P–T–t* paths for these rocks is shown as *F* in Figure 22.12. Although this metamorphic belt has been interpreted as resulting from Hercynian mountain building, it lacks many of the characteristics of mountain belts formed at convergent boundaries (Wickham and Oxburgh, 1985). It lacks metamorphosed ocean floor rocks and has no large-scale nappes and thrusts. All of its rocks belong to the high-*T*, low-*P* series; that is, there is no evidence of high pressures during metamorphism. Wickham and Oxburgh interpret this metamorphism as resulting from an elevated geothermal gradient formed during continental rifting. Instead of a convergent plate boundary, they visualize an extensional basin, or en echelon series of basins, in which elevated temperatures were a result of lithospheric thinning. Large-scale melting of crustal rocks occurred at a relatively shallow depth (see intersection of field gradient with granite solidus in Fig. 22.12), and these rose to higher levels bringing with them the heat to cause metamorphism.

Perhaps the high-*T*, low-*P* facies series is not diagnostic of a single tectonic setting but indicates only regions of Earth where advective heat transfer by magmas or fluids has played an essential role. Although the effects of advective heat transfer are most obvious in high-*T*, low-*P* rocks, such transfer is not limited to low-pressure settings. For example, as we saw in Section 22.5, peak regional metamorphism in the Barrovian type locality involved a brief thermal pulse or pulses (Fig. 22.14). Magma intrusion, and associated fluid flow, probably drove this heat transfer. The thermal peak was then followed by rapid exhumation. Granulite facies hot spots in New Hampshire were almost certainly fossil conduits for advective heating. Granulite facies terranes may record significant heat input from magmas over large regions. Furthermore, LT/HP eclogite facies metamorphism in the Bergen Arcs was potentially driven by earthquakes that triggered extremely short pulses of fluids carrying heat. As with other types of metamorphism, advective *P–T–t* paths provide an important means of constraining tectonic models. Determining the relative roles of conduction and advection remains as a fundamental topic of study, as prograde metamorphism is driven by heat input.

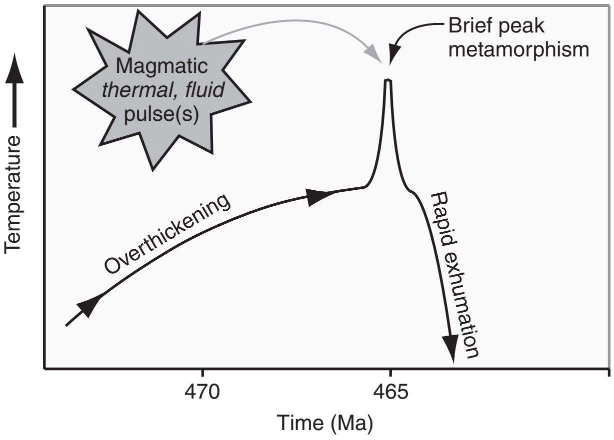


Fig. 22.14 Inferred temperature–time history for Barrovian metamorphism, northeastern Scotland. (After Ague and Baxter, 2007; published by permission of Elsevier.)

22.8 PROBLEMS

22.1 This problem uses the concentration profile measured in an apatite crystal from the Barrovian zones, Scotland, to estimate the total timescale of peak thermal metamorphism. The profile was measured using an electron microprobe and is given in Table 22.1 (data from Ague and Baxter, 2007). Analysis 1 is 2.5 μm from the center of the grain (the center is at $x = 0$ μm), whereas analysis 8 is at the outer edge. The boundary between the detrital core and the metamorphic rim is at 34.5 μm (for example, see Figure 22.11(A)). The sample is from the garnet zone; garnet–biotite thermometry (Chapter 19) indicates that the peak T was 535 °C. Assume an initial concentration of Sr in the core = 0.048 wt%, and an initial rim concentration of 0.115 wt%. The uncertainties on each Sr analysis are ±0.0087 wt% (2σ).

- **(a)** Plot Sr content as a function of distance for the apatite grain. Use the methodology described for Eq. (5.57) and shown in Figure 5.15 to solve for the peak timescale. Use trial and error to find the timescale that best fits the measured profile using the diffusion coefficient expression for Sr in apatite from Cherniak and Ryerson (1993): $D_{\text{Sr}} = 2.7 \times 10^{-7} \exp\left(-\frac{2.72\times10^5}{RT}\right) \text{m}^2\,\text{s}^{-1}$ (activation energy in joules).
- **(b)** Your result for part (a) quantifies the maximum timescale that the grain could have been held at peak temperatures. There could have been one thermal pulse of this total duration, or several shorter ones that add up to this total. Does the diffusion modeling differentiate between these two possibilities? Explain.

Table 22.1 *For Problem 22.1*

Analysis	Position (μm)	Sr content of apatite (wt%)
1 (center)	2.5	0.052
2	9.3	0.048
3	16.1	0.054
4	22.9	0.047
5	29.7	0.046
6	36.5	0.087
7	43.3	0.119
8 (edge)	56.9	0.109

22.2 The object of this problem is to calculate *P–T–t* paths for a crust that has been thickened by a series of thrust faults. Prior to faulting, radioactive heat-producing elements are concentrated in the upper 16.8 km where they give the rock a heat-generating capacity of 2 μW m^{-3}. Three successive thrust slices, each of which penetrates to the base of the heat-producing layer, are emplaced on top of the original crust. The heat-producing layer is then four times as thick (67.4 km). We will assume that the thrust faults form instantaneously at time $t = 0$. Such thickening should cause the geotherm to be raised well above the initial steady-state geotherm. We will calculate the position of the geotherm at 1-Ma intervals. The rate of erosion is assumed to be 0.4 m per 1000 years. All rocks have a thermal conductivity of 2.25 W m^{-1} K^{-1}, a thermal diffusivity of 10^{-6} m^2 s^{-1}, and a heat content ($C_p\rho$) of 2.5×10^6 J m^{-3} K^{-1}. The heat flow at the Earth's surface prior to faulting was 60 mW m^{-2}. Note that because the time increments are to be 1 Ma, the depth increments must be 5.616 km. We will now proceed with this problem in steps.

(a) First, calculate the steady-state geotherm prior to faulting; that is, with a single 16.8-km-thick heat-producing layer (review Problem 1.4); take the surface temperature to be 0 °C. Calculate temperatures for 5.6-km depth increments to a depth of 120 km; this will allow results to be used in the spreadsheet calculation of the numerical solution of the change in temperature with time.

(b) Next we introduce the three thrust faults. Starting at the Earth's surface where the temperature is 0 °C, we proceed as in part (a), but at a depth of 16.85 we reset the temperature to zero and start the calculation over again. This is repeated at each thrust fault. Note that the heat-producing layer is terminated 16.8 km below the lowest thrust fault. Extend your calculations to a depth of 120 km. Plot a graph of the resulting geothermal gradient, which should have three cusps to it.

(c) Using the faulted geotherm for initial temperatures, calculate the geothermal gradient at 1-Ma intervals using the numerical solution of Eq. (22.6). The faulted geotherm is entered in the first column of the spreadsheet calculation with successive columns being used for the geotherms at following 1-Ma intervals. The surface temperature is always 0 °C and the temperature of the deepest cells can be calculated using Eq. (22.7). Because the erosion rate is 4×10^{-4} m a^{-1}, the boundary between the lowest heat-producing layer and the underlying rocks must be moved up after an appropriate amount of time. Plot geotherms for 0, 5, 10, 15, 20, 25, 30, 35, and 40 Ma. This iterative calculation may take the computer an hour to complete.

(d) Construction of *P–T–t* paths. Knowing the rate of erosion, follow the *P–T–t* paths of two rocks that are initially at depths of 18 and 25 km by plotting their depths on the geotherms at successive times. This is best shown in a graph that extends to a depth of only 30 km.

(e) The metamorphic field gradient is plotted by drawing a line through the maximum temperatures on the *P–T–t* paths.

(f) On the same diagram, plot the Al_2SiO_5 phase diagram. The triple point is at a pressure of 0.376 GPa and temperature of 501 °C. The 1-atm equilibrium for the kyanite andalusite reaction is 200 °C, and that for the andalusite–sillimanite reaction is 770 °C. The kyanite–sillimanite reaction passes through a point at 1 GPa and 810 °C (Holdaway, 1971).

(g) To which facies series of Miyashiro (1961) do these rocks belong?

(h) If, after faulting, 10 Ma elapsed before significant erosion began, how deep would a sample have to be for it just to enter the sillimanite field at the maximum temperature along the *P–T–t* path after 40 Ma since the period of faulting?

(i) Despite the great thickening of the crust containing radioactive elements by the three thrust faults and its subsequent generation of heat, rocks show very little likelihood of entering the andalusite field. What geological conditions might raise the geotherm into this field?

(j) Clearly, the diapiric rise of magma bodies from deeper in the section would elevate the geotherms, as would advection of large quantities of hot fluids. Can you think of a simple way of incorporating the rise of magma into the numerical solution?

(k) For an order-of-magnitude calculation of the effect of rising magma on the geothermal gradient of a metamorphic terrane, magma can be treated simply as an advective metamorphic fluid (cf. Eq. (22.14)). In detail, magma would form intrusive bodies that would more locally perturb thermal gradients (see Lux *et al*., 1986) instead of being generally dispersed as would be metamorphic fluids. But such distinction need not concern us if we are evaluating only the overall effect of magmatic intrusion on a metamorphic terrane. Consequently, we will treat the terrane as having a permeability to magma without regard to whether flow is channelized or dispersed. The problem then is to evaluate the permeability.

In the first part of this problem the only movement of rock that occurred after thrusting was the exhumation due to erosion. We must now consider, in addition, the advection of magma. First, how rapidly do calcalkali magmas rise? Marsh and Kantha (1978) calculated that a 6-km-spherical diapir would solidify on its way to the surface if it did not rise more rapidly than 3 m a^{-1} (Section 3.11). We will use this figure for the ascent velocity. Of course, the percentage of a metamorphic terrane that consists of such rapidly moving bodies at any given time would be very small. We will assume that magmatic bodies constitute 5% of a metamorphic area (see Eq. (21.2) for analogy with fluid flow) and that magmatic pulses rising at 3 m a^{-1} occur every 100 years (approximate frequency of pulses in Mount St. Helens, for example). This gives a time- and space-averaged magmatic flux of 1.5×10^{-3} m a^{-1}. This can be added to the advective exhumation due to erosion, but the latter is so small by comparison that it can be ignored.

Equation (22.6) can be used to determine the change in the geothermal gradient resulting from the chosen flux of magma by replacing the advective exhumation term with the magmatic advective quantity of 1.5×10^{-3} m a^{-1}. Take the value of Γ (Eq. 22.15) to be unity. Repeat part (c) of the problem using this new value.

(l) Draw *P–T–t* paths for rocks that are initially at depths of 18 and 25 km, and indicate the resulting field gradient. To what metamorphic facies series would these rocks now belong?

22.3 Follow the general methodology used to derive Fourier's equation (Eq. 5.11) or the transient equation describing mass transport by fluid flow (Eq. 21.41) to derive the transient equation for mass transport heating (Eq. 22.3). How do your results compare with the general balance equation (Eq. 22.2)?

22.4 Two possible explanations for the almost isobaric cooling of many granulite facies terranes involve either the extraction of a magmatic fraction that took with it the heat-producing elements or early flushing of the rocks with hot CO_2-rich fluids. Construct schematic *P–T–t* paths for these two scenarios.

22.5 Exact analytical solutions for Eq. (22.4) exist for certain initial and boundary conditions. These solutions prove extremely valuable for gaining a general understanding of processes and timescales, and for testing numerical models such as those developed in Problem 22.2. One such solution is for the case where there is an initial linear geothermal gradient through the region at $t = 0$. The temperature at the boundary $z = 0$ can vary linearly with time. Thus, the initial condition can be written as: $T = T_0 + az$, where T_0 is the temperature at $z = 0$ and a is a constant that gives the gradient. The boundary condition at $z = 0$ is: $T = T_1 + bt$, where b is a constant that describes the change in temperature with time. With these initial and boundary conditions, the solution for the temperature at position z and time t ($T_{z,t}$) is (Carslaw and Jaeger, 1959, p. 388)

$$
\begin{aligned}
T_{z,t} = {} & T_0 + az + \frac{k_\mathrm{T} A t}{K_\mathrm{T}} - uat + 0.5(T_1 - T_0) \\
& \left[\operatorname{erfc}\frac{z - ut}{2\sqrt{k_\mathrm{T} t}} + e^{\frac{uz}{k_\mathrm{T}}}\operatorname{erfc}\frac{z + ut}{2\sqrt{k_\mathrm{T} t}}\right] \\
& + \frac{1}{2u}\left(b + au - \frac{k_\mathrm{T} A}{K_\mathrm{T}}\right) \\
& \left[(z + ut)e^{\frac{uz}{k_\mathrm{T}}}\operatorname{erfc}\frac{z + ut}{2\sqrt{k_\mathrm{T} t}} + (ut - z)\operatorname{erfc}\frac{z - ut}{2\sqrt{k_\mathrm{T} t}}\right]
\end{aligned}
$$

(a) We can use this expression to estimate the time it would take for a lower crustal rock to reach granulite facies conditions due to magma underplating. Heat transfer takes place up through the crust by conduction. Assume a 40-km-thick crust having a thermal conductivity K_T of 2.25 W m^{-1} K^{-1}, a thermal diffusivity k_T of 10^{-6} m^2 s^{-1}, and an average heat production A of 10^{-6} W m^{-3}. For this problem, the base of the crust is at 0 km, and increases *upward*. Thus, $z = 40$ km corresponds to Earth's surface. Take the temperature at the base of the crust to be 700 °C prior to underplating, and the initial temperature at the surface to be 0 °C. Assume that the underplated magma has a constant temperature T_1 of 1100 °C (the b coefficient is zero). Make a plot of temperature as a function of depth for 10^5, 10^6, 2×10^6, 3×10^6, and 4×10^6 years. How long will it take for a rock 30 km below the surface to heat to granulite facies temperatures of ~770 °C? The temperature at the surface will increase a bit after long times due to radiogenic heating, but this increase is small and can be neglected.

Note that erfc is the complimentary error function (erfc(y) = 1 − erf(y); Section 5.6). The analytical solution given above can be tricky to solve in some cases; see discussion for Problem 21.8 if you run into problems Also, even though there is no advection, u (Eq. 22.3) has to be set to a small value to avoid division by zero. A value of 10^{-6} generally works, although you may need to experiment a little. The calculations for this problem should work using distance in meters and time in years (you will need to convert the values for K_T and so on given above).

(b) Repeat part (a), but now assume that magmas are also ascending into the lower crust. Use the magma flux $u = 1.5 \times 10^{-3}$ m^3 m^{-2} a^{-1} discussed in Problem 22.2. Plot temperature as a function of depth for 10^5, 10^6, and 2×10^6 years. How long will it take for a rock 30 km below the surface to heat to granulite facies temperatures of ~770 °C? How do your results compare with those in part (a)?

(c) Throughout this book we have highlighted the parallels between expressions describing conservation of mass and conservation of energy. As a further illustration, show how the analytical solution given above can be transformed into the mass transfer solution of Eq. (21.48).

22.6 In this problem, we will estimate the duration of advection through the hot spot in New Hampshire studied by Chamberlain and Rumble (1988, 1989). Following Chamberlain and Rumble (1989), we hypothesize that the fluids were flowing more or less vertically; heat would then have been conducted laterally away from the hot spot into the surroundings.

(a) For this part of the problem, we will assume that the fluids flowing through the hot spot fracture zone were well mixed and maintained a constant temperature. We take T_1 as the constant temperature at the contact between the hot spot and the surrounding country rock at $x = 0$, and T_0 as the initial temperature in the country rock. The temperature at position x and time t ($T_{x,t}$) is then given by the following solution to the heat conduction equation, which is appropriate for a planar hot spot country rock contact

$$T_{x,t} = T_1 - (T_1 - T_0)\mathrm{erf}\left(\frac{x}{2\sqrt{k_T t}}\right)$$

Chamberlain and Rumble (1989) used geothermometry to construct an east–west profile of metamorphic temperatures across the western margin of the Bristol hot spot. Temperatures in the hot spot itself average ~700 °C, then fall progressively to 560 °C at a distance of 0.6 km from the hot spot margin, to 525 °C at 1.6 km, and to 480 °C at 4.5 km. Thus, for the analytical solution, we can take T_1 to be 700 °C, and T_0 to be 475 °C. First make a plot of T versus distance for the geothermometry data, starting with 700 °C at $x = 0$ and extending to 480 °C at $x = 4.5$ km. Then, solve the expression given above for times of 2×10^4, 2×10^5, 2×10^6, and 2×10^7 years. Take k_T to be 10^{-6} m^2 s^{-1}.

(b) Repeat part (a), but now assume that the temperature of the hot spot increased linearly with time from 480 °C to 700 °C. You will need to use the general solution given in Problem (22.5). The b coefficient will be nonzero, but a will be zero. The radiogenic heat production term A can be set to zero, and the advective term u should be set to a small number (see part (a) of Problem 22.5) as fluid advection laterally away from the hot spot is ignored. Start by assuming that the total timescale of flow was 10^4 years, and that the boundary T increased from 480 °C to 700 °C over this timespan. Then repeat this procedure for 10^5 years and 10^6 years. Note that you will have to recalculate the b term for each case.

(c) The calculations in parts (a) and (b) represent a likely range of scenarios for heat transfer adjacent to the hot spot. Compare your model results with the measured peak temperature data. What range of timescales fits the data best? Would your conclusions about timescale change if magma had flowed through the hot spot instead of fluid? Using a somewhat different set of equations, Chamberlain and Rumble (1989) concluded that the duration of fluid flow in the hot spot was very short – of the order of 10^5 years. How do your results compare?

22.7 The following analyses are of a zoned garnet porphyroblast and included biotite grains. Using the method of Spear and Selverstone (1983) (Fig. 19.8) construct a P–T–t path for the rock assuming that the average density of the rocks is 2.8 Mg m^{-3}. What tectonic history might this rock have undergone during metamorphism?

Core ⟵			⟶	Rim
$X^{\mathrm{G}}_{\mathrm{Alm}}$	0.865	0.875	0.895	0.910
$X^{\mathrm{Bi}}_{\mathrm{Ann}}$	0.60	0.58	0.67	0.65

22.8 A section of country rock of a given thickness is sandwiched between two igneous dikes. Each dike is at the same temperature, and they hold the temperature constant at the contacts with the country rock. An endothermic devolatilization reaction is proceeding in the country rock. The problem is set up so that the center of the country rock section is at $x = 0$, and the contacts are at $-X$ and X. The difference in temperature between the country rock and the contacts (ΔT) at steady state is given below (Carslaw and Jaeger, 1959). Note that in a natural system, the ΔG of reaction and, thus, the reaction rate will vary with temperature; this important relationship is not accounted for below as the rate is assumed to be constant. Nonetheless, the expression is useful for illustrating some of the general effects of reaction on temperature at steady state

$$\Delta T = -\frac{\Delta H R^{*}\left(X^2 - x^2\right)}{2K_T}$$

(a) Consider a 10-m-thick layer ($X = 5$ m). Using the value 1500 J m^{-3} a^{-1} (Section 22.6) for the average rate of heat consumption across the country rock, plot ΔT as a function of distance from $-X$ to X using $K_T = 2.25$ W m^{-1} K^{-1} = 7.09×10^7 J a^{-1} m^{-1} K^{-1}. Does the reaction have much effect on the temperature at steady state?

(b) Repeat part (a), but now consider a 500-m-thick layer ($X = 250$ m). Why do your results differ from those of part (a)?

(c) The value of 1500 J m^{-3} a^{-1} discussed in Section 22.6 is based on a ΔG of reaction = −1 J mole^{-1}. How would your result from part (b) change if the overstepping increased to −10 J mole^{-1} or decreased to −0.1 J mole^{-1}?

22.9 Figures 22.2(B) and (C) illustrate exhumation by erosion.

(a) Draw similar cartoons showing how exhumation can occur by large-scale extensional normal faulting.

(b) Draw similar cartoons showing how thrust faulting, in and of itself, cannot produce exhumation.

23 Origin of rocks

23.1 INTRODUCTION

In previous chapters, we have dealt with the specific details of the formation of rocks. In this final chapter, we examine the broader question of their ultimate origin. What conditions in the Earth bring about the formation of rocks, and from where does the material come to form them? These are important questions, the answers to which are critical to interpreting Earth's history, for only in rocks is any record of the geologic past preserved. There is certainly no unanimity among petrologists on answers to all aspects of these questions, but the theory of plate tectonics has provided a unifying paradigm that has eliminated much controversy. Difficulty in answering the questions stems from the inaccessibility of the regions in which the controlling processes operate and our limited experimental and theoretical knowledge of the behavior of material under the pressures and temperatures that exist in such regions. Because of the rapid evolution of ideas on this topic, no attempt is made in this chapter to review all aspects of these questions. Instead, some basic principles are discussed which should be of help in evaluating these ideas.

Ever since the formation of the Earth 4.567 Ga ago, heat generated by accretionary processes, radioactive decay, and gravitative differentiation – in particular of the core – has been transferred to the surface of the planet where it has been radiated into space. Conduction, advection, and radiation have all played roles in this transfer. The opacity of rocks in the upper mantle and crust eliminates radiation as a feasible means of transferring heat through the outer half of the planet. Also, the thermal conductivity of rocks is so low that Earth's cooling rate has been determined largely by advective processes, which include mantle and crustal convection, magma intrusion, metamorphism, and fluid transport.

All rocks, whether igneous, metamorphic, or sedimentary, owe their origins directly or indirectly to the advective cooling of the Earth. The rise of magma into the crust or its extrusion onto the surface is an obvious example of such advection. But so is the development of contact metamorphic minerals through endothermic reactions. Similarly, cold crustal rocks that are buried and undergo regional metamorphism, which is largely endothermic, and then return to the surface of the Earth provide another means of advectively removing heat from the planet's interior. Most metamorphic reactions involve the liberation of fluids. The upward flux of these hot fluids from metamorphic terranes is still another means of advectively cooling the Earth. Finally, sedimentary rocks are formed as a result of changes in elevation of Earth's surface, positive changes resulting in erosion and the production of sediment and negative changes producing basins for deposition. The changes in elevation are brought about by plate tectonic motions that result from gravitational instabilities set up by the advective cooling of the planet.

Although cooling of the Earth is a continuous process, the rate of advective heat transfer may have fluctuated throughout geologic time. Evidence for this comes from the rate of production of rocks, which does not appear to have remained constant. Despite the incompleteness of the geologic record, especially as it gets older, most rocks were formed during distinct episodes. According to Moorbath (1977) these occurred between 3.8–3.5, 2.9–2.6, 1.9–1.6, 1.2–0.9, and 0.6–0 Ga ago. Because these episodes appear to be worldwide, they probably reflect major periods of advection in the mantle and possibly the core.

Certain rock types, such as tholeiitic basalt and greenschists, have formed throughout geologic time, but others are restricted to specific rock-forming episodes (see Section 15.8). Greenstone belts, for example, are formed in the first two episodes; massif-type anorthosites characterize the 1.2- to 0.9-Ga episode; and low-T, high-P metamorphic rocks are almost entirely restricted to the most recent episode (the latter group may have formed earlier in the Precambrian but are not preserved due to erosion or continued heating at depth). Elevated geothermal gradients producing regional granulite facies and ultrahigh-temperature metamorphism coincide with the first four episodes, as well as a fifth at about 500 Ma. Some of these episodes may be related to the growth of supercontinents at various times in Earth's history, such as Gondwana at ~500 Ma and Rodinia at ~1000 Ma (Brown, 2006). Ultimately, these changes must reflect evolution of the planet's interior as it progressively cools. This evolution can be expected to continue, with the rate of production of new rocks decreasing as the Earth cools. Eventually, the rate of heat production will be insufficient to maintain advection. The Earth will then cease to be a dynamic planet, and no new rocks will form.

Because the advective cooling of the Earth plays a dominant role in the production of rocks, we begin this chapter with a consideration of the mechanisms by which heat can be transferred advectively. This is followed by a discussion of how advection leads to conditions favorable to the generation of magmas and metamorphic rocks. Next, the actual processes by which partial melts coalesce and form ascending bodies of magma are treated. This, in turn, is followed by a discussion of the factors controlling the compositions of partial melts. Finally, the generation of the three main types of magma – basaltic, andesitic, and granitic – is examined in terms of their plate tectonic settings. For additional information, the book by Yoder (1976) on the *Generation of Basaltic Magma* and the review article by Wyllie (1988) are particularly informative.

23.2 ADVECTIVE HEAT TRANSFER IN THE EARTH

The Earth is believed to have formed by the accretion of particles that condensed from the original solar nebula (Wetherill, 1990) and had a composition similar to that of common chondritic meteorites (Section 6.1). Minerals that may have been present in the original nebula include various Ca-, Al-, and Ti-oxides, metallic Fe and Ni, forsterite, enstatite, feldspar, FeS, and amphibole. As the planet grew, impact energy generated by incoming planetesimals would have increased because of the greater gravitational attraction; this, then, would have steadily raised surface temperatures. At the same time, the increasing mass of the planet would have raised internal temperatures through compression. This accretionary period was essentially complete by 4.567 Ga (Jacobsen, 2003). Toward the close of the accretionary period, temperatures in the Earth ranged from about 600 °C at the center to about 1600 °C near the surface, with a maximum of about 2200 °C at a depth of 1400 km (Ringwood, 1975). These temperatures were high enough to cause partial melting only in the outer 200 km. Below this, the geothermal gradient was progressively farther away from the solidus, which rises rapidly with increasing pressure.

The presence of FeS among the original condensates from which the Earth formed is of great importance to the early development of the planet. FeS and Fe form a low-temperature eutectic (about 1000 °C), at which a dense iron-sulfide liquid is formed that is immiscible in silicate liquids (Fig. 11.4). This liquid would have sunk rapidly through the upper molten silicate layer and then more slowly through the solid but plastic lower part of the Earth. This sinking would have effectively mixed and homogenized any compositional layering that might have developed during the accretionary stage, and the iron-rich melt would have taken with it the siderophile elements (Righter, 2003). Recent experiments indicate that potassium may also have entered this iron-rich melt, which would be of importance because of the radioactive heat producing capacity of ^{40}K (Lee and Jeanloz, 2003). The segregation of this dense liquid to form the core of the Earth was a strongly exothermic process, with enough heat being generated to raise the average temperature of the Earth by 2000 °C (Ringwood, 1975). At the same time, heat was generated from the decay of short-lived radioactive isotopes, such as ^{26}Al. After the core had separated but before 4.45 Ga, Earth was involved in a mega impact that resulted in formation of the Moon (Zhang, 2002). Although Earth may have had a relatively cool beginning, the combined effect of all these early heat-generating processes was to melt the planet and form a magma ocean (Anderson, 1984).

Convection within this magma ocean would have been vigorous and much of Earth's early heat would have soon been lost, possibly in as little as 10^4 years (Solomatov, 2000). The temperature gradient in the convecting layer would have been very nearly adiabatic (see Problem 7.7), and as a result, solidification would have taken place from the base upward, as it does in convecting bodies of magma (Section 14.4 and Fig. 14.8). Today, the core of the Earth is solidifying in this same way, with the solid inner core being overlain by the liquid outer core. Convection in the magma ocean would have been so vigorous that there would have been little chance for crystals to separate and differentiate the liquid. The lower part of this magma ocean is estimated to have become 60% crystallized within 1000 years, and consequently, convection would have stopped because of the increased viscosity (Eq. (2.9)). Subsequently, any differentiation of the lower part of the mantle would have had to be by some process of compaction and redistribution of residual liquid. An upper magma ocean, which might have been ~300 km deep would have remained liquid for > 100 Ma. This upper convecting magma ocean would also have solidified from the bottom up, and it is likely that it would have differentiated to give the initial makings of a continental crust. Again, this differentiation could have continued long after convection stopped through crystal mush compaction. Recall that the oxygen isotopes of Earth's oldest zircons indicate that some type of continental crust and oceans existed as early as 4.4 Ga (Wilde *et al*., 2001). The magma ocean would certainly have had to have solidified by this time. We can conclude that from early in Earth's history, the planet has been differentiated into a core and mantle, which would have been capped by a differentiate that would have formed the proto-continental crust.

Following the removal of the initial surfeit of heat and solidification of the magma ocean, Earth's heat production has since been due largely to the decay of long-lived radioactive isotopes. These would have caused heating until eventually a balance was achieved between the rates of heat production and heat loss. If temperatures rose too high, convection and conduction would have become more effective and removed the excess heat. The Earth has since remained essentially in this state of thermal balance (McKenzie and Weiss, 1975). Of course, the abundances of the radioactive elements have slowly decreased with time because of decay, so heat production would have been four times greater in the early Earth than at present (Problem 23.1). To maintain thermal balance, the heat flux from the early Earth must have been four times as great. This would have required more rapid convection, which may have taken the form of multiple plumes or thin-skinned plate tectonics (Condie, 2005; Cawood *et al*., 2006).

Convection is easily shown to be an important mode of heat transport in the upper mantle today, and thus it must have been still more important in the early Earth. In Section 14.4, the dimensionless Rayleigh number (Eq. (14.10)) was introduced to evaluate the likelihood of convection occurring in bodies of magma. In that case we considered a sheet of liquid whose upper and lower surfaces were held at different temperatures (Fig. 14.7). A Rayleigh number can also be defined for a sheet that generates its own heat internally; this would be the situation in the upper mantle, where heat is generated from radioactive decay. The Rayleigh number in this case is defined as

$$\mathrm{Ra} \equiv \frac{\rho g d^5 \alpha A}{K k \eta} \tag{23.1}$$

where ρ is the average density, g the acceleration of gravity, d the depth of the sheet, α the coefficient of thermal expansion, A the heat productivity per unit volume, K the thermal conductivity, k the thermal diffusivity, and η the viscosity. The critical Rayleigh number for convection when the layer is bounded by rigid surfaces against which there is no slip is 2772, and when the surfaces are free to slip it is 868.

Approximate values of all terms in Eq. (23.1) can be found. The viscosity of the upper mantle is about 7×10^{20} Pa s based on postglacial uplift rates (Peltier and Andrews, 1976). Deep-focus earthquakes reveal that subducted slabs can penetrate to depths of 600 to 700 km. It is reasonable to assume, therefore, that convection might affect a layer to a depth of 700 km. If we assume that the upper mantle is similar to the peridotite given in Table 1.1, its heat-generating capacity, A, is 10^{-8} W m^{-3}, and its thermal conductivity, K, is 3.35 W m^{-1} K^{-1}. The average density of the upper mantle can be taken to be 3.5 Mg m^{-3}, its thermal diffusivity, k, is 10^{-6} m^2 s^{-1}, and its coefficient of thermal expansion, α, is 3×10^{-5} K^{-1}. The Rayleigh number calculated from these values is 7×10^5, which is much greater than the critical Rayleigh number. We can conclude, therefore, that the upper mantle must be convecting today, and that in the past, when more radioactive heat was produced, the Rayleigh number would have been still greater, and convection would have been even more vigorous (Problem 23.2).

Although it is simple to show that convection must occur in the mantle, it is not yet possible to predict the geometry or rate of this convection. In the calculation above, we assumed a depth of 700 km for the convective layer. If the depth had been taken to be the entire thickness of the mantle, the Rayleigh number would have been greater, and we would have concluded that whole-mantle convection occurs. Conversely, shallower depths could have been selected that would still have given Rayleigh numbers in excess of the critical value. Convection, therefore, can occur on many different scales. Whether the mantle convects in large cells that affect the entire depth of the mantle or in tiers of shallower cells depends on many factors, such as the variation in composition and rheological properties with depth, the distribution of radioactive elements, and the amount of heat entering the mantle from the core (here is where the potassium content of the core would be important). Unfortunately, these parameters are poorly known, so no definitive calculation can be made. Most models, however, call on convection to operate at several scales simultaneously.

The most direct evidence for mantle convection is provided by the motion of lithospheric plates. Hot, buoyant lithosphere, created at spreading oceanic ridges, cools and becomes denser as it moves away from the ridge, eventually sinking into the mantle at subduction zones. The dimensions of lithospheric plates define the largest-scale convection that we know of in the Earth. Surprisingly, it was while relating the motion of these plates to horizontal temperature gradients associated with large-scale convection that evidence was found for smaller-scale convection. The gradual deepening of oceans away from divergent plate boundaries is easily accounted for in terms of the progressive cooling and densification of newly formed lithosphere as it moves away from the spreading axis (Sclater *et al*., 1981). As with other processes controlled by conductive cooling (Section 5.3), the deepening of the ocean is proportional to the square root of time, at least for ocean floor younger than 70 Ma (Fig. 23.1). Beyond 70 Ma, the rate of deepening decreases exponentially toward a constant depth. These observations suggest that the cooling rate of plates significantly decreases at times greater than 70 Ma. Parsons and McKenzie (1978) have suggested a model involving small-scale convection that provides a simple explanation for these observations. A description of their model follows.

New, hot lithosphere formed at a spreading axis loses heat to the overlying water as the plate moves away from the axis (Fig. 23.2). The water can be taken to be at a constant 0 °C. The thickness of the rigid lithospheric plate at any time, t, after leaving the spreading axis depends on the depth to which the brittle to ductile transition has extended. This transition can be represented by an isotherm, $T_{\mathrm{B/D}}$. Immediately below the rigid plate is a thermal boundary layer in which the viscosity decreases rapidly as temperatures

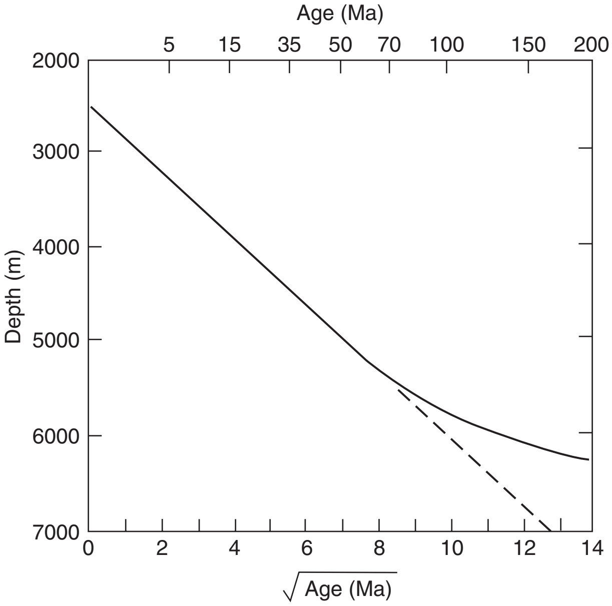


Fig. 23.1 Linear increase in depth of ocean floor with square root of age of ocean floor during first 70 Ma is due to conductive cooling of oceanic lithosphere. (After Sclater

increase. The bottom of this layer is actually gradational into the underlying asthenosphere in which the large-scale convection occurs, which is responsible for transporting the plate. For convenience we identify another isotherm, T_A, as the base of this thermal boundary layer.

Heat transfer through the rigid plate is largely by conduction, but advective hydrothermal fluids may play a role, especially near the spreading axis. The depth, z, to which any isotherm, T, moves in time t is given by a relation similar to Eq. (5.13) but in this case for the cooling of a half-space

$$\frac{T - T_s}{T_0 - T_s} = \text{erf}\left(\frac{z}{2\sqrt{kt}}\right) \tag{23.2}$$

where T_s is the constant temperature of the cooling surface, T_0 the initial temperature, and k the thermal diffusivity. Where the ocean floor temperature is 0 °C ($T_s = 0$), Eq. (23.2) becomes

$$T = T_0\text{erf}\left(\frac{z}{2\sqrt{kt}}\right) \tag{23.3}$$

Parsons and McKenzie assume T_0, the initial temperature of the rock at the spreading axis, to be 1300 °C, the brittle to ductile transition to be at 975 °C, and the base of the thermal boundary layer to be at 1260 °C. Substituting these values into Eq. (23.3) indicates that the thickness of the brittle zone is given by $z_{B/D} = 1.63\sqrt{kt}$, the depth to the base of the thermal boundary layer is given by $z_A = 3.06\sqrt{\text{kt}}$, and the difference between them – the thickness of the thermal boundary layer – is given by $\delta z = 1.43\sqrt{\text{kt}}$ (Fig. 23.2).

Rock within the thermal boundary layer cools against the base of the overlying rigid plate and becomes denser than the underlying rock. Near the spreading axis where the thermal boundary layer is thin, the Rayleigh number for this layer is below the critical value that would allow these density inversions to cause convection. Parsons and McKenzie believe that after 70 Ma, however, the thickness of this layer is great enough that the critical Rayleigh number is exceeded, and from then on convection occurs in the thermal boundary layer. As the layer thickens, convection cells become larger until a balance is achieved between the heat transported vertically by the cells and the heat conducted through the overlying rigid plate. Once convection starts, the temperature gradient in the thermal boundary layer is essentially adiabatic and heat is transferred into the base of the rigid plate much more rapidly. This then accounts for the decrease in the rate of cooling and subsidence of ocean floor older than 70 Ma (Problem 23.3).

Small-scale convection may occur not only within the thin thermal boundary layer immediately beneath the lithosphere but in the upper mantle as a whole. Such convection would still be at a smaller scale than that defined by the plates. Its existence seems necessary to explain the heat flow through older lithosphere, but at present, there is little direct evidence for such convection. The stresses imposed on the base of the rigid plate by this convection are normally too small to break the plate, but they might be great enough to delaminate lower parts of the lithosphere (Kay and Kay, 1993). Such an explanation has been proposed to explain the origin of the rock types and tectonic styles of the Archean basement (Fig. 4.80), with its short greenstone belts infolded between domed areas of gray tonalitic gneiss whose wavelength may reflect the scale of this convection (McKenzie and Weiss, 1975). Their distribution in plan view might have been similar to that formed in the molten wax model shown in Figure 23.3. Zegers and van Keken (2001) suggest that a dense, lower eclogitic part of the Archean crust delaminated and sank into the mantle. The hot mantle that rose to replace it underwent partial melting to form basaltic magma that was either erupted to form the greenstone belts, or was intruded at the base of the oceanic crust where it caused partial melting to form tonalitic magma. With the steeper temperature gradients during the early Precambrian, the lithosphere would have been thinner and might have been broken by the small-scale convective flow. Indeed, the lack of long Archean mountain chains may be evidence that large lithospheric plates were unable to form because of disruptive small-scale convection (Davies, 1999).

The change in tectonic style from greenstone belts in the Archean to long mountain chains in the Proterozoic and

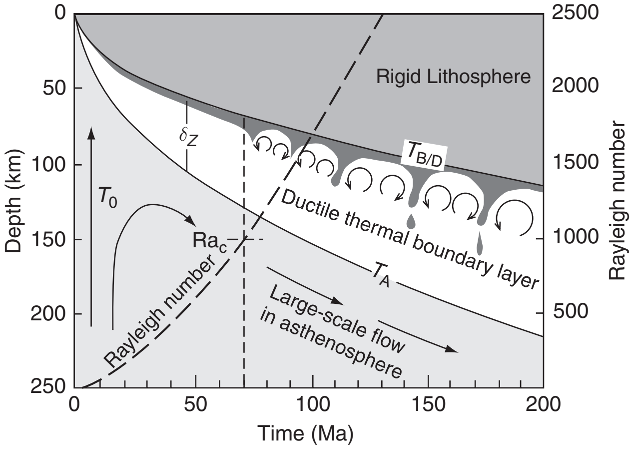


Fig. 23.2 A thermal boundary layer at the base of the cooling and thickening oceanic lithosphere becomes gravitationally unstable after 70 Ma and begins to convect, with cooler, denser material (shaded) near the base of the lithosphere sinking into the boundary layer. This convection is proposed as the means by which ocean floor older than 70 Ma does not continue cooling as rapidly as younger ocean floor. (After Parsons and McKenzie, 1978.)

Phanerozoic indicates that during at least the last half of Earth history, the upper mantle has been involved in large-scale convection, which must have played a dominant role in cooling the Earth. Even in Archean times, small thinner plates must have existed, because paleomagnetism preserves a record of their motion (Cawood *et al*., 2006). Despite its importance, there is no agreement on how the near surface convective flow of plate tectonics extends to depth in the Earth, but seismic tomography, which depends on seismic velocity which, in turn, is a function of temperature, has the potential of unraveling this problem. The geothermal gradient, which is determined in large part by this convection, controls the processes that lead to the formation of crustal rocks.

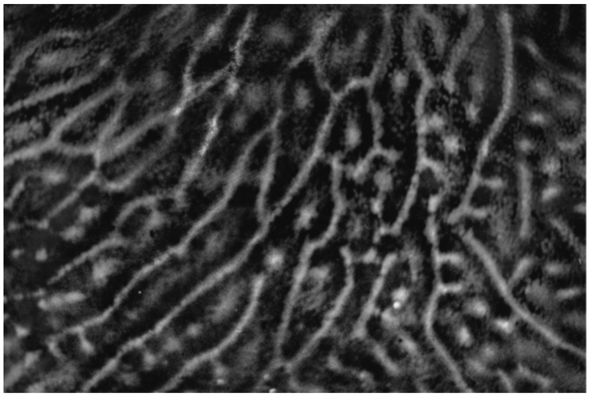


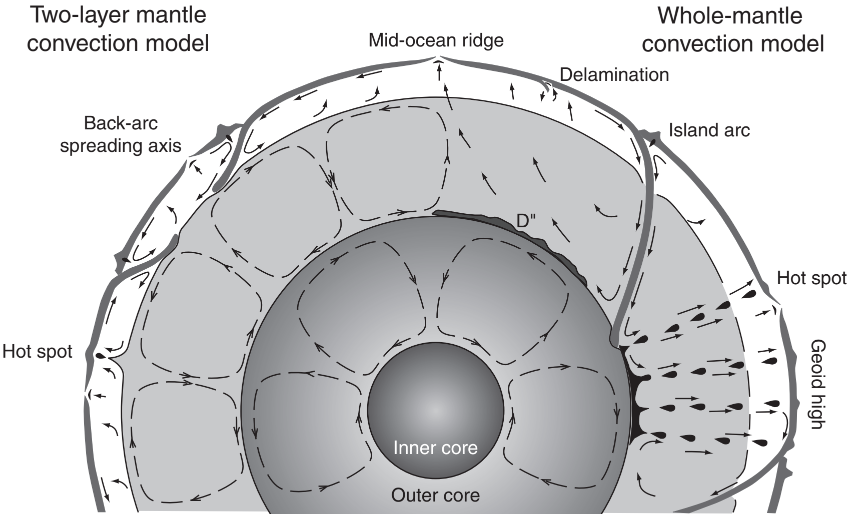
Fig. 23.3 Plan view of convection rolls in 1-cm-thick sheet of paraffin wax caused by cooling of upper surface of wax, as seen from above. Convecting wax cools as it travels horizontally beneath a thin crust, and crystallization begins first where cells descend (continuous white zones). Regions of upwelling are marked either by hot, clear wax or by small patches of white wax crystals that become trapped in eddies at the top of convection cells. The pattern may resemble the convection in the Archean, when the geothermal gradient would have been steeper and the lithosphere thinner and weaker. As the Earth cooled, the lithosphere has thickened and plate tectonic convection has shifted to longer wavelengths.

The main controversy over large-scale convection centers around whether the whole mantle is involved in convective overturn, or the upper and lower mantle convect separately (Silver *et al*., 1988). Deep-focus earthquakes show that subducted lithospheric plates descend to depths of 600 to 700 km – below this seismicity is lacking. Moreover, focal mechanisms indicate that although the subducted slab is in extension at shallow depths, it is in down-dip compression by the time it reaches the depth of the seismicity cutoff. These facts can be interpreted as indicating that the prominent seismic discontinuity at a depth of 670 km marks a compositional boundary separating an upper and lower mantle between which matter is rarely transferred. Conversely, the discontinuity could result from a phase change, across which subducted slabs are able to cross. The lack of seismicity below the discontinuity would then indicate that the slab was simply no longer brittle. These two models are illustrated in Figure 23.4.

The existence of two long-lived isotopically distinct reservoirs in the mantle (Section 13.4) is strong geochemical evidence for there being separate upper and lower mantles. The reservoir from which MORB is derived is strongly depleted in LIL elements. The other is far less depleted and approaches more closely the model bulk Earth composition. The MORB source is the most commonly sampled reservoir, with the less depleted one making significant contributions only at "hot spots." These reservoirs, consequently, are interpreted to be the upper and lower mantle, respectively. Only where plumes rise into the upper mantle from the lower mantle does the less depleted reservoir make a contribution to crustal rocks. This interpretation can be further supported if the continental crust is believed to have formed from the LIL elements that were in the MORB source prior to depletion. If the MORB reservoir is uniformly depleted, its volume would have to be almost exactly that of the upper mantle if it were to account for the entire continental crust.

Although seismicity does not extend below the 670-km discontinuity, the velocity of seismic waves traveling through

Fig. 23.4 Possible patterns of convection in the Earth. On the left, a two-layer mantle shows subducted lithosphere descending only to the 670-km discontinuity, which is interpreted to be a compositional barrier across which heat but not matter is transferred. On the right, whole-mantle convection has subducted plates descending as far as the core–mantle boundary. In this model the 670-km discontinuity is formed by a phase transformation. (From various sources; see text for discussion.)



deeper parts of the Earth can be measured. Differences in velocities allow tomography to be used to investigate the structure of the deep mantle in the same way that X-ray CAT-scans (computerized axial tomography) are used to image the interior of the human body. Because subducted rocks are cold, they have higher seismic velocities than do hot mantle rocks, so their descent into the Earth can be tracked with global mantle tomography (Romanowicz, 2003).

Some subducted slabs can be seen to lie flat on the 670-km discontinuity, but others penetrate deep into the lower mantle (van der Hilst and Kárason, 1999). A belt of higher velocities (hence colder rock), which surrounds the Pacific basin, can be traced from a depth of 1000 km to the core–mantle boundary (Dziewonski, 1984). This belt lies beneath the circum-Pacific zone of subduction and is presumably formed by the sinking of cold lithosphere into the lower mantle. These high-velocity zones may even correlate with depressions on the core–mantle boundary, in which case subducted slabs must be able to sink all the way to the bottom of the mantle (Obayashi and Fukao, 1997; Boschi and Dziewonski, 2000). The 150–350-km-thick D″ layer, which separates the core from the mantle, may be a graveyard of subducted slabs (Garnero, 2000; Lay *et al*., 2004). This layer is marked by seismic anisotropy, which indicates preferred orientation of crystals or of bodies of melt. The ultralow velocity layer at the base of D″ is almost certainly composed of partly molten rock.

If subducted slabs descend into the lower mantle, an equivalent mass of material must be transferred back into the upper mantle. This material would be relatively hot and be characterized by lower seismic velocities. Again, tomography reveals that two regions of low velocity occur in the lower mantle, one underlying the Pacific basin and the other underlying southern Africa. These regions, which are commonly referred to as *superplumes*, underlie parts of the Earth that have the majority of the hot spots, and they form highs on the geoid (Crough and Jurdy, 1980). Hot spots were interpreted by Morgan (1971) to be the surface expression of mantle plumes that brought up hot material from the core–mantle boundary. These plumes are expected to have small diameters as they rise through the mantle, and consequently resolving them seismically has proved difficult. However, tomography has now confirmed their existence (Nataf, 2000). The plumes beneath Hawaii (Montelli *et al*., 2004; Nolet *et al*., 2007) and Iceland (Bijwaard and Spakman, 1999) can be traced all the way from Earth's surface to the lower mantle.

Two models of mantle convection are illustrated in Figure 23.4, one involving separate convection of the upper and lower mantle, and the other involving whole mantle convection. Since publication of the first edition of this book in 1990, the weight of evidence appears to favor the whole-mantle-convection model. Subducting slabs can be traced into the lower mantle even though the 670-km discontinuity may impede their progress in some cases. The return flow from the lower mantle appears to be far more focused and takes the form of mantle plumes, most of which rise from two large regions in the lower mantle beneath the Pacific and southern Afirca.

One extremely important finding from tomography is that although the high velocities associated with subducted slabs can be traced as deep as the core–mantle boundary, the low velocities associated with the hot mantle rising beneath spreading axes, while clearly detectable at depths of 150 km, are not present at depths greater than 350 km (Silver *et al*., 1988). Spreading axes are therefore not the expression of large ascending plumes. The upwelling that occurs along a spreading axis appears to be a passive process where hot mantle simply rises to fill the gap left by the diverging plates. As such, they cannot be interpreted as major ascent conduits that balance the downward flow of material into the mantle at subduction zones. Large plumes, however, may have initiated spreading from triple junctions where large igneous provinces have been created.

This model of passive divergent plate boundaries is supported by the timing of events associated with rifting. For example, the Red Sea rift is now flanked by mountains that rise 2.5 km above sea level, but at the start of rifting, 30 Ma ago, this region was at sea level and did not start uplifting until 20 to 25 Ma ago. As hot mantle rose to accommodate the divergence of the plates, temperatures increased and the land subsequently was uplifted (McGuire and Bohannon, 1989). Had an active mantle plume caused rifting, the uplift would have preceded the rifting. However, at the southern end of this rift, where it meets at the triple junction with the Gulf of Aden and the African rift valley (Fig. 15.19), the surrounding terrane has been domed up to form the highest part of the African continent. This is also the center of a large igneous province, which has been interpreted to have formed over a mantle plume (White and McKenzie, 1989).

The rate of large-scale convection associated with plate motion is sufficiently high that it plays the dominant role in cooling the Earth. At present, ocean floor is being created and subducted at a rate of 3 km^2 a^{-1}. If the average thickness of the plate is 125 km at the time of subduction, 375 km^3 of lithosphere is subducted each year. At this rate, the entire volume of the upper mantle (3×10^{11} km^3) could be processed by plate tectonic convection in 0.8 Ga or the entire volume of the mantle (9×10^{11} km^3) in 2.4 Ga.

Temperatures in the Earth depend mainly on the extent and form of convection systems. In layers, such as the lithosphere, where convection does not occur, temperatures rise steeply as a result of the slow transfer of heat by conduction. Within convecting layers, however, the temperature gradient is extremely shallow, being very nearly adiabatic (Eq. (14.14)). In the mantle just beneath the lithosphere this gradient is about 0.6 °C km^{-1} (Problem 23.5 and 23.6). If the mantle is layered, steep temperature gradients exist in thermal boundary layers at the bottom and top of adjoining convecting layers. Richter and McKenzie (1981), for example, estimate that if the upper and lower mantle convect separately, a 500 °C temperature increase would exist across a narrow zone at the 670-km discontinuity.

Figure 23.5 shows a possible geothermal gradient beneath old oceanic lithosphere. It has an upper 125-km-thick conductive cap in which temperatures rise to 1000 °C. In younger

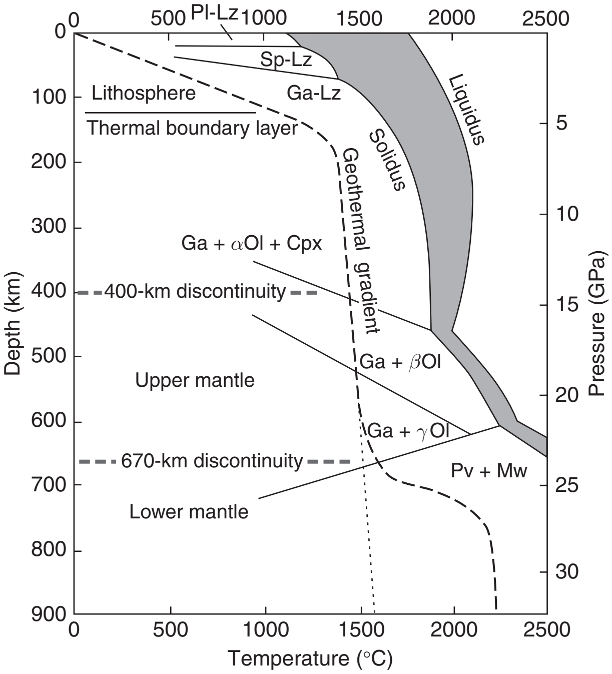


Fig. 23.5 Possible geothermal gradient (dashed line) resulting from separate convecting layers above and below the 670-km discontinuity and conduction in the lithosphere (after Richter and McKenzie, 1981). If there is whole-mantle convection, the geotherm would follow the dotted line. Also shown are the liquidus, solidus, and subsolidus phase relations for a spinel lherzolite (Lz) from Kilborne Hole, New Mexico (Takahashi, 1986). Pl, plagioclase; Sp, spinel; Ga, garnet; Ol, olivine (*α*, *β*, and *γ*); Cpx, clinopyroxene; Pv, silicate perovskite; Mw, magnesiowüstite.

lithosphere this temperature would be reached at shallower depths (Fig. 23.2). Beneath this, temperature continues to rise sharply through a thermal boundary layer at the top of the convecting upper mantle. Below this, however, temperature rises slowly until thermal boundary layers are encountered at the base of the upper mantle (670 km) where temperatures increase by 500 °C. A second geotherm (dotted) shows how the temperature would continue increasing slowly if the upper and lower mantle were not compositionally layered. Below the 670-km discontinuity, temperature again rises slowly until the thermal boundary layers at the core–mantle boundary are reached and there is again a large temperature increase of about 1000 °C. Experiments on the melting of iron under high pressures set the temperature of the core–mantle boundary at ~ 3500 °C, the outer core–inner core boundary at ~4700 °C, and the center of the Earth at ~ 6600 °C (Fig. 1.1). Also shown in Figure 23.5 is the continental geotherm calculated in Chapter 1.

The picture of convection associated with plate motion that seems to be developing is one in which the most prominent flow is associated with the sinking of the plates at subduction zones. These rafts of cold lithosphere are metamorphosed to dense eclogite (Fig. 22.12) on sinking into the mantle, which adds substantially to the gravitational imbalance resulting from the temperature differences. These slabs clearly descend to the base of the upper mantle and probably to the core–mantle boundary. Because they affect such a large fraction of the mantle, they must be long-lived phenomena. By contrast, the upwelling at divergent plate boundaries is passive and determined simply by the position of breaks in the plate. Spreading axes are therefore able to shift at any time. The upward flow that balances the downward flow at subduction zones could be distributed over broad areas, or concentrated in mantle plumes. The type of convection illustrated by the wax model in Figure 23.3 is therefore similar, with the downward flow being concentrated into narrow zones while the upward flow is spread out across broad zones.

Although the mineralogical constitution of the upper mantle is discussed in Section 23.5, it is worth commenting here on some important phase changes that may affect convection in the mantle and thus the shape of the temperature gradient. We do this in terms of a spinel lherzolite, which is commonly thought to have a composition similar to that of much of the upper mantle. This rock type has been investigated at high pressures by numerous workers, the data in Figure 23.5 being those of Takahashi (1986).

At low pressure, lherzolite contains plagioclase, but above 1 GPa, spinel becomes the stable aluminous phase and plagioclase disappears. Above 2.5 GPa, spinel disappears and is replaced by garnet as the stable aluminous phase. Below a depth of 100 km, then, the upper mantle consists of garnet lherzolite; that is, olivine + garnet + clinopyroxene + orthopyroxene. With increasing pressure, each of these minerals changes to denser phases, which may account for some of the seismic discontinuities.

The first mineral to change is olivine. At a pressure corresponding to a depth of about 400 km, *α* olivine transforms to a denser *β* phase, and if the rock also contains a spinel phase the reaction is univariant; that is, for a given temperature, the reaction would occur at a particular depth. Because olivine typically constitutes more than 50% of the rock, this transformation significantly changes the properties of the rock and probably causes the 400-km seismic discontinuity. Because the transformation takes place rapidly, it does not inhibit the flow of rock across the discontinuity and therefore does not affect convection. Pyroxenes also undergo a transformation to the *β* phase plus the high-pressure silica polymorph stishovite at about the same pressure, and this may contribute further to the change in properties at the 400-km discontinuity. With increasing pressure, both olivine and pyroxene undergo further changes to a spinel phase. At the same time pyroxene begins to dissolve in the pyrope-rich garnet as the majorite component [$Mg_3Al_2Si_3O_{12}$ + $M_3(MSi)Si_3O_{12}$, where M = Mg, Fe, Ca] as a result of some of the silica entering octahedral coordination and combining with a divalent cation to replace Al in the garnet.

The largest change in properties in the mantle occurs at the 670-km discontinuity. At approximately this pressure, the coordination of oxygen about silicon changes from four to six with the formation of a silicate perovskite [$(Mg,Fe,Ca)(Si,Al)O_3$] from olivine and garnet; the excess Mg and Fe in the rock forms magnesiowüstite [$(Mg,Fe)O$]. These two phases presumably are the major phases in the lower mantle. A still higher-pressure polymorph of perovskite, named post-perovskite, may exist at the base of the mantle (Murakami *et al*., 2004). Although the appearance of perovskite at pressures corresponding to the 670-km discontinuity appears to provide a simple explanation for this discontinuity, perovskite forms through a number of continuous reactions that would likely spread the transition out over a range of depth. This is not consistent with the sharpness of the seismic discontinuity, which may therefore require a compositional change for its explanation. At present, then, the experimental data do not resolve the question of whether convective currents can cross the 670-km discontinuity the same way they can the 400-km discontinuity.

Also shown in Figure 23.5 are the liquidus and solidus for this anhydrous peridotite. The geotherm, along much of its length, is within about 500 °C of the solidus. If the solidus were to be lowered by fluxing with water, the geotherm would first intersect the solidus near a depth of 150 to 200 km. This could certainly account for the position of the low-velocity zone beneath the lithosphere. If the lower and upper mantle convect separately, then the sharp rise in temperature through the boundary layers separating them brings the geotherm close to the solidus at this depth. Such a sharp rise might cause melting at this depth and give rise to a second low-velocity layer. The lack of such a low-velocity layer has been used as evidence for whole mantle convection. However, recent experimental work (Fig. 23.5) shows that the solidus rises steeply once silicate perovskite becomes a stable phase, reducing the possibility of the geotherm reaching the solidus at this depth.

Lherzolite melts over a 650 °C temperature interval at atmospheric pressure. With increasing pressure, the liquidus and solidus approach each other and at 17 GPa are separated by less than 100 °C. This indicates that lherzolite is almost a eutectic composition at this pressure. Is this a coincidence? Probably not (Walker, 1986). It seems likely that early in Earth history the upper mantle was formed as a near-minimum melt in equilibrium with solids at a depth of at least 500 km. This melting, which would have been related to the early release of heat from the Earth (core formation etc.), would have produced the magma ocean (Sasaki and Nakasawa, 1986). It is of interest to note that the liquidus temperature decreases as the pressure increases from 10 to 15 GPa (Fig. 23.5). Olivine, which is the mineral on the liquidus under these conditions, would therefore have floated (see Herzberg, 1987). This part of the mantle might therefore have solidified from the top down, leaving a residual liquid at its base.

23.3 CONDITIONS NECESSARY FOR ROCK GENERATION

Rocks are formed when changes in environmental conditions (T, P, X_{fluid}) cause changes in the phases constituting the Earth. This requires that material cross reaction surfaces in T–P–X_{fluid} space. Changes in the environment that do not result in reaction surfaces being crossed do not produce rocks. Surface rocks on the Moon, for example, undergo daily fluctuations in temperature but no new rocks are formed because no reactions take place under those conditions. Only one reaction surface, the solidus, is of importance in the generation of igneous rocks, whereas many reaction surfaces, most involving devolatilization, are crossed in the formation of metamorphic rocks. Reaction surfaces are functions of T, P, and fluid composition. Changes in any of these variables can therefore result in the formation of new rock. It is all too easy to lose sight of this important fact. For example, because magma is hot it seems reasonable to suppose that it is the product of heating. We will see, however, that most magma is formed not by heating but by decompression, which actually involves adiabatic cooling.

Reactions that form igneous rocks and most metamorphic rocks are accompanied by increases in entropy and volume because of the production of liquid or fluid. They consequently have positive slopes in terms of P and T, at least at low pressures (Eq. (8.3)). At higher pressures, the change in volume is smaller and the reactions become largely functions of T. At still higher pressures, the volume change may become negative, and the reaction will have a negative slope, as is seen to be the case for the liquidus of lherzolite (Fig. 23.5) or the stability limit of amphibole and phlogopite (Fig. 23.14). Because of these changes in slope, variations in environmental conditions can have different effects at different depths in the Earth. A decrease in pressure at one depth, for example, could cause melting, whereas at another it could cause solidification.

We will now examine how changes in temperature, pressure, and fluid composition can occur. But first, it is important to keep in mind that the Earth is a dynamic planet that tends toward a state of *dynamic equilibrium* or *steady state*. Thus, when we talk of changes in the environment we are not referring to a change from one set of static conditions to another, but rather to a disturbance of a set of conditions that over an extended period of time have established a dynamic equilibrium. Once the conditions are perturbed, fluxes of heat, stress, and matter try to reestablish the initial dynamic equilibrium. Rocks are products of the relaxation of these perturbations.

Let us start by considering how changes in temperature might occur. Nowhere is the principle of dynamic equilibrium more evident than in the development of the geothermal gradient. Heat is continuously generated in the Earth and transferred to the surface, where it is radiated into space. Because of the self-regulating nature of heat transfer, the rate of heat production in the Earth is balanced by the rate of heat loss. At any given depth, the rate at which heat is transferred depends on the mechanism of heat transfer and the thermal properties of the material at that depth. Where heat is transferred by the slow mechanism of conduction, as in the lithosphere for example, the temperature gradient must be steep in order to transfer the heat coming from below and being generated from within. In the mantle, where heat is transferred by the more rapid mechanism of convection, the temperature gradient is correspondingly less. To change the temperature at any given depth, the geothermal gradient must be perturbed.

At dynamic equilibrium the temperature at any depth in the Earth remains constant as a result of the balance between the flux of heat out, the flux of heat in, and the quantity of heat produced at that depth (see heat conservation Eq. (1.7)). It is this balance that allowed us to calculate a steady state geotherm in Section 1.6. Changes in temperature can therefore result from changes in any one of these three components. We will first examine the heat production component.

Heat production can be positive or negative, depending on whether heat is generated or consumed. But only one form of heat production is independent of all other variables, and that is radiogenic heat. Despite the slow decrease in this quantity through radioactive decay, it can be considered a constant at the timescale of normal geologic processes. It therefore cannot be responsible for changing the temperature of a rock. Certain sedimentary and igneous processes may create local concentrations of radioactive elements that may perturb the gradient, but this involves a transfer of material and is not strictly a change in the heat-producing capacity of the initial rock. Similarly, thickening of the radioactive crust results in higher temperature, but this is because of a blanketing effect and not because of a change in the heat-producing capacity of a given volume of rock.

A number of reactions take place in rocks that either generate or consume heat. These certainly can affect the temperature of a rock, but their effect is secondary because they first require changes in other environmental factors for the reactions to take place. Melting of rock and dehydration of minerals both absorb heat, whereas crystallization of magma and coarsening of grain size liberates heat. In addition, mechanical energy can be converted to heat. In magma, for example, this occurs through the dissipation of viscous forces, but the effect is small unless velocities are very high (see ash flows in Section 4.4). In solid rocks, shearing generates heat, but this too is usually small unless concentrated in major shear zones (Section 17.5). Harrison *et al.* (1997) have calculated that, at the present slip-rate of 20 mm a^{-1} on the Main Central Himalayan thrust, shear heating above the thrust surface raised temperatures by as much as one hundred degrees. This could have caused high-grade metamorphism and, ultimately, melting to form granite by either of the first two reactions shown in Figure 23.19.

Changes in temperature must be caused mainly by changes in heat flux. In the rigid lithosphere heat is transferred mainly by conduction, but advection can be important locally or regionally. Circulating meteoric waters can cause cooling of near-surface rocks, as for example near mid-ocean ridges, and ascending metamorphic fluids can cause heating. Magma may also advectively introduce considerable heat into the lithosphere. Advective flow, however, is commonly channelized, so the final distribution of heat is still by conduction. The large-scale transfer of heat in the mantle is by convection, but this flow is laminar, so that conduction must still play a role in transferring the heat in and out of any given volume of rock. The rate of heat conduction depends on the temperature gradient. Changes in the gradient will change the temperature of the rock. The gradient above a given volume of rock can be increased if the crust is thinned, which can result from erosion or extension. This would lower the temperature of the volume of rock in question. Thickening of the crust, on the other hand, would decrease the gradient and cause heating. Intrusion of magma or increased rates of mantle convection beneath a given volume would increase the flux into the volume and cause heating. We conclude from the heat conservation equation that changes in temperature must result primarily from changes in the conduction of heat into or out of the volume of interest. But heat conduction through rocks or magma is extremely slow ($k_T = 10^{-6}\ \mathrm{m^2\ s^{-1}}$), so that phase changes resulting from temperature changes must be equally slow.

Next we examine how changes in pressure may occur. Most pressure changes in the outer part of the Earth result from the motion of lithospheric plates. These changes can be positive or negative. At convergent plate boundaries, subduction and lithospheric thickening result in increases in pressure. Where the lithosphere is thinned by extension, the pressures are decreased. Erosion and sedimentation redistribute surface loads, which also cause changes in pressure at depth. The advance and retreat of continental ice sheets can also cause small changes in pressure. Explosion craters formed by large meteorite impacts can result in rapid decreases in the load pressure beneath impact sites. This must have been an important process early in Earth history when the frequency of such impacts was far greater.

Because of the relative weakness of rocks to long-term stress (<30 MPa), changes in pressure take place almost immediately, as evidenced by the Earth's close approach to isostatic equilibrium. Because most changes are a consequence of plate tectonics, we expect pressure changes to occur at comparable rates to plate motion, which is of the order of 0.01 m a^{-1}. This corresponds to changes in pressure of 350 Pa a^{-1} if we assume an average density of 3.5 Mg m^{-3}. Changes in pressure are consequently expected to be far more rapid than changes in temperature.

We have seen that the presence and composition of a fluid can have profound effects on the temperatures and pressures at which rocks melt and metamorphic reactions occur (Chapters 11 and 16, respectively). The rates at which advection can introduce or change the composition of fluids depend on such factors as the pressure gradients on the fluids, the wetting of the minerals by the fluids, and the presence or absence of cracks. These matters were discussed at length in Section 21.2. We have also seen in Chapter 22 that advective fluids are capable of transferring heat (thermal Péclet number) as well as of mass at rates that are greater than those for heat conduction. Fluid flow regimes likely vary with depth in the crust. At shallow to moderate depths, convection of metamorphic fluids is possible (Fig. 23.6(A)). At deeper levels, flow is thought to be dominantly upward, although it could be diverted, for example, along permeable zones or layers (Fig. 23.6(B), (D)). Moreover, if fluids from a large area are focused into a conduit such as a regional fracture zone, fluxes in the conduit can be large and transport significant heat (Fig. 23.6(C)).

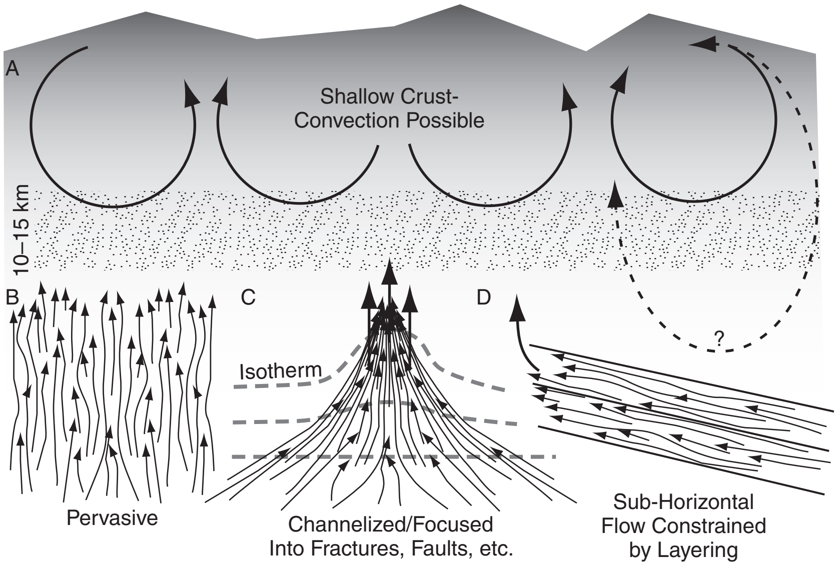


Fig. 23.6 Fluid flow regimes in the crust. (**A**) At upper crustal levels, convection is possible. Whether or not convection can penetrate into deeper levels of the crust (dotted path with question mark) is actively debated. (**B**) At deeper levels, flow is thought to be largely upward. (**C**) If flow is channelized into conduits such as regional fracture networks or fault zones, advective fluxes can be large enough to transport heat upward. Isotherms (dashed) would be distorted in areas of high flux. (**D**) Flow constrained by permeability contrasts along subhorizontal layering. (Modified from Ague, 2003b; published by permission of Elsevier.)

Advective fluids can therefore be important in the formation of rocks. However, it is important to point out that although fluid generation and flow in the lower crust are considerable during active metamorphism, once metamorphism stops the remaining fluids will migrate upward and leave the rock mass. Therefore, the lower crust in the quiescent parts of continents is probably dry (Yardley, 1986b).

The ultimate source of most fluids, at least during the last half of geologic time, has been the Earth's surface. Plate motion has transported this fluid in the form of hydrous and carbonate minerals into the Earth at subduction zones where rising temperature has caused its release into the overlying rocks during metamorphism. The isotopic evidence for metasomatism prior to the emplacement of magmas in oceanic islands (hot spots) may point to flushing with fluids enriched in incompatible elements derived from even deeper in the mantle. Some fluids may be primordial and come from outgassing of the planet. These are likely to have been more important in the early history of Earth.

We will now consider five different environments in each of which one of the variables above plays a dominant role in generating rocks. Lithospheric plate motion provides the ultimate cause for the changes which lead to the formation of rocks, but the rate of rock production depends on which of the environmental factors changes, with pressure changes producing rocks fastest, followed by advective changes, and finally, by thermal changes.

By far the largest production of igneous rocks in the world occurs at divergent plate boundaries. This normally involves the eruption of about $20\,\mathrm{km}^3\,\mathrm{a}^{-1}$ of MORB at mid-ocean ridges, but on numerous occasions throughout geologic time, voluminous flood basalts have also formed when the divergent boundaries have traversed continents to form large igneous provinces. In both environments magmas are formed as a result of decompression melting in zones of lithospheric extension (McKenzie and Bickle, 1988).

Figure 23.7(A) shows the distribution of isotherms prior to lithospheric extension. The geotherm (Fig. 23.7(B)) rises steeply through the lithosphere because of conductive heat transfer, but then flattens in the thermal boundary layer beneath the lithosphere, reaching a temperature of about 1480 °C at the top of the convective mantle. The geotherm approaches the lherzolite solidus but does not reach it under the conditions considered here.

When the lithosphere undergoes extension, it does so by thinning (Fig. 23.7(C)) and in the case of oceanic lithosphere this is accompanied by actual rupture at the spreading axis (Fig. 23.7(E)). On the lower surface of the lithosphere, mantle rises to compensate for the lost volume, whereas on its upper surface a sedimentary basin may form. The thinning, which in the continental lithosphere may occur by stretching (pure shear) or by movements on normal faults (simple shear) (White, 1989), results in the conductive geotherm being steepened and the convective geotherm being brought closer to the surface. This happens because the rate of tectonic transport is faster than the rate of heat transfer. Thus, each point on the preextensional geotherm moves vertically to establish the new geotherm. Eventually, this perturbed geotherm will relax and form a new steady-state geotherm, but that takes a considerable length of time (tens of Ma). In the interim, the perturbed geotherm has intersected the lherzolite solidus and produced magma.

The melting accompanying lithospheric extension is caused by the decrease in pressure as the mantle rises; it is not produced by heating of the mantle. Indeed, as the mantle rises it undergoes adiabatic cooling (Eq. (14.14)), which decreases its temperature by about $0.6\,°\mathrm{C\,km}^{-1}$ if the coefficient of thermal expansion is $4\times10^{-5}\,\mathrm{K}^{-1}$ and the heat capacity is $10^3\,\mathrm{J\,kg}^{-1}\,\mathrm{K}^{-1}$. The temperature is lowered still more once melting occurs because of the latent heat of fusion, which is $\sim 420\,\mathrm{kJ\,kg}^{-1}$ (Problem 23.7).

The amount of melting that takes place depends on how far the perturbed geotherm rises above the solidus. Extension in oceanic lithosphere is concentrated at spreading axes, and as a result the mantle rises a considerable distance with a

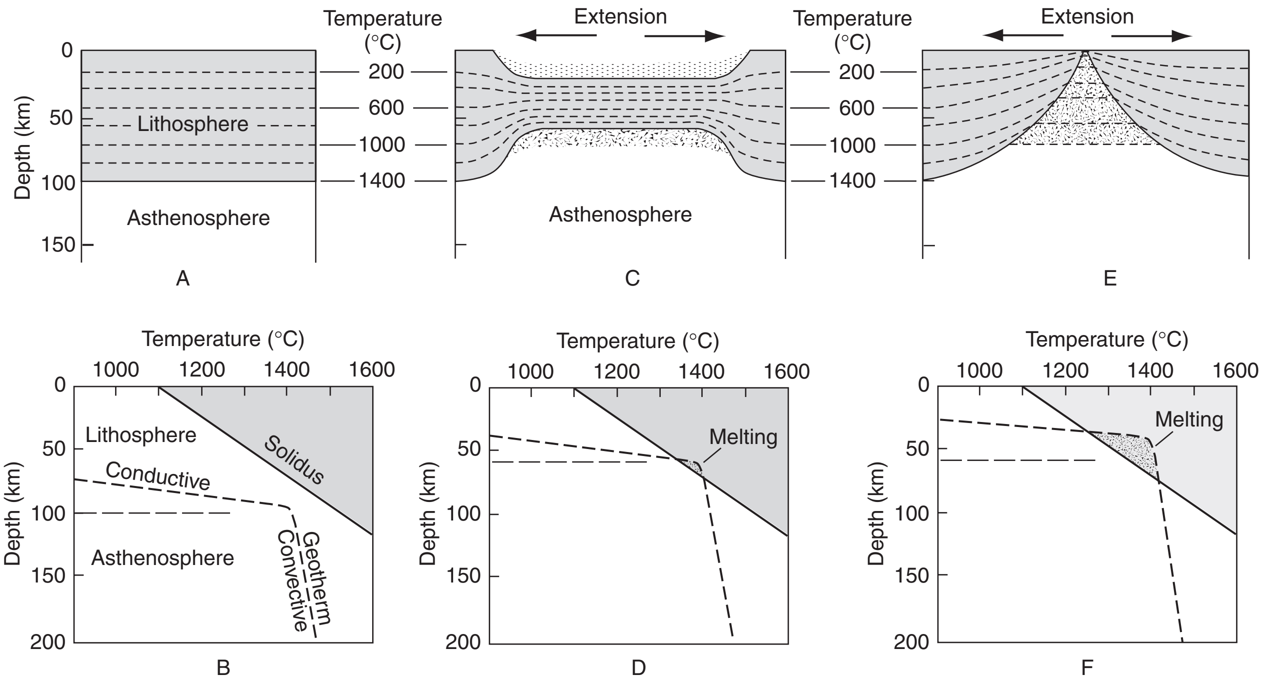


Fig. 23.7 Development of magma by decompression melting beneath regions of lithospheric extension. (**A**) Position of isotherms in a conductive lithosphere and convective asthenosphere prior to extension. (**B**) Geotherm prior to extension and simplified solidus of lherzolite. (**C**) Extension of the continental lithosphere results in the rise of the asthenosphere; this brings isotherms closer to the surface, and melting takes place through decompression (shaded area). (**D**) Rise of the geotherm beneath the zone of extension results in temperatures exceeding the lherzolite solidus. (**E**) If the lithosphere is ruptured during extension, the asthenosphere rises more, but over a narrower zone; this results in more extensive melting, as seen in (**F**).

consequent production of large volumes of melt (Fig. 23.7(E) and (F)). Extension of the continental lithosphere tends to be spread over greater distances. Consequently, the mantle does not rise as far, and correspondingly less melt is produced (Fig. 23.7(C) and (D)). The amount of melt formed in this case clearly depends on how much the continental lithosphere is stretched (Problem 23.8).

The compositions of the first-formed melt along the solidus varies with depth, as will be seen in Section 23.5. Where the geotherm intersects the solidus depends on the adiabatic temperature of the convecting mantle. This probably ranges from about 1300 to 1500 °C, depending on whether the mantle rises passively beneath the extending lithosphere or has a convective component rising from a deep mantle plume. In any case, temperatures, and consequently depths, cannot vary widely. This probably explains the limited range of compositions of MORB and continental flood basalts. Differences between MORB and continental flood basalts can probably be attributed largely to slight differences in the depth of melting and the degree of melting. In addition, the assimilation of felsic crustal material can exert important compositional controls on basalts in continental settings (Reiners *et al*., 1995).

If sufficient melt is produced in the extensional zone by adiabatic decompression, magma is likely to ascend buoyantly through the lithosphere to produce lavas and intrusive bodies. This advectively introduces heat into the lithosphere, which causes contact metamorphism. In addition, the steepened gradient caused by lithospheric extension results in an increased heat flux that is independent of the intrusion of magma. Because this heat flux depends on conduction through the lithosphere, it operates on a much slower timescale than the advective heat transfer associated with the emplacement of magmas. Extension of the lithosphere consequently sets up a number of perturbations that relax at different rates. This variability produces the complexity in the geologic record. We can illustrate this by considering the type of history that might be recorded in an extensional sedimentary basin, such as the Mesozoic ones formed in eastern North America just prior to the opening of the central Atlantic.

Once extension of the lithosphere begins, sediment starts accumulating in surface depressions (Fig. 23.8). Sedimentary rocks therefore provide the earliest record of extension. This is because of the speed with which isostatic adjustments take place. The surface changes are, of course, a reflection of movement of the mantle at depth, which, after rising far enough, intersects the solidus and melting begins. After sufficient melt has accumulated, magma rises into the basins to form lava flows and intrusive bodies around which contact metamorphic rocks are formed. Temperatures in the sedimentary basins are initially low and are only locally raised by the intrusion of magma. Eventually, however, the thinned lithosphere with its steepened gradient begins to raise temperatures at the base of the sedimentary pile. However, pore fluids in the sediments advectively remove this heat and distribute it throughout much of the basin, causing low-grade zeolite facies metamorphism, in particular where fluid

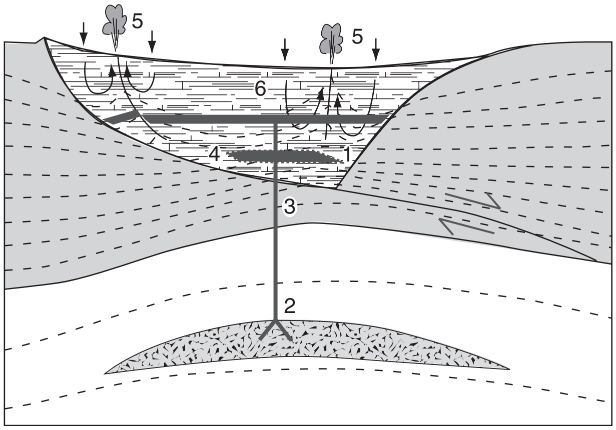


Fig. 23.8 Chronological record of extension of continental lithosphere. (1) Sedimentary fill of extensional basins. (2) Decompression melting of ascending asthenosphere. (3) Ascent of magma after sufficient segregation of melt in source region. (4) Eruption of lavas and intrusion of dikes and sills, and development of contact metamorphic rocks. (5) Elevation of regional temperatures as a result of lithospheric thinning causes zeolite facies metamorphism, in particular near fractures where convective hydrothermal cells develop. (6) Eventually, the thinned continental crust develops a geotherm that is shallower than the preextensional one. Final sedimentation occurs at this time.

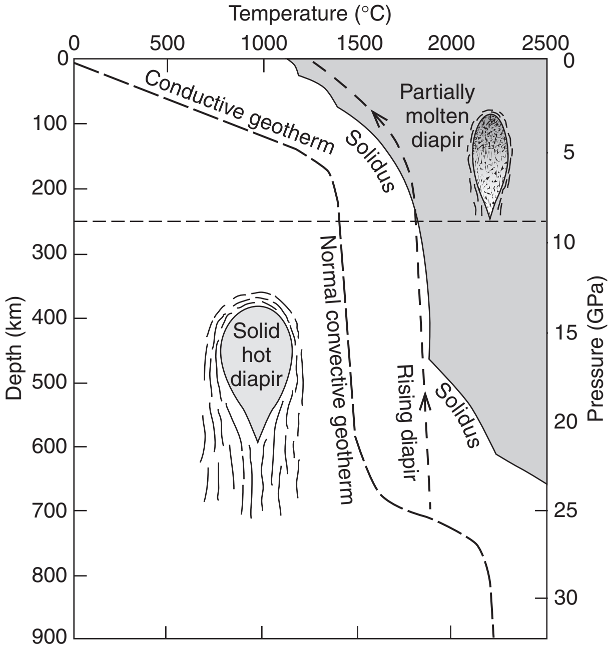


Fig. 23.9 A hot mantle diapir rising from the top of the lower mantle intersects the lherzolite solidus at a depth of 250 km. As it continues rising, the heat of fusion causes its temperature to decrease.

flow is channelized along faults. With time, temperatures gradually decrease until a steady-state geotherm is reestablished during which time the youngest sediments are deposited. Because the lithosphere is now thinner, radiogenic heat production is less, and the steady state geotherm is shallower than the preextensional geotherm.

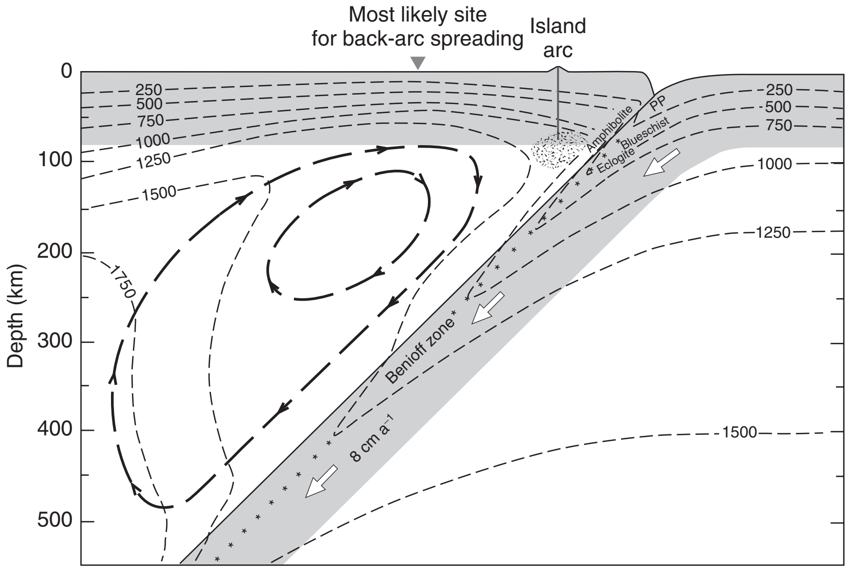
Because of the shallow slope on the convective geotherm in the upper mantle (Fig. 23.9), material starting from anywhere on the geotherm in the upper 670 km is not likely to intersect the solidus unless there is sufficient lithospheric extension and then the point of intersection is limited to the region where the convective geotherm becomes conductive (< 100 km). If, however, convection is driven by plumes rising from deep in the mantle, as presumably happens beneath oceanic islands such as Hawaii, melting can occur at greater depth. If diapirs are able to rise from below 670 km or from the thermal boundary layers between the upper and lower mantle, their higher temperatures will cause them to intersect the lherzolite solidus at greater depth. Once partial melting takes place, the lowered viscosity allows the diapir to rise more rapidly. Large amounts of melting, however, do not take place, because the adiabatic cooling keeps the magma near the solidus, which at these greater depths has a shallow slope. Because of the greater depth of melting, these magmas have a different composition from the shallow MORB. Here again the cause for melting is the change in pressure on the ascending mantle and not heating.

We will now consider the case where advection of fluids plays the dominant role in forming rocks. At convergent plate boundaries, cold lithosphere is subducted into the mantle, thus lowering the average temperature in these regions (Fig 23.10). It seems paradoxical, therefore, that such regions are almost always marked by volcanoes and high-temperature metamorphic rocks. Why should regions that would be expected to be cool show such clear evidence of heat? The explanation lies in the mechanism of heat transfer. Throughout most of a lithospheric plate, heat is transferred mainly by the slow process of conduction. Above subduction zones, advective fluids released from metamorphic reactions bring about partial melting, and bodies of magma advectively transfer heat into the overlying lithosphere. Thus, although the average temperature of the Earth at convergent plate boundaries may be low, rising bodies of magma and metamorphic fluids advectively transfer heat to much higher levels in the crust than would be possible through conduction alone.

The ocean floor subducted at convergent plate boundaries has had a long history of cooling and hydrothermal alteration. Following generation at an oceanic ridge, the basalt and intrusive rocks are intensely altered by convecting hydrothermal waters, which circulate through the newly formed crust. Ophiolite suites presumably provide an example of the end product of this alteration (Section 15.2). When this oceanic crust is subducted, increased pressures and temperatures lead to its progressive metamorphism through the blueschist to eclogite facies (Section 16.4). In doing so, most of the water that was taken into the oceanic crust during its alteration is released and rises into the mantle above the Benioff zone, where it causes metamorphism and melting (Figs. 23.10, 23.11). For a full discussion of this topic, see the review by Wyllie (1988).

Before tackling the problem of melt generation above Benioff zones, it is important to have some understanding

Fig. 23.10 Calculated isotherms (short dashes) resulting from convective flow (heavy long dashes) in the mantle wedge overlying a subduction zone (Toksoz and Hsui, 1978). Convection in the mantle wedge is driven by coupling with the subducting plate which is moving at 8 cm a^{-1}. Locus of earthquakes (Benioff zone) is shown by dots. Ocean floor rocks are progressively metamorphosed as they are subducted. Melting occurs (shaded area) when Benioff zone exceeds 100 km, probably as a result of water rising into the mantle wedge from the subducted plate. This magma rises to form island-arc volcanoes. If convection in the wedge is great enough, the oceanic lithosphere can be thinned and eventually ruptured at the top of the convection cell to produce back-arc spreading. PP, prehnite–pumpellyite facies.



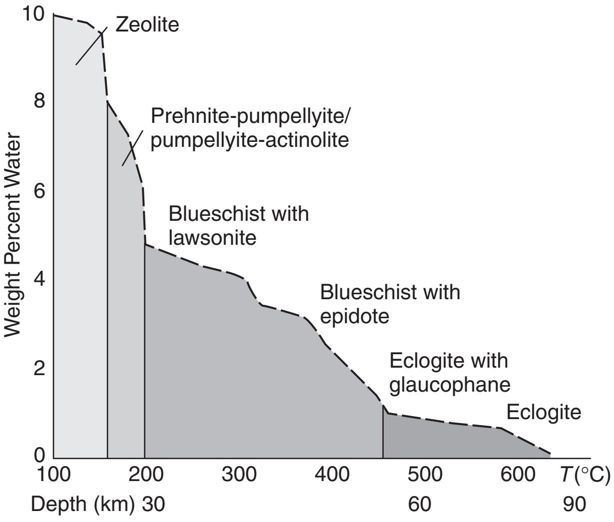


Fig. 23.11 Highly schematic metamorphic water release curve for subducted metabasalt. Weight percent water is the water content of the rock, which decreases with progressive dehydration. Initial water content is high at low temperatures in the zeolite facies; this water was added during hydrothermal metamorphism in the mid-ocean ridge environment. Water would be retained to greater depths if the geothermal gradient was smaller than depicted here, and it would be depleted at shallower depths if the gradient was larger. (Modified from Fyfe *et al.*, 1978.)

of the thermal structure of such regions. As already seen in Section 23.2, the temperature distribution in an oceanic plate can be adequately described in terms of conductive cooling of a half space, with convective heat influx at the base of the lithosphere for ocean crust older than 70 Ma. Cooling stops, however, once the plate begins to subduct, and its temperature steadily rises as heat flows in from the mantle above and below the plate.

An important question relating to convergent plate boundaries is whether the subducting plate shears past the overriding plate or is coupled to it. In the first case, the rock in the overlying mantle wedge remains stationary and slowly loses heat to the subducting slab, a condition that is hardly conducive to the generation of magmas. On the other hand, if the overlying mantle is coupled to the subducting plate, a convective motion is set up, which continuously cycles new hot mantle into the wedge (Fig. 23.10). This mechanism has the advantage that in addition to keeping the mantle wedge hot, it provides an explanation for back-arc spreading, which is present behind some but not all island arcs (Toksoz and Hsui, 1978). The convection in the mantle wedge may put sufficient traction on the base of the lithosphere to cause extension and rupture even though compression is occurring simultaneously near the oceanic trench. Analogous motion has been witnessed in cooling Hawaiian lava lakes when slivers of solid crust break off and are pulled toward large slabs of crust that are sinking into the still molten lava lake (Duffield, 1972).

Figure 23.10 shows the calculated distribution of isotherms in the convecting mantle wedge above the Benioff zone following 75 Ma of subduction at a rate of 8 cm a^{-1} (Toksoz and Hsui, 1978). Although the isotherms in the wedge are much nearer the surface than in the subducted plate, nowhere do they intersect the anhydrous solidus of lherzolite (Fig. 23.12). If convection in the wedge were strong enough to rupture the lithosphere, the upwelling mantle would intersect the lherzolite solidus and MORB-like rocks would be formed above the convecting cell at a back-arc spreading axis. But this would not produce melting beneath the island arc, nor would the subduction of the oceanic basalts, which are much too cold to melt.

In Section 14.3 we saw that igneous activity above the Benioff zone does not start until the subducted slab has sunk

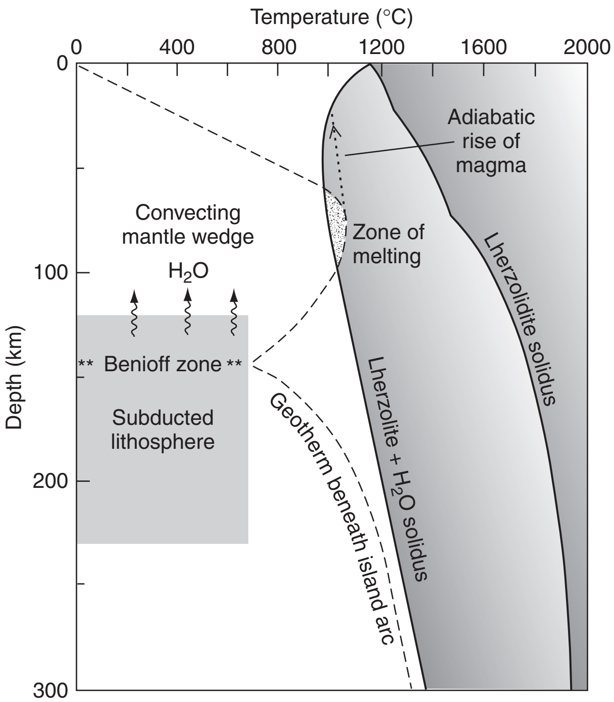


Fig. 23.12 Geotherm beneath the island-arc volcanoes in the model of Figure 23.10. The water-saturated solidus of lherzolite (Kushiro *et al.*, 1968) intersects the geotherm between depths of 60 to 100 km. The water necessary for melting is liberated by dehydration reactions in the subducted oceanic lithosphere.

to a depth of 100 km, with most island arc volcanoes being located about 120 km above the Benioff zone. This relation is independent of the rate or angle of subduction, but if the angle decreases below about 25° there is no volcanism at all. These characteristics provide important evidence about the way in which magmas are generated at convergent plate boundaries. Because the position of the volcanoes is independent of the rate of subduction, the temperature of the subducted slab cannot be a critical factor in the generation of the magmas. The strong correlation with depth indicates that pressure may be an important factor. Metamorphic reactions, in particular those giving rise to the eclogite facies (Fig. 16.6), are pressure sensitive, but they are also functions of temperature and thus should be dependent on the rate of subduction. Also, if pressure were the controlling factor, plates subducting at angles of less than 25° would eventually reach the same pressure as those sinking more steeply and magmas should be generated; but they are not.

Because the correlation between the position of volcanoes and the depth to the Benioff zone is so strong, other important factors can be overlooked. If heat is needed for the associated magmatism, it must come from convection within the mantle wedge. This convection produces a sharp bend in the isotherms above the Benioff zone where it goes below a depth of 100 km, with temperatures exceeding 1000 °C between depths of 60 and 100 km (Fig. 23.10). The position of this temperature maximum is determined mainly by the depth below the Earth's surface rather than the angle of subduction. If the subducting plate plunges more steeply, the bend in the isotherms is gentler but occurs at about the same depth. As the angle of subduction decreases, the bend in the isotherms becomes sharper, and at some critical angle, the convecting mantle would not be able to rise into the wedge, and temperatures would fall rapidly. This could explain why volcanoes are not present when the subduction angle is less than 25°.

Temperatures slightly in excess of 1000 °C at depths of 60 to 100 km are not high enough to cause melting of dry mantle, but they are if water is present (Fig. 23.12). By the time the subducted plate has reached a depth of ~120 km it has been metamorphosed to eclogite assemblages, and the water that was initially present in hydrous minerals has been released and migrated into the overlying wedge (Figs. 23.10, 23.11). The formation of magma in this region can therefore be attributed to the advection of water, but the elevated temperatures caused by convection also play an important role. As the Benioff zone becomes still deeper, temperatures in the overlying wedge increase, but by this time the subducted slab has been dehydrated, and thus no further melting occurs.

Although melts can form at the water-saturated solidus of lherzolite at temperatures near 1000 °C, they cannot rise to the surface. A melt formed at a depth of 75 km and 1060 °C, for example, adiabatically cools to a temperature of 1005 °C on rising to a depth of 20 km, where it intersects the solidus and crystallizes (Fig. 23.12). This has led some petrologists to invoke still higher temperatures in the mantle wedge, so that melts that are undersaturated in water can be formed (Marsh, 1979). It should be pointed out, however, that the calculated isotherms in Figure 23.10 are based on convection in a solid, viscous mantle with no account taken of the effects of melting. Once melting occurs, the decreased viscosity modifies the position of the isotherms and promotes still more melting. Magma rising from 75 to 20 km, for example, elevates the isotherms in its path and raises the temperature of the source region. If the lithosphere is ruptured and back-arc spreading starts, isotherms in the lithosphere would again be raised. Also, if the lithosphere contains thickened continental crust containing radioactive elements there would be additional heat. Although it is not certain how all these factors affect the production of magmas at convergent plate boundaries, there is little doubt that water plays a dominant role (Ulmer, 2001).

Advective transfer of heat by ascending bodies of magma is the most important means of raising lithosphere temperatures above the values that normally result from conduction. Most large bodies of silicic magma in the continental crust are now recognized to derive their heat from basaltic magmas that rise from the mantle (Section 14.11 and 15.3, and Fig. 15.8). Because of density, many basaltic magmas probably spread laterally at the base of the crust to form large sills. Heat liberated by the cooling and crystallization of these bodies raises the temperature of overlying rocks, producing upper amphibolite to granulite facies metamorphism. The heating is also sufficient to generate melts of granitic composition that are less dense than the basaltic magma and form gravitationally stable layers on top of the basalt with minimal mixing (Huppert and Sparks, 1988; Annen *et al.*, 2006). The granitic melt, however, is unstable with respect to the overlying crust and may rise into it, thus advectively transferring the mantle-derived heat into the crust. This heat brings about further regional metamorphism at middle to upper crustal levels during magma ascent (Sections 16.4, 22.7).

The rate at which the basaltic magma transfers heat to the granitic melt is geologically rapid, being in the tens to hundreds of years (Huppert and Sparks, 1988). This is because all but the thinnest basaltic sills (< 10 m) convect, thus keeping their roof rocks at high temperatures. Initially, the roof rocks are simply heated, but eventually melting of a granitic fraction takes place. The actual temperature at which this occurs depends on the availability of H_2O. As the zone of melted rock thickens, its Rayleigh number increases to the point where convection occurs in the granite as well (Eq. (14.10)). This then allows the granitic layer to transfer heat more rapidly to its upper surface and cause more melting and metamorphism.

The heat required for melting the roof of the granitic layer is supplied initially by the underlying convecting basaltic magma. This, however, cools the basalt, which, on becoming more than 60% crystallized (Eq. (2.9)), becomes too viscous to convect. From then on, the granitic layer must cool, but it is sufficiently superheated that convection continues for some time. Even melting of the roof continues as heat is liberated by the crystallizing granitic magma. Convection of this crystal mush continues until it is 60% solidified. After this, heat can only be transferred by conduction.

If the sheets of granitic melt formed above basaltic sills are thick enough, they become unstable relative to the denser overlying crust, and diapiric bodies rise from them. This advectively transfers heat still higher in the crust. The height to which the diapirs rise depends on many factors (Section 3.11), but eventually they are halted by the cooler, more viscous crust, and then heat transfer into the surroundings is largely by the slow process of conduction (advective meteoric waters may play a role near the surface). The rate of heat transfer from the mantle by advection through the basalt and granite (order about 10^2 to 10^3 years) is so much faster than the rate of heat loss by conduction (order about 10^4 to 10^5 years) that the thermal gradient in the vicinity of the diapirs is greatly steepened and high-T, low-P metamorphic rocks are formed.

Geochemical and isotopic data can be used to set timescales for the processing of material through convergent plate boundaries (Turner *et al.*, 2000). Although magmas may be generated by partial melting of peridotite in the mantle wedge above the Benioff zone, they show enrichment in LIL elements relative to HFS elements (e.g. Ba/Th), which points to addition of material from either the hydrothermally altered oceanic crust or the sedimentary cover. Because the sediments have relatively high $^{87}Sr/^{86}Sr$, and the magmas with the highest LIL/HFS elements have the lowest $^{87}Sr/^{86}Sr$, the contaminants are thought to consist largely of fluids expelled from the subducting oceanic crust rather than sediments. However, the presence of ^{10}Be in some lavas is unequivocal evidence of contamination with sediments (Morris *et al.*, 1990). ^{10}Be is formed in the atmosphere by the nuclear interaction of cosmic rays with oxygen and nitrogen. It then becomes concentrated in sediments. It decays to ^{10}B with a half-life of 1.5 Ma. The source of the sedimentary geochemical signature is actively debated. It may be fluids released during metamorphic dehydration (particularly if there were interactions with ultramafic mélange; Section 21.12.2), partial melts of subducted sediments, or some combination.

When fluids are released from the subducted slab, they should be enriched in the soluble U relative to the less soluble Th. If the fluids contribute to the melting in the overlying mantle wedge, they will disturb the U–Th decay series, which will take about 380 ka to reestablish equilibrium (see Section 15.3 and Fig. 15.11). Before equilibrium is reestablished, the series define isochrons that give the time since the release of fluids from the subducted slab. Isotopic analyses from a large number of arcs indicate that this time varies from 10 ka (Aleutians) to 200 ka (New Britain), with 60 ka being a common number (Turner *et al.*, 2000). The relatively short time between the release of metamorphic fluids and the eruption of lava is strong evidence that fluid flux is critical to the generation of these magmas.

The addition of metamorphic fluids can also produce ^{226}Ra–^{230}Th isotopic disequilibria. Like U, Ra is more soluble in fluids than Th. Recent studies show that island arc lavas commonly have significant excess ^{226}Ra (Turner *et al.*, 2001). Turner *et al.* infer that fluid addition to the mantle wedge increases Ba/Th. Thus, because the extent of ^{226}Ra–^{230}Th disequilibria in the lavas is correlated with Ba/Th, Turner *et al.* conclude that fluid addition also gave rise to the isotopic disequilibria. ^{226}Ra has a half-life of only 1600 years, so the lack of equilibrium suggests extremely rapid fluid addition, melt generation, and melt migration to the surface. For some arcs, the time between fluid release and eruption of lava may be as little as a few hundred years (Turner *et al.*, 2001). The ^{226}Ra–^{230}Th disequilibria thus imply magma ascent rates as great as $1000\,\mathrm{m\,a^{-1}}$. Such rapid rates would require channelized flow through fractures, instead of slow percolation through a porous network (Turner *et al.*, 2001).

Timescales of a few hundred years are shorter than those inferred based on the U–Th system. One explanation for this discrepancy is that Ra and U were introduced into the mantle wedge by different fluids separated in space and time. The excess ^{238}U would reflect early fluid addition whereas the excess ^{226}Ra would reflect the last increment prior to melt ascent (Turner *et al.*, 2001).

The above discussion hinges on the assumption that the mantle is in secular isotopic equilibrium for ^{226}Ra–^{230}Th prior to the addition of metamorphic fluids. However, Feineman and DePaolo (2003) argue that steady-state ^{226}Ra–^{230}Th disequilibrium can exist in the mantle wedge. This would mean that the disequilibrium measured in the lavas does not record the time of fluid infiltration; instead, it would record the time of partial melting. Thus, there could be a time lag between fluid addition and partial melting. If such lags exist, then the time between fluid addition and eruption could be considerably longer than a few hundred years. In any case, the presence of excess ^{226}Ra in the lavas still requires that melts ascend rapidly once they are formed.

Finally, let us consider the case where conditions in the Earth are changed by crustal thickening. This has already been considered in Chapter 22 in connection with *P–T–t* paths of metamorphic rocks. Initially, crustal thickening lowers the geotherm, because tectonic processes are faster than the rate at which heat conducts through rock. Eventually, however, the geotherm is elevated by the thickened crust, which behaves like a layer of insulation on the Earth. Furthermore, thickening of the layer containing abundant radioactive elements provides additional heat. The elevated geotherm eventually returns to its former steady state once erosion or tectonic processes have thinned the crust to its original thickness.

Calculated *P–T–t* paths followed by rocks during crustal thickening by a single thrust fault are shown in Figure 22.8. At depth, these paths increase in temperature while pressure decreases due to erosion of the thickened crust. They cross numerous metamorphic reactions, each of which involves a heat of reaction and possibly the evolution of fluid that could advectively remove heat from the rock. Neither effect was considered in calculating the paths, in part because of complexity, but also because of uncertainties in some of the quantities. The effects on solid metamorphic rocks are still matters of research and debate, but there is little doubt that partial melting will have a large impact. First, the enthalpy change on melting is large – about 300 kJ kg^{-1}, and second, once the rock is melted, it may rise and advectively transfer heat. *P–T–t* paths will therefore be strongly affected if significant amounts of melt are formed.

The lowest temperature at which melt can form is defined by the water-saturated granite solidus (Fig. 22.12). The amount of melt formed at this temperature is unlikely to be great, simply because large quantities of water are not available. Whatever water is present, however, enters this minimum melt. Because of the slope on the water-saturated solidus, these melts are unable to rise in the crust without solidifying. Water-saturated granitic melts are therefore not likely to be volumetrically or thermally important.

If *P–T–t* paths reach higher temperatures, fewer water-saturated melts can form. If the amount of water available in the rock at grain junctions is negligible, the next important melting curve traversed by *P–T–t* paths is that of “dry” muscovite granite (Huang and Wyllie, 1973). The only water present in this rock is bound up in muscovite (0.6 wt% of rock), which is released on melting. The solidus is consequently for a granitic liquid containing 0.6 wt% H_2O (review Section 11.5 and Fig. 11.5). Unlike the water-saturated solidus, this solidus slopes in the opposite direction, so that the fraction of liquid in partially melted rock steadily increases with decreasing pressure, at least to a high level in the crust (about 5 km).

The melting process buffers temperatures near the solidus. *P–T–t* paths cannot go far beyond this until all rock that is capable of melting has done so. The attainment of temperatures near 800 °C, but no higher, in many granulite facies rocks may be explained by this buffering effect. Note that after all the melting constituents are removed temperatures can continue to rise if sufficient heat is present, ultimately forming ultrahigh-temperature metamorphic rocks at 900–1000 °C. Magmas that rise from the zone of melting adiabatically cool by about 0.4 °C km^{-1}; in addition, further cooling accompanies the decompression melting due to the latent heat of fusion (about 300 kJ kg^{-1}). Consequently, magma remains close to the solidus during ascent. On reaching a depth between 10 and 5 km it becomes water saturated and must crystallize, releasing the heat that was imparted to the magma by fusion during ascent. This results in an advective influx of heat and fluids into the upper crust, where they cause high-*T*, low-*P* metamorphism.

23.4 GENERATION AND ACCUMULATION OF MELTS

Because seismic evidence reveals that the Earth’s mantle and crust are almost everywhere solid, the formation of bodies of magma requires a deviation from the norm. This may result from changes in pressure, advective heat transfer, fluid fluxing, or thermal insulation – their importance probably decreasing in that order. Given sufficient change in any of these factors, a rock crosses the solidus, and a series of steps ensue that may lead to the development of a body of magma. First, melt is formed where the phases involved in the solidus reaction come together. The enthalpy of fusion is sufficiently large that only small fractions of the rock melt – usually much less than 30%. For a partial melt to form a body of magma, it must first form a continuous phase throughout the rock so that the melt can flow. Finally, the melt must be able to separate from the source rock in a reasonable period of time, which requires that permeabilities be sufficiently high. Only then are bodies of magma formed that are large enough to have sufficient buoyancy to exceed the yield stress of the overlying lithosphere. In this section, we examine these magma-forming processes.

The first melt in a rock must form where minerals involved in the solidus reaction come together. The reaction will involve eutectic, peritectic, or cotectic assemblages. In each case, it involves the maximum number of phases in the rock, which typically come together at grain corners (normally four) or grain edges (three); grain faces juxtapose only two minerals. The first melt appears as small globules at grain corners or along grain edges. As the amount of melt increases, especially if it flows, the site of melting is less specific. At first, however, melt is restricted to multigrain corners.

Once melt is formed between the grains of appropriate composition, equilibrium requires that the globules of melt have shapes that minimize the interfacial free energies with the juxtaposed grains. This is achieved by the grains adjusting their area and angle of contact with the liquid by dissolving and reprecipitating from the melt. At equilibrium, a characteristic dihedral angle develops, which is determined by the ratio

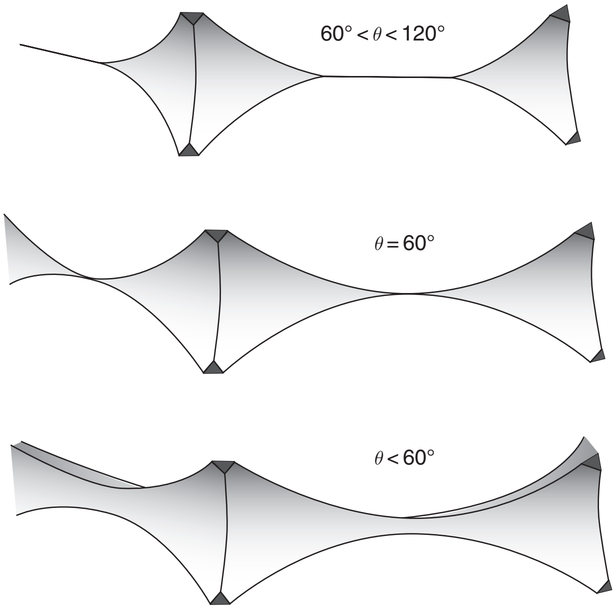


Fig. 23.13 Shapes of pockets of liquid formed between grains during partial fusion as a function of dihedral angle, θ, between grains and liquid. Silicate melts form dihedral angles with most minerals of less than 60°. As a result, even small fractions of melt form an interconnected network of channels through the rock along grain edges.

of the surface energies on each of the different interfaces (Eq. (12.48)). If the dihedral angle is greater than 120°, the melt globules remain as globules at grain corners with no penetration along grain edges (Fig. 23.13). If the dihedral angle is less than 120° but greater than 60°, the melt partly extends along grain edges; separate globules connect only after they have grown large enough as a result of higher degrees of melting. If the dihedral angle is less than 60°, melt extends all the way along grain edges, thus connecting corners with other grain corners to form a continuous network. Which of these configurations exists when rocks partly fuse is of obvious importance in determining how much melting is necessary before melt can be extracted.

A number of experimental studies have been carried out to determine the dihedral angle between silicate melts and common minerals, and in almost all cases the angle is less than 60°. The fact that these experiments were successful is, in itself, important because it indicates that if equilibrium textures can be achieved in the laboratory in hundreds of hours, nature most certainly would have achieved them in a timescale of years. Melts as varied as olivine basalt, ultramafic liquid, and granite have been investigated at pressures up to 2 GPa (Cooper and Kohlstedt, 1984; Jurewicz and Watson, 1985; Waff, 1986). Only orthopyroxene–orthopyroxene junctions with basaltic melt have a large dihedral angle (70°), but this decreases to 52° when the melt is water saturated (Fujii and Osamura, 1986). Quartz–quartz junctions with granitic melt (<0.2 wt% H_2O) were found to have the relatively large angle of 59° (Jurewicz and Watson, 1985). Holness (2006) measured dihedral angles in clumps of phenocrysts in natural melts and found that the angles varied from a maximum of 38° in clinopyroxene to a minimum of 18° in quartz (Section 12.7). These are all much less than the critical value of 60°. These results indicate that most silicate melts will form a continuous network at very small degrees of melting.

Despite the preponderance of experimental evidence that most dihedral angles are less than 60°, a number of natural rocks contain evidence to the contrary. Spinel lherzolite nodules in nephelinite from Dreiser Weiher, Germany, contain small isolated glassy or microcrystalline blebs that are almost spherical, suggesting that dihedral angles in this case must exceed 120° (Maaløe and Printzlau, 1979). Nicolas (1986) describes similar-shaped aggregates of orthopyroxene, clinopyroxene, and spinel in many lherzolite nodules from basalt. There is no obvious explanation for this discrepancy, other than that field observations are always open to various interpretations. As will be seen below, geochemical arguments indicate that some alkali basalts are formed by very low degrees of melting, which strongly suggests that the dihedral angles are less than 60°.

Where dihedral angles are less than 60°, even the smallest degree of fusion results in melt being distributed throughout the rock along a network of channels along grain edges, except, perhaps around orthopyroxene grains. Melt, however, is not able to penetrate along grain faces, and thus grains remain welded together while melt is able to flow through the channels. Only after the fraction of melt reaches 0.2 to 0.4 (depending on dihedral angle) can melt penetrate along grain faces and disaggregate the rock. Note that if the dihedral angles are less than 60°, the melt forms a continuous phase throughout the rock, regardless of how small the fraction of melt present. If this melt experiences a pressure gradient, it can theoretically be extracted from the rock, but the permeability would be very low if the fraction of melt were small. Also, if dihedral angles are less than 60°, melt will be dispersed along all grain edges regardless of where the melt is formed (Watson, 1982). In this respect, the rock behaves in an analogous manner to a piece of blotting paper or sponge. This also means that a certain amount of melt is needed along grain edges to minimize surface free energies, and this melt cannot be extracted from the rock. This fraction is on the order of 3%.

A body of magma forms by the separation and collection of melt from a partly fused source rock. The rate of this accumulation depends on permeability, which in turn is strongly dependent on dihedral angles and the degree of melting. If these angles are larger than 60°, the permeability will be zero up to a critical fraction of fusion where melt globules in neighboring grain corners first touch (Fig. 23.13). When the dihedral angle is less than 60°, however, the channels along grain edges always form a continuous network. As the dihedral angle decreases or the amount of melt increases, the cross-sectional area of the channels increases and so does the permeability (Fig. 23.13). Just as the flow of fluid through a cylindrical conduit increases with the square of its radius (Eq. (3.11)), so does the velocity through the trigonal prisms between grains increase with their cross-sectional area. Small increases in degree of melting result in large increases in permeability. For example, if grain diameters are 1 mm and the dihedral angle is 50°, the bulk permeability when only 0.01% melt is present is 10^{-16} m^2, whereas it is 5×10^{-15} m^2 when 0.05% melt is present, and 2.5×10^{-14} m^2 when 1% melt is present (von Bargen and Waff, 1986).

Because of the small dihedral angles, the permeability of partly molten rock is moderately high, even at small degrees of melting. For example, partly molten rock is far more permeable than metamorphic rock is to fluids (about 10^{-18} m^2; review Section 21.3). Knowing the viscosity of the melt and the driving pressures (buoyancy or volume expansion on melting), flow velocities can be calculated using Darcy's law (Eq. (21.2)). These velocities are greater than those at which plates move, and therefore significant bodies of magma can form in geologically reasonable times from small amounts of partial melt extracted from the source region (Problem 23.11). In Section 3.10, we saw that deformation of partially melted rock can cause the melt to segregate into small channels, which also increases the permeability (Holtzman *et al*., 2003, 2005). These small channels may then coalesce and form larger veins, as seen in some ophiolite complexes (Fig. 3.12 and Braun and Kelemen, 2002).

Although the grains in a partly molten rock remain in contact until the degree of melting exceeds 20% to 40%, they are not rigid particles but can deform both by creep and dissolution and redeposition through the melt phase (see pressure solution, Section 12.6). Thus because the average density of the solids is normally greater than that of the melt, the solids tend to compact and displace the melt upward. The rate at which the solid matrix deforms in large part determines the rate at which the melt is expelled (Section 3.9). Quantitative treatment of this process (McKenzie, 1984) indicates that in regions of upwelling mantle where there is more than a few percent of fusion, melt separates from matrix rapidly. Indeed, the process is so rapid that the percentage of melt in the source region at any given time is never likely to exceed more than a few percent, although up to 3% may have to remain in the source to satisfy surface energy requirements (the blotting paper effect).

The ease with which melts separate from a partly molten source plays an important role in determining the major compositional range of magmas and their trace element contents. Because the melt separates from the source after only a few percent melt has formed, equilibrium melting does not take place. Instead, the process approaches very closely that of extreme fractional melting (Section 10.4), and the trace element behavior obeys Rayleigh fractionation (Section 14.12 and Eq. (14.34)).

We conclude that melts separate from their source rapidly after only small degrees of fractional melting to form bodies of magma that rise toward the surface of the Earth. Near the source, rocks are ductile, so magma probably moves as diapirs, whereas at higher levels where the lithosphere is brittle dikes will be more important, especially if the lithosphere is undergoing extension (see Chapter 3). It has been suggested that in the source region magma may even move as waves of liquid passing through a porous medium. Analogous behavior can be created in fluidized sand at high pore-fluid flow rates. At some critical flow rate of water through sand, particles become fluidized and separate from one another (the quick condition). If the flow rate is increased still more, large bodies of relatively sand-free water rise through the sand. Particles fall rapidly through these regions, allowing waves to propagate upward without lifting the sand particles. This motion is different from that of a diapir because there is no net movement of sand particles during the passage of a wave. The supposed analogous waves of magma, which are named *magmons*, have wavelengths of kilometers and velocities of centimeters per year (Scott and Stevenson, 1986). The episodic activity in many volcanic regions may be due to the rise of magmons through the mantle.

23.5 COMPOSITION OF THE SOURCE OF MAGMAS

The compositions of magmas and the temperatures and pressures at which they form are strongly dependent on the chemistry and mineralogy of the rocks in the source region. For granitic magmas generated in the continental crust these source rocks are well known and form the metamorphic core of any deeply dissected mountain range. Magmas of basaltic composition, however, have deeper sources that are not accessible for direct sampling. Our knowledge of these rocks is therefore deduced largely from petrological, geochemical, and geophysical evidence. This section examines the limits that can be placed on the nature of the source of basaltic magmas.

Estimates of the composition of source regions can be made in three ways. First, rare samples of deep-seated rocks are brought to the surface as xenoliths in certain volcanic rocks. These indicate the types of rock penetrated during ascent of the magma and may actually provide samples of the source region. Second, experimental studies on basaltic magmas at high pressures indicate the phases that must be stable in the source region. Third, model compositions can be constructed that are capable of generating basaltic magmas while satisfying the general physical and chemical properties that the mantle is thought to have based on geophysical evidence.

Certain volcanic rocks of alkaline affinity, in particular those emplaced rapidly by explosive activity, may contain xenoliths of rocks through which the magma passed en route to the surface. While some fragments are clearly of local crustal rocks, others are of rocks that are completely foreign to the upper crust and have much deeper sources. Pressure-sensitive mineral assemblages indicate that these have come from depths ranging from the lower crust to as much as 200 km in the upper mantle.

These xenoliths provide the most direct evidence of the rocks from which magmas may be derived, but care must be taken in interpreting this evidence. Xenoliths may give a biased sample of the mantle. For example, they do not occur in all types of basalt, nor are all xenoliths likely to be preserved to the same extent. Mantle-derived xenoliths are restricted to alkaline rocks, especially to the ultra-alkaline

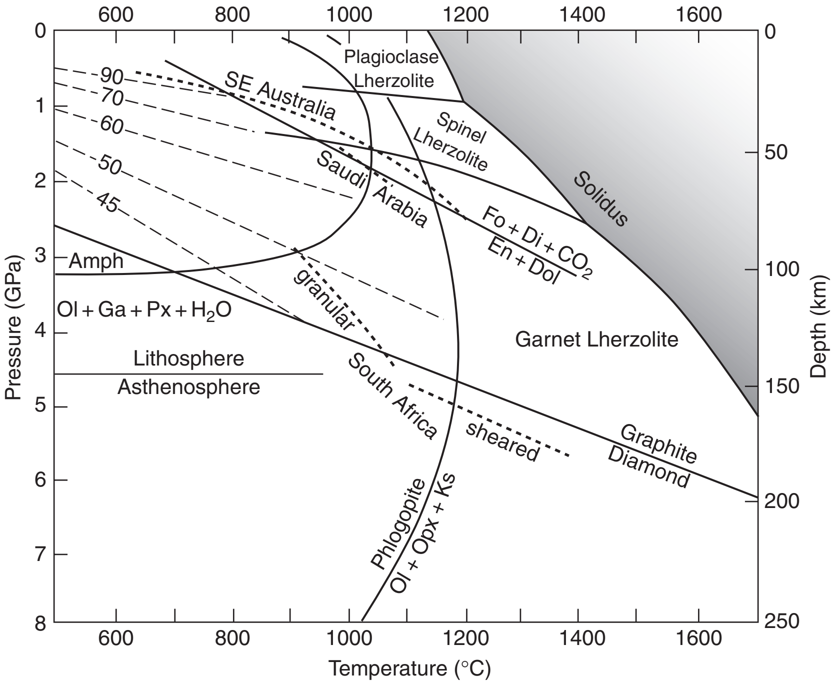


Fig. 23.14 Pressure–temperature conditions in the upper mantle based on xenoliths brought up in alkaline magmas (dotted lines). Data are from southeastern Australia (Griffin *et al*., 1984), Saudi Arabia (McGuire and Bohannon, 1989), and southern Africa (Boyd and Gurney, 1986). Conductive geotherms in the lithosphere are based on surface heat fluxes ranging from 90 to 45 mW m^{-2}. The phase boundaries are from the following sources: solidus (Takahashi, 1986); plagioclase- to spinel-lherzolite (Kushiro and Yoder, 1966); spinel- to garnet-lherzolite (Danckwerth and Newton, 1978; O'Neill, 1981); forsterite (Fo) + diopside (Di) + CO_2 to enstatite (En) + dolomite (Dol) (Eggler, 1976); graphite to diamond (Kennedy and Kennedy, 1976); amphibole and phlogopite stability limit (Kushiro, 1970).

silica-poor varieties, such as kimberlite, alnoite, and melilite nephelinites; they never occur in tholeiitic rocks. The ultra-alkaline rocks also tend to be restricted to continental regions, where there is rifting. All rock types passed through by these magmas may not be sampled equally, because of differences in physical properties. Moreover, some rocks may survive the trip to the surface more easily than others. Finally, although the xenoliths indicate the rock types picked up by a magma during ascent, they may not include samples from the source region itself.

Mantle xenoliths fall into two general types, extremely abundant peridotites (about 95%) and less abundant eclogites (about 5%) (Carter, 1970). The peridotites in kimberlites typically range from garnet lherzolite (ol + cpx + opx + gar) to harzburgite (ol + opx), whereas those in the alkali basalt–nephelinite series range from spinel lherzolite to harzburgite. The garnet and spinel lherzolites overlap in composition and can be shown to be equivalent rocks equilibrated at different pressures – the garnet lherzolite being the higher-pressure rock (see discussion below and Fig. 23.14). The eclogite xenoliths occur mainly in kimberlites and consist essentially of two minerals, pyrope-rich garnet and omphacite (diopside–jadeite).

In some areas, sufficiently large numbers of xenoliths have been derived from a range of depths that it is possible, through the use of geothermometers and geobarometers, to construct geothermal gradients. These allow us to examine the stratigraphy of the upper mantle. Because the xenoliths are brought to the surface so rapidly, their minerals are literally quenched, leaving no time for reequilibration.

Figure 23.14 shows the geotherms beneath southeastern Australia, Saudi Arabia, and southern Africa deduced from the mantle xenoliths. The lherzolite xenoliths from southeastern Australia indicate a steep geothermal gradient, especially at shallow depths where the indicated temperatures would correspond to a surface heat flux of 90 mW m^{-2} (Griffin *et al*., 1984). Temperatures indicated by the deeper garnet lherzolite xenoliths correspond to a more normal surface heat flux of 60 mW m^{-2}. The high temperatures of the shallow xenoliths are interpreted to result from the rise of diapirs from mantle plumes. The xenoliths from Saudi Arabia also indicate higher than normal temperatures. The mantle here has risen passively from depth in response to the opening of the Red Sea rift (McGuire and Bohannon, 1989). By contrast the xenoliths in the kimberlites from southern Africa (Boyd and Gurney, 1986) indicate a much shallower geotherm, but one that is consistent with ancient shield areas (Fig. 1.5 and Pollack *et al*., 1993).

Mantle xenoliths vary considerably in composition, and most are not suitable source rocks for basaltic magmas. Indeed, many appear to be refractory residues formed by the removal of basaltic fractions; this is particularly true of those of harzburgitic composition. It will be recalled from Section 15.2 that harzburgites that underlie ophiolite suites are also interpreted to be the residual mantle from which MORBs have been extracted. As the contents of clinopyroxene and garnet or spinel increase, the more likely is the rock to be able to generate a basaltic fraction. Eclogite xenoliths would therefore appear to be ideal candidates for basaltic source rocks. Experiments, however, indicate that eclogites have near-eutectic compositions and are simply formed from basaltic magma that has crystallized at high pressure. They can also be formed from basaltic rock that has been transported to depth where it has recrystallized into an eclogite assemblage. Eclogites are consequently not the source of basaltic magmas (Yoder and Tilley, 1962). The spinel and

Table 23.1 *Composition of mantle materials*[a]

			Composition of first-formed melts in Kilborne Hole lherzolite at:					
	Pyrolite	Kilborne Hole lherzolite	1 atm	1 GPa	3 GPa	8 GPa	14 GPa	Peridotitic komatiite
SiO_2	45.20	44.48	54.2	49.2	46.9	46.6	45.2	44.9
TiO_2	0.71	0.16	0.7	0.6	0.9	0.2	0.2	0.2
Al_2O_3	3.54	3.59	15.1	17.7	11.0	4.6	4.2	5.3
FeO	8.47	8.10	5.6	6.7	7.8	8.8	7.9	10.4
MnO	0.14	0.12	0.1	0.1	0.2	0.2	0.2	0.2
MgO	37.48	39.22	8.4	9.5	19.2	34.9	37.8	33.6
CaO	3.08	3.44	12.3	11.4	12.2	3.9	3.7	5.0
Na_2O	0.57	0.30	2.1	2.9	1.2	0.3	0.3	0.4
K_2O	0.13	0.02	0.0	0.0	0.0	0.0	0.0	0.0
P_2O_5	0.06	0.03	0.0	0.0	0.0	0.0	0.0	0.0
Cr_2O_3	0.43	0.31	0.1	0.1	0.4	0.4	0.5	0.3
NiO	0.20	0.25	0.0	0.0	0.0	0.0	0.0	0.0
Total	100.01	100.02	98.6	98.2	99.8	99.9	100.0	100.3
CIPW norms								
Q	0.0	0.0	5.22	0.0	0.0	0.0	0.0	0.0
or	0.77	0.12	0.0	0.0	0.0	0.0	0.0	0.0
ab	4.82	2.54	17.77	24.54	10.15	2.54	2.54	3.38
an	6.72	8.39	31.78	35.28	24.63	11.20	10.11	12.67
cpx	6.46	6.70	23.60	17.21	28.63	6.45	6.51	9.65
opx	15.81	12.57	18.76	1.89	1.46	26.92	17.36	13.39
ol	63.31	68.86	0.0	17.99	32.63	51.82	62.37	60.39
chr	0.63	0.46	0.15	0.15	0.59	0.59	0.74	0.44
il	1.35	0.30	1.33	1.14	1.71	0.38	0.38	0.38
ap	0.14	0.07	0.0	0.0	0.0	0.0	0.0	0.0

[a] Model mantle composition: pyrolite after Ringwood (1975); spinel lherzolite, Kilborne Hole, New Mexico, and composition of first-formed melts, after Takahashi (1986); peridotitic komatiite after Arndt *et al*. (1977).

garnet lherzolites are far more likely candidates. An analysis of a particularly *fertile* spinel lherzolite from Kilborne Hole, New Mexico, is given in Table 23.1 (Takahashi, 1986).

If a basaltic magma is derived from its source without undergoing any compositional change (primary magma), and the solidus phases can be experimentally determined as a function of pressure, the minerals in the source region can be determined if the depth of origin can be independently estimated. Note that this does not mean that we would know the bulk composition of the rock, only the minerals that must be present in the source. Although the technique is simple in principle, most basalts, even those with high magnesium numbers, are found to have undergone some fractionation en route to the surface, and thus their high-pressure solidus phases may not be those found in the source. For example, consider an olivine tholeiite (no normative quartz) that is fractionated into the quartz-tholeiite field by early separation of olivine. High-pressure experiments on the quartz tholeiite would indicate erroneously that the source contains coesite instead of olivine.

Experiments reveal that the typical mineral assemblages of basaltic rocks change to different assemblages with only moderate increases in pressure. The most common constituent of crustal rocks, plagioclase feldspar, disappears by reacting with olivine to produce pyroxene and spinel (Kushiro and Yoder, 1966). The precise reaction depends on the proportions and compositions of the minerals (Problem 23.13). For example, here are three possible reactions between plagioclase and olivine:

$$\begin{aligned} &CaAl_2Si_2O_8 + 2Mg_2SiO_4 \\ &\quad = CaMgSi_2O_6 + 2MgSiO_3 + MgAl_2O_4 \end{aligned} \tag{23.4}$$

$$\begin{aligned} &CaAl_2Si_2O_8 + Mg_2SiO_4 \\ &\quad = CaAl_2SiO_6 + 2MgSiO_3 \end{aligned} \tag{23.5}$$

$$\begin{aligned} &\underset{\text{(plagioclase)}}{NaAlSi_3O_8} + \underset{\text{(olivine)}}{Mg_2SiO_4} \\ &\quad = \underset{\text{(clinopyroxene)}}{NaAlSi_2O_6} + \underset{\text{(enstatite)}}{2MgSiO_3} \quad \text{(spinel)} \end{aligned} \tag{23.6}$$

The clinopyroxene in Eq. (23.4) is diopside, whereas that in Eq. (23.5) is the Ca-Tschermak's molecule, and that in Eq. (23.6) is jadeite. These form a continuous solid solution known as *omphacite*. Because all of the minerals involved in these reactions belong to solid solution series, the conversion of a plagioclase-bearing rock to a pyroxene-spinel-bearing rock takes place over a range of pressures and temperatures. However, the range is small and is not particularly sensitive to temperature. For natural rock compositions, the transformation takes place at approximately 1 GPa at 1200 °C (Fig. 23.14).

At still higher pressures, the assemblage spinel + pyroxene becomes unstable and is replaced by the assemblage garnet + olivine (Danckwerth and Newton, 1978; O'Neill, 1981). The reaction can be written in terms of the magnesian end members as

$$\underset{\text{(spinel)}}{MgAl_2O_4} + \underset{\text{(enstatite)}}{4MgSiO_3} = \underset{\text{(pyrope)}}{Mg_3Al_2Si_3O_{12}} + \underset{\text{(forsterite)}}{Mg_2SiO_4} \tag{23.7}$$

The reaction in a natural spinel lherzolite (Takahashi, 1986) is found to take place at 2.5 GPa at 1400 °C (Fig. 23.14).

What can we conclude about the mineralogy of the source region of basalts from these experiments? At low pressures, olivine is normally the primary mineral on the liquidus of basalts, and above 2.5 GPa it should again be present along with garnet. Between 1 and 2.5 GPa, however, olivine could be absent, depending on the bulk composition of the rock. The common occurrence of xenoliths of garnet lherzolite in kimberlites and of its lower-pressure equivalent spinel lherzolite in alkali basalts, consequently, makes this type of peridotite the most likely parental material from which basaltic magmas are derived.

Because many mantle xenoliths, especially those of harzburgitic composition, were thought to be refractory residues from which basaltic fractions had been extracted, model compositions were invented that are capable of producing basaltic liquids and leaving a peridotitic residue. The most widely used model composition, which is known as *pyrolite* (= pyroxene + olivine ± pyrope garnet), was introduced by Ringwood (1975). It consists of three parts alpine peridotite (harzburgite – 79% Ol, 20% Opx, 1% Sp – see Section 15.2) and one part Hawaiian tholeiite (analysis 1 in Table 23.1).

Although fertile mantle xenoliths of pyrolite composition are rare, some do occur. Analysis 2 in Table 23.1 is of a spinel lherzolite xenolith from Kilborne Hole, New Mexico, which almost exactly matches pyrolite. This xenolith has been the subject of careful melting experiments up to pressures of 14 GPa (Takahashi, 1986). The results give a good idea of the composition of melts that can be expected from a pyrolite-like mantle. They are discussed in the following section.

Finally, we must consider the composition of possible fluid phases in the source regions of basaltic magmas. In the upper mantle, porosities and permeabilities are not likely to be high, and thus only extremely small quantities of a free fluid phase could exist. Larger quantities of fluid can be bound in hydrous and carbonate minerals, and should these be involved in devolatilization reactions, substantial volumes of fluid could be evolved. As was seen in Chapter 11, even small quantities of fluid can dramatically affect the solidus temperature of rocks and the composition of melts.

Surprisingly, the olivine crystals in many mantle xenoliths contain fluid inclusions, which are composed largely of CO_2. Care must be taken in assigning a mantle origin to these inclusions. Fractures in xenoliths formed during the decompression accompanying ascent from the mantle can be filled with volatiles derived from the surrounding magma at shallow depths. The fractures may then heal, leaving isolated fluid inclusions. These, however, form planar arrays that can be distinguished from inclusions trapped on the surface of crystals as they grow. There is no doubt that some of the CO_2 in xenoliths originated as a separate fluid phase in those parts of the mantle sampled by alkali basalts. However, the presence of CO_2 in the upper mantle is limited by its ability to react and form carbonates, and in the deeper parts of the upper mantle where conditions are more reducing, CO_2 may combine with H_2O to form CH_4 and C, the carbon existing as diamond.

Both amphibole and phlogopite occur in some mantle xenoliths, and carbonatites are certainly derived from the mantle. Both H_2O and CO_2 can therefore be derived from mantle minerals by appropriate reactions (review Section 11.7). Amphibole is stable to depths of about 100 km at high activities of water and low geothermal gradients, whereas phlogopite is stable to depths of about 180 km under similar conditions (Fig. 23.14). Phlogopite is of interest because it provides a source not only of water but also of potassium, which is otherwise not present in any significant amount in other minerals of lherzolite.

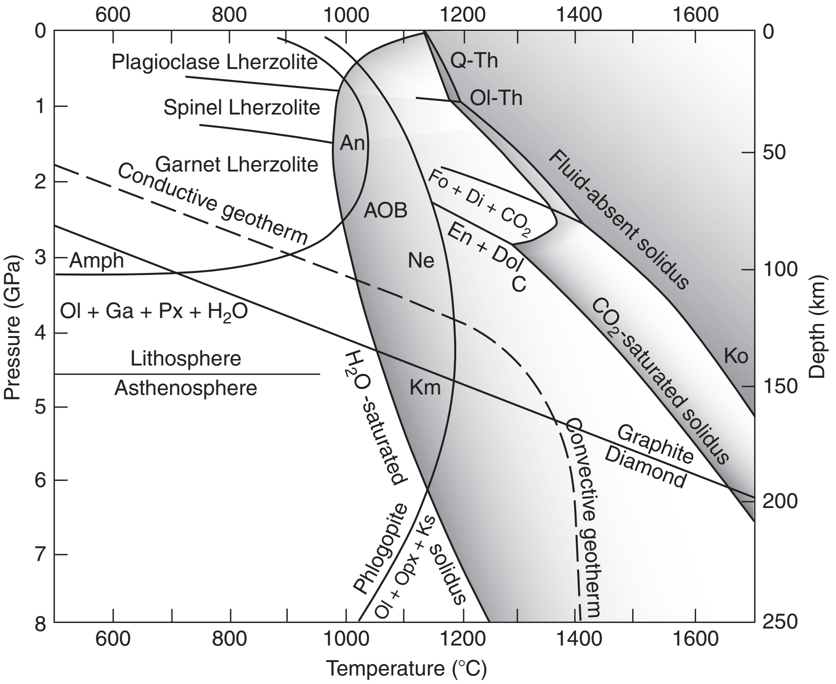
In the presence of a CO_2-rich fluid the common mineral assemblage of basaltic rocks – olivine + augite – reacts to form enstatite + dolomite (Eq. (11.37)) at pressures just inside the garnet lherzolite field (Fig. 23.14). Because this reaction lies at pressures above most geothermal gradients, CO_2 should react in mantle rocks to form carbonate. Why then should mantle-derived xenoliths contain CO_2 fluid inclusions? One possible explanation is that the magmatic event responsible for bringing xenoliths to the surface may also elevate the geotherm above the decarbonation reaction. This would not only produce free CO_2 but also provide a flux of gas capable of driving the explosive volcanism with which mantle xenoliths are so commonly associated. As will be seen in the following section, volatile-bearing magmas can form from mantle containing hydrous or carbonate minerals at temperatures well below those of the anhydrous lherzolite solidus (see also Section 11.7).

23.6 PARTIAL MELTING IN THE SOURCE REGION

Now that we have some idea of the materials present in the source regions of magmas, we are in a position to consider the actual melting process and the compositions of magmas formed. To do this we must know the phase relations between the minerals at the pressures of interest and the effects a fluid phase has on the composition and temperature of the magma. We discuss basaltic magmas first because they are the most abundant and because they may bear a parental relation to other magmas. This is followed by a treatment of granitic magmas, and finally of andesitic ones.

Melting in the mantle probably results from decompression rather than from heating. Decompression can result from the passive rise of the mantle into zones of lithospheric extension or from the rise of diapirs from deeper in the

Fig. 23.15 Approximate phase relations for a lherzolite mantle under fluid-absent, H_2O-saturated, and CO_2-saturated conditions. Sources of information same as in Figure 23.14, with H_2O-saturated solidus from Kushiro *et al*. (1968), and CO_2-saturated solidus from Eggler (1976) and Wyllie and Huang (1976). Probable source regions are indicated for the following rocks: quartz tholeiite (Q-Th), olivine tholeiite (Ol-Th), komatiite (Ko), andesite (An), alkali olivine basalt (AOB), nephelinite (Ne), carbonatite (C), and kimberlite (Km).



mantle, possibly emanating from mantle hot spots or plumes. In either case the temperature of the source region of basaltic magmas must be nearly that of the convective geotherm in the upper mantle, which on rising adiabatically to the base of the lithosphere has a temperature of 1400 to 1500 °C. If the temperature in the upper mantle cannot vary greatly, the actual depths at which basaltic magmas are generated depends largely on the position of the solidus, which, in turn, is strongly dependent on the presence and composition of fluids.

Figure 23.15 shows schematically the solidus for a lherzolitic mantle under fluid-absent, H_2O-saturated, and CO_2-saturated conditions. The highest solidus temperatures are obtained when no fluid is present (Takahashi, 1986). With addition of water, the solidus decreases dramatically and intersects the stability fields of amphibole and phlogopite (Kushiro *et al*., 1968). The solubility of CO_2 in basaltic melts is not great at low pressures, and consequently, the solidus of lherzolite is not lowered much by the presence of CO_2 up to pressures of 2 GPa. However, the solubility increases as the pressure approaches 2.9 GPa (90 km) and the solidus drops rapidly, reaching a low temperature at the reaction of forsterite and diopside with CO_2 to produce enstatite and dolomite (Eggler, 1976; Wyllie and Huang, 1976). At depths below this reaction, the solidus temperature again increases. For fluid compositions containing both H_2O and CO_2, the solidus is at intermediate temperatures between the curves for the H_2O- and CO_2-saturated solidus.

The solidus in a *P–T* diagram can never be a straight line if the assemblage of solid phases changes with pressure. This is evident from a consideration of Schreinemakers rules (Section 8.4). The solidus in a *P–T* diagram is a univariant line marking the first appearance of liquid. If it is intersected by a univariant reaction between the solid phases, the field occupied by the liquid must subtend an angle of less than 180° at the invariant point. Thus we expect to see cusps on the solidus where plagioclase lherzolite changes to spinel lherzolite, and spinel lherzolite changes to garnet lherzolite, and forsterite, diopside, and CO_2 react to form enstatite and dolomite (Fig. 23.15). These cusps are of importance because it is here that a perturbed geotherm that is raised toward the surface is most likely to intersect the solidus. The decreased variance at the cusps also restricts the compositional range that melts can have.

We have seen that the mantle beneath spreading oceanic ridges probably rises passively in response to lithospheric extension. If the geotherm in Figure 23.15 is raised toward the surface, and the mantle lacks significant quantities of fluid, the perturbed geotherm is likely to intersect the fluid-absent solidus at the cusp formed by the reaction between plagioclase and spinel lherzolite. The restricted range of composition of MORB may be due to the decreased variance at this cusp (Presnall *et al*., 1979).

As discussed in Section 11.7, the marked cusp in the CO_2-saturated solidus may play a role in the development of the ultra-alkaline, strongly silica-undersaturated igneous rocks and their explosive emplacement. If a perturbed geotherm rises in a mantle that contains a CO_2 or CO_2–H_2O fluid phase, melt first forms in the vicinity of the prominent cusp at the carbonation reaction. Experiments reveal that these melts are indeed strongly silica undersaturated. As the magma begins to rise buoyantly, the solubility of CO_2 decreases rapidly and the ensuing exsolution of gas could bring about explosive activity and the development of diatremes. If sufficient H_2O is present in the fluid, the cusp intersects the stability fields of phlogopite and amphibole.

These hydrous minerals would then appear as magmatic phases, as for example in kimberlites and alnoites.

The composition of melts formed at, or near, the solidus varies with pressure. This is best illustrated with the use of phase diagrams, but the effect is readily apparent from electron microprobe analyses of first-formed liquids. Table 23.1 includes analyses and CIPW norms of glasses formed in melting experiments at different pressures on the fertile spinel lherzolite xenolith from Kilborne Hole (Takahashi, 1986). The first-formed melt at atmospheric pressure has the composition of a quartz tholeiite, but above 1 GPa it is an olivine tholeiite. As pressure increases, so does the normative olivine content, and between 5 and 8 GPa, near-solidus melts containing from 30% to 35% MgO are identical to komatiitic magmas. Such high MgO contents can be formed at lower pressures only by large degrees of partial melting (> 50%), and as we have seen in Section 23.4, melts separate and segregate long before they attain such high fractions. We can conclude, therefore, that komatiites are most likely formed by small degrees of partial melting at depths of 150 to 250 km (Ko in Fig. 23.15). The convective geotherm would have to be at a much higher temperature than it is at present in order to intersect the lherzolite solidus in this depth interval. This probably explains the restriction of most high-magnesian komatiites to the Archean.

Another consequence of the increasing MgO content of the first-formed melts with increasing pressure is that by 14 GPa the melts have compositions that are nearly the same as that of the lherzolite as a whole. This indicates that the composition of the lherzolite must be near a eutectic or minimum at this pressure, which agrees with the observation made in Section 23.2 that the liquidus and solidus of lherzolite are separated by less than 100 °C toward the base of the upper mantle.

The composition of first-formed melts in fluid-saturated lherzolite also changes with pressure. The addition of water to lherzolite not only lowers the solidus dramatically, but it increases the silica content of the melt into the andesitic range (An in Fig. 23.15; see also Fig. 11.19). As pressure increases above 2 GPa, especially when the fluid contains CO_2 as well as H_2O, melts become nepheline normative and resemble alkali olivine basalts (AOB in Fig. 23.15). At still higher pressures and mole fraction of CO_2, the first-formed melts become increasingly undersaturated in silica, passing through nephelinitic to carbonatitic compositions (Ne and C in Fig. 23.15), as was discussed in Section 11.7. Finally, at very high pressures and with mixed CO_2–H_2O fluids, kimberlitic magmas form in equilibrium with phlogopite (Km in Fig. 23.15).

The compositional variation of melts with pressure and fluid composition can only be understood clearly through a study of multicomponent phase diagrams (e.g. Wyllie, 1988). These tend to be complex and are beyond the scope of this book. We will, however, consider a relatively simple system, that of silica–forsterite–nepheline, which contains a number of important minerals and invariant points that play roles in the generation of major rock types.

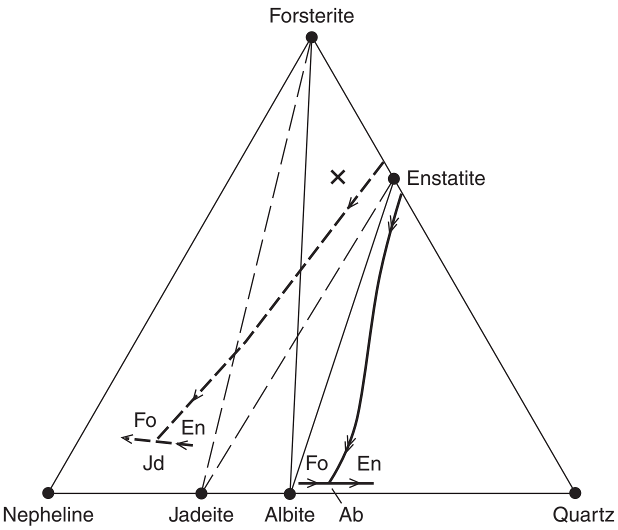


Fig. 23.16 System forsterite–nepheline–quartz showing position of boundary between liquidus fields of forsterite and enstatite at atmospheric pressure (Schairer and Yoder, 1961) and at 3 GPa (Kushiro *et al.*, 1968). Solid tie lines between phases refer to low pressure and dashed ones to high pressure. At low pressure, forsterite reacts with peritectic liquid to form enstatite and albite; at high pressures this reaction is reversed, with enstatite reacting with peritectic liquid to form forsterite and jadeite.

At low pressure in the system silica–forsterite–nepheline (Fig. 23.16), the join forsterite–albite is a thermal divide, with tholeiitic rocks plotting on the quartz side of the join and alkaline ones on the nepheline side. For clarity, the only low-pressure phase relations shown are those of the boundary curve between forsterite and enstatite and the ternary peritectic involving forsterite, enstatite, and albite (Schairer and Yoder, 1961). A plagioclase lherzolite that plots at the point marked X would produce its first melt at the ternary peritectic, which because it lies in the triangle enstatite–albite–silica would give rise to a quartz tholeiite magma.

With increasing pressure, the peritectic involving forsterite, enstatite, and albite shifts toward nepheline, and by 1.0 GPa it lies to the nepheline side of the forsterite–albite join. At these higher pressures, the peritectic reaction involves enstatite plus liquid going to forsterite plus albite. Above 2.5 GPa the mineral jadeite is stable, and at 3 GPa, the peritectic, which now involves reaction of enstatite with liquid to form forsterite and jadeite, plots in the triangle forsterite–nepheline–jadeite, and the join enstatite–jadeite is now the thermal divide in the system. The lherzolite of composition X consists of olivine, enstatite, and jadeite at 3 GPa and on first melting would produce a liquid at the ternary peritectic involving these three minerals. Intrusion of this liquid to a region of low pressure would result in formation of a strongly undersaturated nepheline normative magma.

Clearly, melting of a simple lherzolitic composition in this ternary system produces melts that range from quartz tholeiite to strongly undersaturated liquids depending on the pressure. The higher the pressure, the more undersaturated the melt is likely to be. Also, at constant pressure, introduction of a fluid phase rich in CO_2 has the same effect as increasing the pressure; that is, melts become more undersaturated in silica. Introduction of a hydrous fluid, on the other hand, has the same effect as decreasing the pressure; that is, the melts become richer in silica.

These changes in composition of the first liquid formed in this simple system from a lherzolitic starting material appear to explain the general compositions of a number of major rock types. MORB, for example, is formed at relatively shallow depths beneath spreading ocean ridges and has olivine tholeiite compositions (Ol–Th in Fig. 23.15). Basaltic magmas that come from greater depth, as evidenced by mantle xenoliths with high-pressure mineral assemblages, are alkaline, and those that are strongly undersaturated in silica contain high carbonate contents and presumably are formed in equilibrium with a CO_2-rich fluid. Magmas developed above Benioff zones are probably formed at depths near 100 km (about 3 GPa), and might therefore be expected to be undersaturated in silica if formed from a lherzolitic mantle. However, water released from the subducted slab would shift the first-formed melts into the quartz-saturated field and produce andesitic magmas. The water would also promote crystallization of amphibole in these magmas.

We turn now to the question of how the magma evolves as melting takes place. Does melting proceed in an equilibrium manner, or does it take place fractionally or perhaps through batch melting? The composition of the segregated magma that eventually rises into the lithosphere depends critically upon this stage. We will illustrate this process by considering melting in the simple pseudoternary system forsterite–diopside–pyrope at 4 GPa (Davis and Schairer, 1965).

Let us consider first the equilibrium melting of a garnet lherzolite of composition X (Fig. 23.17). On reaching a temperature of 1670 °C the first melt forms with the eutectic composition, E. With continued heating, the temperature remains constant while the fraction of eutectic liquid increases. Simultaneously, the composition of the remaining solid changes from X toward s_1. When the solids reach s_1, the rock is 33% melted (s_1X/s_1E) and no garnet remains. Consequently, the eutectic liquid begins to change its composition along the cotectic toward F, while the solids change their composition from s_1 toward fosterite. For example, when the liquid has composition l_2, the solids have composition s_2. When the liquid reaches l_3, the solid, s_3, reaches forsterite and thus no diopside remains. Consequently, the melt leaves the cotectic and moves toward olivine. As it does so, the temperature rises rapidly, and melts are unlikely to progress far in this direction because of the large amounts of heat required. For example, when the melt reaches l_4, the original rock is 66% melted (s_3X/s_3l_4) and the temperature is already 1800 °C. The result of equilibrium melting is a range of possible melt compositions along the line El_3X, with the actual composition depending on the temperature attained.

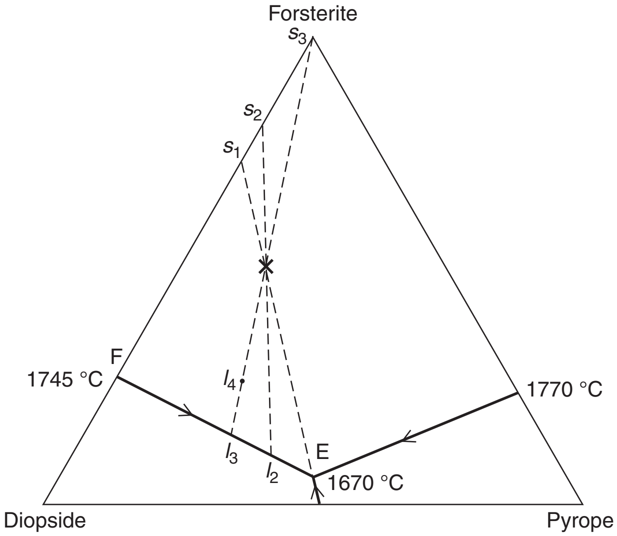


Fig. 23.17 Pseudoternary liquidus diagram for the system forsterite–diopside–pyrope at 4 GPa. The ternary eutectic, *E*, is actually a piercing point. See text for discussion of equilibrium and fractional melting of composition X. (After Davis and Schairer, 1965.)

Consider next the fractional melting of the same garnet lherzolite. Melting again begins at 1670 °C with the formation of a eutectic liquid at *E*, but this time the melt is removed as soon as it is formed. This results in the bulk composition of the rock changing from X toward s_1. One-third of the original rock can be extracted as eutectic liquid before the bulk composition of the rock reaches s_1. At this stage, the remaining rock consists only of forsterite and diopside; no garnet remains. But olivine and diopside can melt only at the binary eutectic, F, which has a temperature of 1745 °C. Melting therefore ceases after the removal of the last drop of ternary eutectic liquid and does not begin again until the temperature reaches 1745 °C. The liquid that forms at this temperature is very different from the first one and has the composition of the binary eutectic, F. Melting continues at this eutectic until all of the diopside has been consumed and then it ceases. Melting would not occur again until the temperature reached the melting point of pure forsterite, which is most unlikely. The result of fractional melting is a series of magmas with distinct compositions corresponding to invariant points. No magmas of intermediate composition are formed.

Melting in the mantle is likely to proceed under conditions that are intermediate between these two extremes. However, it will be recalled from the discussion of compaction (Section 23.4) that fractions of melt as small as a few percent are likely to separate from the source rock. We can conclude therefore that fractional melting provides a better model of natural melting than does equilibrium melting. Primary magmas, that is, ones derived directly from the source with no change in composition, are expected to have relatively fixed major element compositions, although several different compositions may be present due to melting at successive invariant points. Magmas that show a gradational range of compositions are likely to be the product of differentiation of the magmas after they leave the source region.

Although the major element composition of successive melts extracted from a source rock undergoing fractional melting may remain relatively constant, the same is not true of trace elements with small distribution coefficients (Section 14.12). Incompatible elements, such as P, K, and Zr, enter the first melt and are removed from the system; that is, they undergo Rayleigh fractionation (Section 14.12). The abundance of such elements could therefore be used to monitor the extraction of magma from a source region whereas the major elements could not (Problem 23.15).

Fractional melting of the mantle provides a means of producing magmas that are rich in volatiles. Although a separate fluid phase may exist in the solid mantle, it can constitute only an extremely small fraction of the rock. Equilibrium melting could therefore not produce a significant amount of fluid-saturated magma. On the other hand, if segregation of liquid takes place after only a few percent of melting, the fluid-saturated melt, which is bound to form first because of its lower melting point, will be able to collect and form bodies of magma.

Granitic melts can form directly from the partial fusion of a variety of continental crustal rocks, including pelites, arkosic sandstones, and granites (i.e. any rock containing quartz + plagioclase + K-feldspar or muscovite). This partial melting process (*anatexis*) is relatively well understood, in part because of extensive experimental studies on relevant synthetic systems (Section 10.19), but also because of field evidence. Most high-grade metamorphic terranes contain rocks known as *migmatites*, which are intimate mixtures of igneous granite and refractory metamorphic rocks (Fig. 23.18). Most migmatites are thought to form by partial fusion, but some may form through injection of granite into metamorphic rocks. Migmatites provide a glimpse of the partial fusion process frozen in progress, and though the melts are viscous and silicic, basaltic melts may segregate from mantle peridotites in a similar manner (see Fig. 3.12).

Fig. 23.18 Migmatite showing dark refractory amphibolite that was partly disrupted by the flow of the surrounding granite. Contrast this deformation with the brittle behavior of the rocks at a later time when a diabase dike was intruded (background). Coast of Labrador, near Hopedale.

Migmatites are striking rocks with light-colored granite veins anastomosing through dark-colored refractory metamorphic rocks (Figs. 3.13, 23.18). The contrast in color and composition indicates that partial fusion results in almost total segregation of the melt from the refractory residue. Granitic material is not dispersed along grain boundaries throughout the rock. Instead, it is segregated into veins and sheets, most of which parallel the metamorphic foliation. Migration of the melt is therefore channelized rather than dispersed. The distribution of the metamorphic rocks within the migmatite normally provides clear evidence that migmatites flow as coherent masses, with the solid fraction being stretched and pulled apart during flow. This movement may, in fact, be important to the segregation process, with liquid migrating into fractures or low-pressure regions where competent layers are necked down by stretching (Holtzman *et al*., 2003).

During progressive metamorphism, migmatites first appear toward the top of the amphibolite facies where temperatures exceed 650 °C (Fig. 16.6). Granite must contain about 4 wt% H_2O to melt at these temperatures (Whitney, 1988). Intergranular pore spaces can contribute no more than 0.3 wt% H O; the fluid must therefore be derived largely from other sources. Dehydration reactions provide the largest amount. Water is released locally by progressive metamorphic reactions, such as those accompanying the transition from amphibolite to granulite facies. The abundance of migmatites at this boundary is probably related to the availability of this water. In Benioff zones, water is released into the overlying mantle wedge by metamorphic dehydration reactions taking place in the subducted plate. Because the fluid may not wet mineral grains (Section 21.2), it may be forced upward through fractures generated by the pressure on the fluid evolved during dehydration and decarbonation reactions.

Because the amount of intergranular water in metamorphic rocks is very small, no significant amount of water-saturated melt can form from it. Instead, melting is delayed until dehydration reactions take place or water infiltrates the rock from an external source. When melting accompanies a dehydration reaction, the melt is undersaturated in water. The degree of undersaturation depends on the amount of water released by the reaction.

Three reactions are capable of releasing large quantities of water at temperatures where melting of a granitic fraction can occur in the crust (Whitney, 1988). The lowest temperature of these involves the reaction of muscovite and quartz to form sillimanite and K-feldspar (Eq. (16.1)). In Figure 18.4(A) this reaction is seen to intersect the water-saturated beginning of melting curve of granite at about 0.3 GPa and about 680 °C. Above this pressure, the reaction proceeds with the formation of an undersaturated melt ("Dry" musc. granite solidus in Fig. 22.12). If the rock also contains albitic feldspar, the melting and breakdown of muscovite will occur at a slightly lower temperature according to the reaction

$$\begin{aligned}&\text{muscovite} + \text{quartz} + \text{Na-feldspar}\\&\quad = \text{sillimanite} + \text{K-feldspar} + \text{liquid}\end{aligned} \tag{23.8}$$

This melting curve, which is shown in Figure 23.19, gives rise to peraluminous granites with K/Na ratios greater than 1 at temperatures below 750 °C.

The next reaction is similar but involves the breakdown of biotite according to the reaction

$$\begin{aligned}&\text{biotite} + \text{quartz} + \text{Na-feldspar}\\&\quad = \text{K-feldspar} + \text{pyroxene} + \text{liquid}\end{aligned} \tag{23.9}$$

Because this reaction does not involve muscovite, the melts are less aluminous and typically have K/Na ratios of 1 or less. The melts also range from granitic to granodioritic in composition and have temperatures between 750 and 850 °C.

The final reaction involves the breakdown of hornblende to pyroxene and a meta-aluminous to peralkaline melt with a low K/Na ratio at the amphibolite–granulite facies boundary. These melts have temperatures of 900 to 1000 °C, which probably require an influx of heat from a basaltic magma. With increasing pressure, the composition of the granite minimum (or eutectic) shifts to less quartz-rich compositions (Fig. 10.27). As a result, partial melting of crustal rocks at depths of about 50 km gives rise to syenitic magmas (Wyllie, 1977). Fractional melting of an amphibole-bearing source can produce both quartz and nepheline normative melts (Presnall and Bateman, 1973). Extraction of the first-formed melt, which is silica oversaturated, leaves a critically undersaturated residue, which in the next stage of melting produces nepheline normative magma.

The temperatures necessary to bring about the dehydration reactions might result from crustal thickening during orogenesis. However, as was seen in the discussion of *P–T–t* paths followed by metamorphic rocks, temperatures are not likely to go much above the temperature of the solidus associated with the breakdown of muscovite unless there is an additional heat source (Fig. 22.12). Most melting in the crust must therefore be associated with the advection of basaltic magmas from the mantle (Huppert and Sparks, 1988; Annen *et al*., 2006).

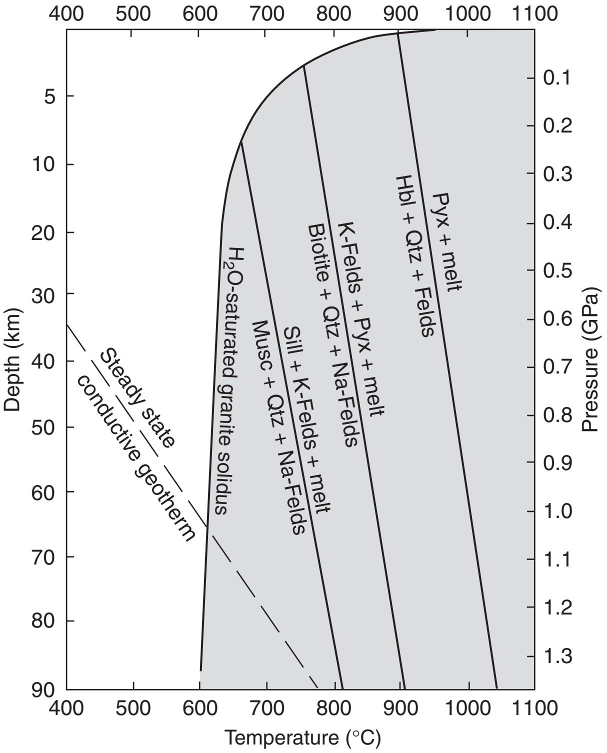


Fig. 23.19 Beginning of melting curves for granitic rocks. The lowest temperature melting is for water-saturated conditions. Little melting is likely to occur at this temperature because of the low porosity of the rocks. Instead, melting is more likely to occur under water-undersaturated conditions at temperatures marking the breakdown of muscovite, biotite, or hornblende. (After Whitney, 1988.)

Ascending basaltic magma on reaching the base of the continental crust is likely to spread laterally because of the lower density of the overlying rocks. The temperature of the crust at this depth is about 500 °C (Fig. 23.9). As the basaltic magma cools and crystallizes, the temperature of the overlying rocks rises and partial melting of a granitic fraction occurs (Huppert and Sparks, 1988). The temperature and amount of this melt depend on how much water is released by the crystallizing basaltic magma and how much is generated in the rocks from dehydration reactions. Formation of a cap of granitic magma over a basaltic magma prevents any subsequent intrusions of basalt penetrating to higher levels in the crust. Later intrusions of basalt pond at the base of the granite, possibly forming pillow-like structures (Fig. 14.38). If the basaltic magma mixes with the granitic one (Section 14.11), intermediate magmas of granodioritic to tonalitic composition may form (intermediate magmas may also form directly by partial fusion of crustal rocks). The low-density granitic cap to these bodies explains why in regions of granitic magmatism, basaltic dikes are either emplaced before the granitic magmas are formed, or they intrude only in peripheral regions, presumably beyond the extent of the granitic cap. Once the cap becomes large enough, bodies of silicic magma rise into the crust advectively transferring heat to higher levels and thus steepening the regional geothermal gradient.

Finally, we come to the question of the origin of calcalkaline andesites. No rock type has a more predictable mode of occurrence – it forms volcanoes near convergent plate boundaries where the depth to the Benioff zone exceeds 100 km, provided that the subduction angle is greater than 25°. Despite this certainty, there is no agreement on the source of andesites. Are they formed in the subducting slab, in the overlying mantle wedge, or even in the crust? Are they formed by partial melting of subducted MORB, ocean floor sediments, lherzolitic mantle above the Benioff zone, or do they require assimilation of crustal material? Geochemical evidence relating to these questions, which is given in Section 13.4, suggests, but does not prove, that andesites are derived from a mantle that is less depleted than the type from which MORBs are derived, and that a certain fraction of oceanic water must be involved and possibly some ocean floor sediment. This evidence favors the mantle wedge as the source of this magma. We will now consider whether this interpretation is supported by any other lines of evidence.

As discussed in Section 23.3, convergent plate boundaries must, in general, be the coolest parts of the Earth, because of the subduction of cool lithospheric plates. To develop magmas in the overriding plate, the subducting one must induce convection in the overlying mantle wedge. The calculated results of such convection are given in Figure 23.10 (Toksoz and Hsui, 1978), but the region in which magmas may form is given in more detail in Figure 23.20. First, note that temperatures in the subducted ocean floor rocks beneath the island arc are less than 750 °C and are consequently at least 200 °C below the water-saturated solidus at this depth. We can conclude, therefore, that the ocean floor basalts (eclogitic at depth) and their underlying sheeted dikes, layered gabbroic intrusions, and residual harzburgitic mantle rocks could not melt. Nor are the ocean floor rocks likely to rise diapirically into the overlying wedge where temperatures are higher, because once they are metamorphosed to eclogite their average density of 3.6 Mg m^{-3} is significantly greater than that of garnet peridotite, which is only 3.4 Mg m^{-3}. If sediment is subducted, a granitic component could melt; this would not account for basalts and andesites but might account for the sedimentary contaminant found in some andesites. Moreover, most sediment accumulates as an accretionary wedge in front of the arc rather than being subducted.

Temperatures in the mantle wedge above the Benioff zone are not high compared with normal steady-state geotherms, but nor are they low, despite their proximity to the subducting plate. They are prevented from falling by the convection induced by the subducting plate (Fig. 23.10). The temperatures in the wedge are not high enough to generate melts if no water is present. The fact that melts do exist in this region is therefore evidence that water must be present, and the subducted ocean floor rocks are the most obvious source of this fluid. It is significant that volcanism occurs where the Benioff zone goes below the depth to which amphibole is stable (dotted line in Fig. 23.20). Below this, essentially all water that had been bound in minerals is released as the rocks are converted to eclogite assemblages. The pressures generated by the dehydration reactions would cause the fluids to rise into the overlying wedge and fracture the rock if poor wetting ability hindered its flow. The short time interval between the release of these fluids and the eruption of lava (hundreds to tens of thousands of years) indicates that fracture flow may well play a role (Turner *et al.*, 2000, 2001).

Fluids rising from the Benioff zone enter progressively hotter rocks, and at approximately the 1000 °C isotherm they cross the water-saturated solidus of garnet lherzolite, and melting begins (region 1 in Fig. 23.20). The up-dip extent

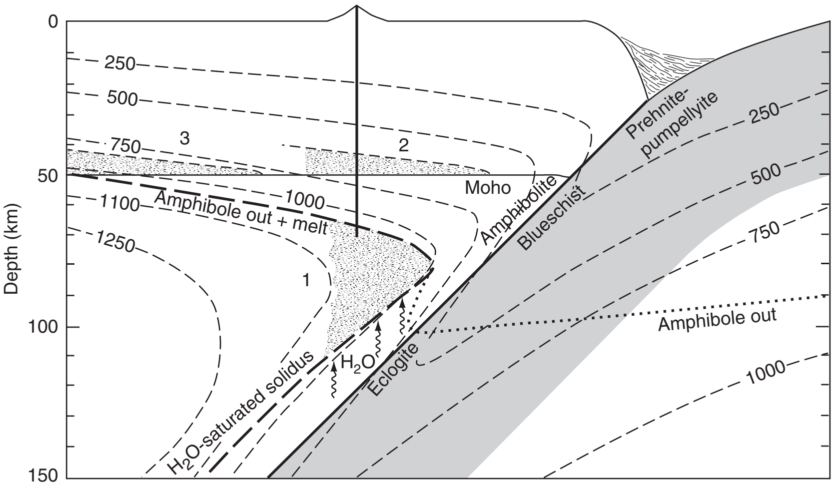


Fig. 23.20 Three possible regions of melting near convergent plate boundary based on positions of isotherms in convection model of Toksoz and Hsui (1978) (see Fig. 23.10). Region 1, water liberated from the subducted plate allows garnet lherzolite to melt under water-saturated conditions. Region 2, granitic melt could form in continental lithosphere only under water-saturated conditions (not likely; see text). Melting in this region is more likely to occur at higher temperatures as a result of the breakdown of mica or amphibole (see Fig. 23.19); this occurs when magma rises from region 1. Region 3, granitic melt formed during breakdown of muscovite can form in the lower continental crust at the top of convection cell; this might be associated with continental rifting.

of this melting region is limited by the position of both the isotherms and the stability field of amphibole (note the sharp bend in the dotted line in Fig. 23.20). The down-dip extent, however, is limited only by the availability of water. The melting region is bounded on its upper side by the solidus involving the breakdown of amphibole to a water-undersaturated melt. Thus the melts range from water-saturated in the lower part of this region to water-undersaturated at its top. Note that the isotherms in Figure 23.20 take into account only the convection induced by the subducting slab. Once significant volumes of melt are formed in the mantle wedge, the buoyant rise of this magma will distort the isotherms and extend the zone of melting vertically.

If the overriding plate has a continental crust, another region of possible melting occurs just above the Moho where water-saturated granite melts can form (region 2 in Fig. 23.20). This region is bounded on its lower side by the compositional Moho and on its upper side by the water-saturated granite solidus. Free water must be available for melts to form in this region. Water is not likely to enter directly from the Benioff zone because melting in the underlying mantle wedge consumes all available H_2O. However, these mantle-derived melts, on rising into the base of the crust and crystallizing, liberate water and heat, which promotes metamorphism and crustal melting. The lateral extent of this melting thus depends on the flux of magmas from the mantle.

A third possible region of melting occurs in the lower crust above the top of the convecting cell in the mantle wedge (region 3 in Fig. 23.20). This is the same place that crustal rifting and back-arc spreading may occur. The melting region is again bounded on its lower side by the Moho, but its upper surface is the solidus marked by the breakdown of muscovite to water-undersaturated melt. Fusion here is, therefore, not dependent on an influx of H_2O. Again, the positions of these regions of granite melting change as diapiric bodies rise into the overlying crust and distort the isotherms.

Estimated magmatic temperatures for andesites based on geothermometers and experimental studies range from a low of 850 °C for hornblende-bearing varieties to 1100 °C for two-pyroxene varieties (Gill, 1981). Allowing for adiabatic cooling during ascent, these temperatures indicate a likely magma source in the mantle wedge. This must also be the source of the commonly associated basaltic rocks.

Basalts were initially thought to be parental to andesites. This idea was based mainly on the almost continuous chemical variation of calcalkali rocks from basalt, through andesite, and dacite, to rhyolite. Experiments on andesites, however, have failed to provide any simple means of differentiating basaltic liquids into andesitic ones. If andesite is derived from basalt by fractional crystallization, minerals crystallizing in the basalt must be present as primary liquidus phases in the andesite. Magnetite is one mineral that must fractionate if basalt is to differentiated to andesite; magnetite crystallization reduces the iron content while increasing the silica content of the residual melt. Experiments indicate that under normal oxygen fugacities magnetite is a late-crystallizing mineral in andesite (e.g. Eggler and Burnham, 1973), and many andesites do not contain magnetite phenocrysts at all. These facts are clearly unreconcilable with andesite being derived from basalt.

In many island arcs, andesites contain lower crustal xenoliths of a variety of ultramafic to mafic rocks, including hornblende-bearing olivine clinopyroxenite, hornblende gabbro, and hornblende-free gabbro (Conrad and Kay, 1984). These cumulates record the differentiation of basaltic magma in the lower crust or upper mantle. The fact that andesite so commonly brings up these xenoliths shows that andesite itself has a deeper source than the site of this basaltic differentiation. Again this would argue in favor of andesite having an independent origin from basalt.

Experiments indicate that unless andesites contain small but significant amounts of H_2O ($>2\%$), crystallization sequences do not match those found in the natural rocks. This volatile content would certainly account for the large volumes of andesitic magma that are erupted explosively (see Section 15.3). Could the volatile content determine whether andesitic or basaltic melts are formed by partial fusion of the mantle? Melting experiments on garnet lherzolite indicate that it might (Kushiro, 1972). With increasing pressure, the composition of the first-formed melt becomes progressively less silicic, with nepheline normative compositions forming above 1 GPa (Fig. 23.16). However, if water is present, the melts are richer in silica, and at 2 GPa the first-formed water-saturated melt is andesitic in composition (Fig. 11.19). Thus basalt and andesite could be derived from a common source but under different activities of water.

The amount of metamorphic water available for melting must decrease as the Benioff zone deepens. The prevalence of andesitic volcanoes along the volcanic front of the arc could thus be explained if andesites require high activities of water for their generation. In deeper zones, where there is less water, magmas would be basaltic, with compositions becoming progressively more undersaturated in silica with increasing depth. This would explain the common distribution of basaltic rocks in island arcs with tholeiitic ones being closest to the convergent plate boundary and alkaline ones farthest from it, as is found in going from east to west across Japan (Kuno, 1960).

The origin of andesites remains uncertain, but fluids undoubtedly play a critical role in their formation. Fluids may also be important in the genesis of other igneous and metamorphic rocks. Isotopic data (Section 13.4) indicate that hot-spot magmatism is preceded by mantle metasomatism, which introduces incompatible elements into otherwise relatively depleted source regions (Menzies and Hawkesworth, 1987). If these elements are transported by fluid, as seems most likely given the low rates of solid diffusion, the fluid may at the same time act as a flux and cause melting. In this example rocks may be formed in response to mantle outgassing. If the upper mantle were completely melted during formation of the core, most of its fluids would have been expelled. Fluids in the upper mantle at subsequent times would therefore have to come from either subducted sediments or the lower mantle.

23.7 SUMMARY AND CONCLUSIONS

The formation of a rock, whether it be igneous, metamorphic, or sedimentary, is a rare geological event that requires extraordinary conditions. All rocks are formed directly or indirectly in response to perturbations in the steady-state geotherm, and they represent just one of the ways the Earth has of dissipating its internal heat.

Throughout most of the Earth, the geotherm is determined mainly by convection, except in the lithosphere where conduction is the principal means of heat transfer. Even here the oceanic part of the lithosphere is involved in convection through ocean-floor spreading and subduction. Perturbations in the geotherm can have shallow or deep origins. Continents, for example, preserve a complex record of changing lithospheric plate configurations, each of which would have had its own shallow thermal pattern. Seismic evidence indicates that subducted lithospheric slabs may descend as far as the core–mantle boundary, which must cause a counter flow from this same depth, perhaps in the form of mantle plumes. The ascent of this material not only transfers heat from the core to the upper mantle but may also transfer elements that would otherwise remain in an untapped lower mantle reservoir.

The upper mantle, although now largely solid, behaves plastically and is able to convect. Temperature gradients in the upper mantle are consequently nearly adiabatic and extremely shallow ($< 1\,^{\circ}\mathrm{C\ km^{-1}}$). This means that no matter from where material rises in the upper mantle its temperature, on reaching the base of the lithosphere, is approximately the same; that is, about 1400 °C. Throughout the upper mantle, the adiabatic gradient is about 400 °C below the fluid-absent solidus of garnet lherzolite, the rock type most likely comprising this part of the Earth. Because of the limited range of temperatures in this part of the mantle, melting is not likely to occur as a result of temperatures rising to the solidus. Instead, the solidus is more likely to be lowered to the geotherm by changes in composition. These changes could result from the rise of mantle diapirs of a different composition or by infiltration metasomatism of CO_2 or H_2O fluids.

In the conductive lithosphere and thermal boundary layer at the top of the convecting mantle the temperature gradient is steep, and consequently rocks that are displaced upward find themselves at temperatures that may be far enough above the geotherm to intersect the fluid-absent solidus of lherzolite and cause melting by decompression. This type of melting takes place in zones of lithospheric extension as a result of passive upwelling of the mantle.

Basaltic magmas are formed by partial fusion of peridotite in the upper mantle. Melting first takes place at grain corners where low-melting mineral assemblages are in close proximity. The dihedral angle between the melt and crystals is sufficiently small that the melt flows by capillary action along all grain edges. Once it forms a continuous network, it is able to rise buoyantly through porous flow while the residual solids undergo compaction. Channels or veins of melt probably develop, as happens with granitic melts in migmatites. Once large enough bodies of melt have coalesced, they may rise diapirically through the mantle if their density is less than that of the surroundings. Their ascent is normally halted at the base of the rigid lithosphere, and only when fractures can form easily is basaltic magma likely to rise higher. This most commonly occurs in regions of lithospheric extension, such as mid-ocean ridges and continental rift valleys.

The rise of magma into the base of the lithosphere elevates the conductive thermal gradient in the overlying rocks and causes high-temperature regional metamorphism of the lower crust. Fluids released by metamorphic reactions rise into the upper part of the lithosphere, advectively transferring mass and heat as they do so. The heat from the basaltic magma at the base of the lithosphere also causes melting of granitic fractions in regions of continental crust. These silicic melts form diapiric bodies that rise and advectively introduce heat into the middle and upper crust, bringing about contact and regional metamorphism.

The main mechanism by which the Earth has cooled throughout most of geologic time has been through creation of new, hot ocean floor and the subduction of old, cold, hydrated ocean floor. Few of these rocks are preserved in the geologic record – low-temperature, high-pressure metamorphic rocks and obducted ophiolites being some of the rare exceptions. The water released during the metamorphism of subducted ocean floor, however, has led to the development of magmas above Benioff zones at convergent plate boundaries that have, throughout geologic time, slowly added to the rocks of the continental crust.

Heat production in the Earth was greater in the early Precambrian than it is now. Convection would therefore have been more rapid, and the lithosphere would have been thinner. Despite these changes, certain rocks, such as tholeiitic basalt and granite, have formed with remarkably constant composition throughout geologic time. Other rocks, however, may owe their origins to changes in the nature of convection – komatiite in the Archean, widespread granulites in the Archean and Proterozoic, alkaline magmas in the early Proterozoic and Cretaceous, massif-type anorthosite in the Proterozoic, and blueschists and eclogites in the Phanerozoic. Some of these rocks that are prevalent at particular times may also involve changes occurring in the lower mantle or core. As our understanding of the physics of the deeper parts of the Earth increases, it may be possible to use the petrologic record to interpret a truly global history of the Earth.

23.8 PROBLEMS

23.1 If the present average mantle of the Earth contains 0.025 ppm ^{238}U, 0.0002 ppm ^{235}U, 0.103 ppm ^{232}Th, and 0.033 ppm ^{40}K, plot graphs of the heat production from each of these isotopes and their total heat production per kilogram of rock throughout the past 4.5 Ga. Compare the present heat production of this composition with that of 4.0 Ga ago. Which were the most important heat-producing isotopes during the early history of the Earth? The decay energies in mW kg^{-1} are 0.0937 for ^{238}U, 0.569 for ^{235}U, 0.0269 for ^{232}Th, and 0.0279 for ^{40}K. Use the decay constants in Table 13.1.

23.2 Using the rates of heat production from the mantle at the present and 4.0 Ga ago, as calculated in Problem 23.1 (i.e. 0.61×10^{-11} and 2.20×10^{-11} W kg^{-1}, respectively), and assuming a mantle density of 3.5 Mg m^{-3}, calculate the Rayleigh numbers at these two times for a 700 km-deep layer. The physical properties of this layer are as follows: $\alpha = 3 \times 10^{-5}$ K^{-1}, $K = 3.35$ W m^{-1} K^{-1}, $k = 10^{-6}$ m^2 s^{-1}, and $\eta = 7 \times 10^{20}$ Pa s. Would you expect the mantle to convect, and if so, how would the convection 4.0 Ga ago have compared with that of today?

23.3 If early in Earth history, magma was emplaced at spreading oceanic ridges with a temperature of 1400 °C, how long would you expect the ocean to deepen at a rate proportional to $t^{1/2}$ according to the Parsons and McKenzie (1978) model for small-scale convection in a thermal boundary layer? All conditions, other than the initial intrusion temperature, are assumed to be the same as the present day; that is, the thermal boundary layer has a density of 3.33 Mg m^{-3}, α of 3×10^{-5} °C^{-1}, k of 8×10^{-7} m^2 s^{-1}, and viscosity of 7.76×10^{19} Pa s, the temperatures at the top and bottom of the boundary layer are 975 and 1260 °C, respectively, and the temperature of the ocean floor is 0 °C. Use a critical Rayleigh number (Eq. (14.10)) of 10^3.

23.4 Given that the average radius of the Earth is 6.371×10^6 m, that of the core is 3.486×10^6 m, and the depth to the upper mantle–lower mantle boundary is 650 km, calculate the volumes of the Earth, core, lower mantle, and upper mantle. How many times greater is the volume of the lower mantle than the volume of the upper mantle?

23.5 If the upper mantle has a coefficient of thermal expansion of 4×10^{-5} °C^{-1}, a heat capacity of 10^3J kg^{-1} °C^{-1}, and a temperature of 1500 °C, what would the adiabatic temperature gradient be in °C km^{-1}?

23.6 If freely convecting mantle at a depth of 200 km has a temperature of 1400 °C, construct an adiabatic geothermal gradient to a depth of 600 km. Use the same physical properties for the mantle as in Problem 23.5. (*Hint:* Use Eq. (14.14) to express T as a function of z.)

23.7 If adiabatic decompression during lithospheric extension caused 10 wt % of the mantle to melt, how much lowering of temperature would this cause if the latent heat of fusion of the mantle were 420 kJ kg^{-1} and the heat capacity were 10^3 J kg^{-1} K^{-1}?

23.8 The conductive geotherm through a 100-km-thick lithosphere is described by the relation $T_z \approx 14z$, where z is depth in kilometers. Below this, the temperature rises adiabatically in the convecting mantle. For simplicity, the thermal boundary layer has been ignored.

(a) Calculate expressions for the temperature gradient immediately after the lithosphere has been stretched by factors of 1.5 and 2.0.

(b) The solidus of the mantle is described by the relation $T_s = 1100 + 4.4z$ where z is depth in kilometers. At what depth and temperature would the geotherm intersect the solidus if the lithosphere is stretched by factors of 1.5 and 2.0?

(c) If the geotherm is assumed to be at its maximum temperature above the solidus at the point where it switches from being conductive to convective, calculate by how much the geotherm exceeds the solidus temperature for lithospheric stretching by factors of 1.5 and 2.0. How do the degrees of melting compare in these two cases?

23.9 A 500-m-thick basaltic sill at its liquidus temperature of 1200 °C is emplaced into country rocks of granitic composition that are at a temperature of 500 °C. If during convection all of the heat lost in becoming 60% solidified goes into making an overlying molten layer of convecting granitic magma that is 50% crystalline, how thick is the layer of granite? The basaltic liquid has a latent heat of crystallization of 400 kJ kg^{-1} and a density of 2.7 Mg m^{-3}. The granitic liquid has a latent heat of crystallization of 300 kJ kg^{-1} and a density of 2.3 Mg m^{-3}. The heat capacity of all materials is 1.34 kJ kg^{-1} K^{-1}. The fraction of basalt crystallized as a function of temperature in degrees Celsius is $X_b = 7200T^{-1} - 6$ and that of the granite is $X_g = 0.65(1000 - T)/150$ (Huppert and Sparks, 1988).

23.10 If in Problem 23.9 the base of the granitic layer is kept at a temperature of 1175 °C by the convecting basaltic magma, and its upper surface is kept at the temperature at which the rock contains 65% crystals (acts as a solid), that is, 850 °C, how thick must the layer be in order to convect? The coefficient of expansion of the granitic magma is 5×10^{-5} K^{-1}, its thermal diffusivity is 10^{-6} m^2 s^{-1}, and its viscosity is 2.5×10^5 Pa s.

23.11 If, following 1% partial melting of the mantle, the melt is able to buoyantly rise through the solid matrix along the interconnecting channels along grain edges, calculate the velocity of rise if the permeability is 2.5×10^{-14} m^2 (this assumes dihedral angles for the melt phase of 50°), the viscosity of the melt is 1 Pa s, and the density contrast between the melt and solid is 600 kg m^{-3}. Would this velocity be rapid enough to supply magma at a divergent plate boundary if the rate of plate motion is taken to be on the order of 10 mm a^{-1}? (Review Section 21.3 and Eq. (21.2).)

23.12 If a dike of alnöite contains spherical xenoliths of eclogite, what must the minimum rate of intrusion have been for the xenoliths to rise from the mantle? The densities of the xenoliths and magma were 3.6 and 2.6 Mg m^{-3}, respectively, the viscosity of the magma was 10 Pa s, and the diameter of the largest xenolith is 10 cm.

23.13 Do reactions 23.4 and 23.5 account for all possible reactions that can eliminate anorthite from assemblages involving the six minerals in these two reactions if we specify that forsterite must be present? Consider Ca-Tschermak's molecule as a separate mineral.

(*Hint:* How many components do you need to account for all of the phases? A tetrahedral plot can be simplified to a ternary one by projecting from a phase common to all assemblages, that is, olivine. Then determine if there is degeneracy.)

23.14 What composition magmas will be formed by **(a)** the equilibrium partial fusion and **(b)** the fractional partial fusion of a garnet peridotite composed of 60% forsterite, 30% pyrope, and 10% diopside at 4 GPa? (Use Fig. 23.17.) The total fraction of the rock that melts in both cases is 47 wt %. **(c)** What compositions do the residual solids have in the fractional melting case? **(d)** Which mechanism requires the higher temperature?

23.15 If a source rock contains 10 ppm Eu and the bulk distribution coefficient (rock/liquid) is 0.1, what concentration of Eu would be found in melts formed by 2%, 5%, 15%, and 20% fusion of this material under complete equilibrium and complete fractional melting? (Review Eq. (12.21), (14.33), and (14.36). Be careful of what the fraction *F* refers to and how the distribution coefficient is defined.) How might the fraction of melting be determined in rocks?

Answers to selected numerical problems

Chapter 1

1.2 1.174 km **1.3** 0.557 μW m^{-3} **1.7** 68 mW m^{-2}

Chapter 2

2.1 $\bar{V}_{\mathrm{SiO_2}} = 2.611 \times 10^{-5}\mathrm{m^3 mol^{-1}}$, $\bar{V}_{\mathrm{CaO}} = 1.868 \times 10^{-5}\mathrm{m^3 mol^{-1}}$

2.3 $\rho_{\mathrm{basalt}} = 2705\mathrm{kg\,m^{-3}}$, $\rho_{\mathrm{granite}} = 2322\mathrm{kg\,m^{-3}}$

2.4 $\rho_{\mathrm{basalt}} = 2720\mathrm{kg\,m^{-3}}$, $\rho_{\mathrm{granite}} = 2340\mathrm{kg\,m^{-3}}$

2.6 $\eta_{\mathrm{basalt}} = 18\,\mathrm{Pa\,s}$, $\eta_{\mathrm{granite}} = 1.7 \times 10^5\,\mathrm{Pa\,s}$

2.9 Maximum velocity at top = 6.27 m s^{-1}

Chapter 3

3.1 600 m **3.3 (a)** 112 km **3.3 (b)** 69.5 km

3.6 (a) 47 m s^{-1} **3.6 (b)** 1.65 m **3.9** 3.726 m

3.14 0.68

Chapter 4

4.1 0.055 m s^{-1}, laminar **4.6 (a)** 27.25 **4.6 (b)** 0.004 m s^{-1}

4.6 (c) Intrusion $v = 73.8\,\mathrm{m\,s^{-1}}$, subsidence $v = 2.7\,\mathrm{m\,s^{-1}}$

4.8 (a) 17.7 m **4.8 (b)** at 60°, $z = 53$ m; at 20° $z = 11$ m

Chapter 5

5.1 Continental 30 °C km^{-1}, ocean ridge 160 °C km^{-1}

5.3 5.85 years **5.5** 390 years **5.7** 125.2 days

5.14 453 kJ mol^{-1} **5.16** 64.8 ka

Chapter 6

6.1 diapir ~27%, host ~57%

Chapter 7

7.1 1×10^4J **7.8 (a)** 1.22×10^{-8} K Pa^{-1} **7.8 (b)** 2.874×10^{-4} K m^{-1}

7.10 0.23 J mol^{-1}, near equilibrium **7.12** 79 K

Chapter 8

8.2 1895 K **8.4 (a)** 6350 J mol^{-1} **8.8 (a)** 561 K

Chapter 9

9.4 15.319 kJ mol^{-1} **9.5 (a)** $X_{\text{Or}}^{\text{Ab}} = 0.133,\ X_{\text{Ab}}^{\text{Or}} = 0.887$

9.5 (b) $X_{\text{Or}}^{\text{Ab}} = 0.255,\ X_{\text{Ab}}^{\text{Or}} = 0.745$

Chapter 10

10.1 (b) 1337 °C, 65 wt% diopside **10.2 (b)** $\gamma_{\text{Di}} = 0.713,\ \gamma_{\text{An}} = 0.54$

10.4 (c) 84% K-feldspar **10.6** 122.57 kJ mol^{-1}

10.12 (c) 55% Ne, 45% Ab **10.19** 1100 °C **10.25** 31.3%

Chapter 11

11.10 87.5% **11.13 (a)** 60% FeO

Chapter 12

12.1 $r_c = 0.25\ \mu\text{m}$ **12.5** 0.86 **12.6(c)** 34.3 K

Chapter 13

13.2 0.193 **13.5** At 2.5 Ga, 0.5094 **13.7** At 2.5 Ga, 0.7015

Chapter 14

14.5 Fo_{77} **14.8 (a)** 311.7 Pa s **14.8 (b)** 267.2 Pa s

14.11 2700 kg m^{-1}, 3.32 Pa **14.12 (b)** 19 km day^{-1}

14.13 (b) 0.436 mm day^{-1} **14.18** 1246 °C **14.20 (c)** -0.316 mm

14.24 (c) 60 wt% plagioclase

Chapter 15

15.2 Alkali basalt 3.6%, tholeiite 21.5% **15.7** 72.5

15.8 $\log_{10} a_{\text{SiO2}} = -0.634$ **15.12** 5.2 cm **15.13 (b)** 1.96 cm s^{-1}

15.14 8.16 km

Chapter 16

16.1 $V_{\text{fluid}} = 0.2\ V_{\text{rock}}$ **16.3 (a)** 418 °C

Chapter 17

17.1 (a) 10% **17.1 (b)** 25% **17.3** 31 MJ

Chapter 18

18.4 (a) Alumina index = $(Al_2O_3 - K_2O)/[(Al_2O_3 - K_2O) + FeO + MgO]$

Chapter 19

19.2 570 °C, 0.4 GPa **19.3 (a)** 565°C, 0.783 GPa **19.4 (c)** 604 °C

19.5 ~0.7 GPa

Chapter 20

20.4 (a) 0.18 GPa, $\gamma = 1.7$ **20.12 (a)** 0.133 **20.12 (b)** 0.0605

20.14 (c) 1108 cm^3

Chapter 21

21.1 (a) 116 m a^{-1} **21.2** 7.67×10^{-16} m^2 **21.8 (b)** 500 m^3 m^{-2}

21.9 7.7×10^5 m^3 m^{-2} (at 1.0 GPa) **21.10 (a)** 1.6×10^4 m^3 m^{-2}

21.13 (b) Mean K loss = −96%; −92% to − 98% (2 sigma range).

Chapter 22

22.1 (a) 3.3×10^5 years **22.2 (h)** 30 km **22.5 (a)** ~ 4×10^6 years

Chapter 23

23.3 191 Ma **23.5** 0.6°C km^{-1} **23.7** − 42°C

23.9 339 m **23.11** 0.5 m a^{-1} **23.12** 0.54 m s^{-1}

References

Abart, R. (1995). Phase equilibrium and stable isotope constraints on the formation of metasomatic garnet-vesuvianite veins (SW Adamello, N Italy). *Contributions to Mineralogy and Petrology*, **122**, 116–133.

Abbott, R. N., Jr. (1978). Peritectic reactions in the system An–Ab–Or–Qz–H_2O. *Canadian Mineralogist*, **16**, 245–256.

Abe, T., Tsukamoto, K., and Sunagawa, I. (1991). Nucleation, growth and stability of $CaAl_2Si_2O_8$ polymorphs. *Physics and Chemistry of Minerals*, **17**, 473–484.

Adams, N. K., Houghton, B. F., and Fagents, S. A. (2006). The transition from explosive to effusive eruptive regime: the example of the 1912 Novarupta eruption, Alaska. *Geological Society of America Bulletin*, **118**, 620–634.

Agee, C. B., and Walker, D. (1993). Olivine flotation in mantle melt. *Earth and Planetary Science Letters*, **114**, 315–324.

Ague, J. J. (1994a). Mass transfer during Barrovian metamorphism of pelites, south-central Connecticut: II. Channelized fluid flow and the growth of staurolite and kyanite. *American Journal of Science*, **294**, 1061–1134.

Ague, J. J. (1994b). Mass transfer during Barrovian metamorphism of pelites, south-central Connecticut: I. Evidence for changes in composition and volume. *American Journal of Science*, **294**, 989–1057.

Ague, J. J. (1995). Deep crustal growth of quartz, kyanite, and garnet into large-aperture, fluid-filled fractures, north-eastern Connecticut, USA. *Journal of Metamorphic Geology*, **13**, 299–314.

Ague, J. J. (1997). Crustal mass transfer and index mineral growth in Barrow's garnet zone, northeast Scotland. *Geology*, **25**, 73–76.

Ague, J. J. (2000). Release of CO_2 from carbonate rocks during regional metamorphism of lithologically heterogeneous crust. *Geology*, **28**, 1123–1126.

Ague, J. J. (2002). Gradients in fluid composition across metacarbonate layers of the Wepawaug Schist, Connecticut, USA. *Contributions to Mineralogy and Petrology*, **143**, 38–55.

Ague, J. J. (2003a). Fluid infiltration and transport of major, minor, and trace elements during regional metamorphism of carbonate rocks, Wepawaug Schist, Connecticut, USA. *American Journal of Science*, **303**, 753–816.

Ague, J. J. (2003b). Fluid flow in the deep crust. In *The Crust*, ed. R. L. Rudnick, vol. 3 of *Treatise on Geochemistry*, ed. H. D. Holland, and K. K. Turekian. Amsterdam: Elsevier, 195–228.

Ague, J. J. (2007). Models of permeability contrasts in subduction zone mélange: implications for gradients in fluid fluxes, Syros and Tinos Islands, Greece. *Chemical Geology*, **239**, 217–227.

Ague J. J., and Baxter, E. F. (2007). Brief thermal pulses during mountain building recorded by Sr diffusion in apatite and multicomponent diffusion in garnet. *Earth and Planetary Science Letters*, **261**, 500–516.

Ague, J. J., and Brimhall, G. H (1988). Magmatic arc asymmetry and distribution of anomalous plutonic belts in the batholiths of California: effects of assimilation, crustal thickness, and depth of crystallization. *Geological Society of America Bulletin*, **100**, 912–927.

Ague, J. J., and Rye, D. M. (1999). Simple models of CO_2 release from metacarbonates with implications for interpretation of directions and magnitudes of fluid flow in the deep crust. *Journal of Petrology*, **40**, 1443–1462.

Ague J. J., and van Haren, J. L. M. (1996). Assessing metasomatic mass and volume changes using the bootstrap, with application to deep crustal hydrothermal alteration of marble. *Economic Geology*, **91**, 1169–1182.

Ague, J. J., Park, J., and Rye, D. M. (1998). Regional metamorphic dehydration and seismic hazard. *Geophysical Research Letters*, **25**, 4221–4224.

Ague, J. J., Baxter, E. F., and Eckert, J. O., Jr. (2001). High f_{O2} during sillimanite zone metamorphism of part of the Barrovian type locality, Glen Clova, Scotland. *Journal of Petrology*, **42**, 1301–1320.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

Aki, K., and Koyanagi, R. (1981). Deep volcanic tremor and magma ascent mechanism under Kilauea, Hawaii. *Journal of Geophysical Research*, **86**, 7095–7109.

Albee, A. L. (1965). A petrogenetic grid for the Fe-Mg silicates of pelitic schists. *American Journal of Science*, **263**, 512–536.

Allègre, C. J., Hart, S. R., and Minster, J.-F. (1983). Chemical structure and evolution of the mantle and continents determined by inversion of Nd and Sr isotopic data: II. Numerical experiments and discussion. *Earth and Planetary Science Letters*, **66**, 191–213.

Allen, C. C. (1979). Volcano-ice interactions on Mars. *Journal of Geophysical Research*, **84**, 8048–8059.

Alt, J. C., and Teagle, D. A. H. (1999). The uptake of carbon during alteration of ocean crust. *Geochimica Cosmochimica Acta*, **63**, 1527–1535.

Alvarado, G. E., Denyer, P., and Sinton, C. W. (1997). The 89 Ma Tortugal komatiitic suite, Costa Rica: implications for a common geological origin of the Caribbean and East Pacific region from a mantle plume. *Geology*, **25**, 439–442.

Amstutz, G. C. (1974). *Spilites and Spilitic Rocks*. Berlin: Springer-Verlag.

Andersen, O. (1915). The system anorthite–forsterite–silica. *American Journal of Science*, **39**, 407–454.

Anderson, A. T. (1975). Some basaltic and andesitic gases. *Reviews of Geophysics and Space Physics*, **13**, 37–55.

Anderson, A. T., Jr., and Morin, M. (1968). Two types of massif anorthosites and their implications regarding the thermal history of the crust. In *Origin of Anorthosites and Related Rocks*, ed. Y. W. Isachsen. *New York State Museum and Science Service Memoir*, **18**, 57–69.

Anderson, D. L. (1984). The earth as a planet: paradigms and paradoxes. *Science*, **223**, 347–355.

Anderson, D. L. (1998). The helium paradoxes. *Proceedings of the National Academy of Sciences*, **95**, 4822–4827.

Anderson, D. L. (2005). Scoring hotspots: the plume and plate paradigms. *Geological Society of America Special Paper*, **388**, 31–54.

Anderson, E. M. (1936). The dynamics of the formation of cone-sheets, ring-dikes, and caldron-subsidences. *Royal Society of Edinburgh Proceedings*, **56**, 128–163.

Annen, C., Blundy, J. D., and Sparks, R. S. J. (2006). The genesis of intermediate and silicic magmas in deep crustal hot zones. *Journal of Petrology*, **47**, 505–539.

Anovitz, L. M., and Essene, E. J. (1987). Phase equilibria in the system $CaCO_3$-$MgCO_3$-$FeCO_3$. *Journal of Petrology*, **28**, 389–414.

Arndt, N. T., and Nisbet, E. G. (1982). What is a komatiite? In *Komatiites*, ed. N. T. Arnt and E. G. Nisbet. London: George Allen & Unwin, 19–27.

Arndt, N. T., Naldrett, A. J., and Pyke, D. R. (1977). Komatiite and iron-rich tholeiite lava of Munrow Township, northeast Ontario. *Journal of Petrology*, **18**, 319–369.

Aranovich, L. Y., and Newton, R. C. (1999). Experimental determination of CO_2–H_2O activity-composition relations at 600–1,000 °C and 6–14 kbar by reversed decarbonation and dehydration reactions. *American Mineralogist*, **84**, 1319–1332.

Asimow, P. D., and Ghiorso, M. S. (1998). Algorithm modifications extending MELTS to calculate subsolidus phase relations. *American Mineralogist*, **83**, 1127–1131.

Asimow, P. D., Hirschmann, M. M., Ghiorso, M. S., O'Hara, M. J., and Stolper, E. M. (1995). The effect of pressure-induced solid-solid phase transitions on decompression melting of the mantle. *Geochimica et Cosmochimica Acta*, **59**, 4489–4506.

Atherton, M. P. (1977). The metamorphism of the Dalradian rocks of Scotland. *Scottish Journal of Geology*, **13**, 331–370.

Austrheim, H. (1987). Eclogitization of lower crustal granulites by fluid migration through shear zones. *Earth and Planetary Science Letters*, **81**, 221–232.

Bailey, D. K. (1974). Continental rifting and alkaline magmatism. In *The Alkaline Rocks*, ed. H. Sørensen. London: John Wiley, 148–159.

Baker, B. H., and Wohlenberg, J. (1971). Structure and evolution of the Kenya Rift Valley. *Nature*, **229**, 538–542.

Baker, B. H., Crossley, R., and Goles, G. G. (1978). Tectonic and magmatic evolution of the southern part of the Kenya Rift Valley. In *Petrology and Geochemistry of Continental Rifts*, ed. E.-R. Neumann, and I. B. Ramberg. Dordrecht: D. Reidel, 29–50.

Baker, P. E. (1968). Comparative volcanology and petrology of the Atlantic island-arcs. *Bulletin of Volcanology*, **32**, 186–206.

Baldwin, S. L., Monteleone, B. D., Webb, L. E., Fitzgerald, P. G., Grove, M., and Hill, E. J. (2004). Pliocene eclogite exhumation at plate tectonic rates in eastern Papua New Guinea. *Nature*, **431**, 263–267.

Barley, M. E. (1986). Incompatible-element enrichment in Archean basalts: a consequence of contamination by older sialic crust rather than mantle heterogeneity. *Geology*, **14**, 947–950.

Barron, L. M. (1972). Thermodynamic multicomponent silicate equilibrium phase calculations. *American Mineralogist*, **57**, 809 823.

Barrow, G. (1893). On an intrusion of muscovite-biotite gneiss in the southeast Highlands of Scotland, and its accompanying metamorphism. *Quarterly Journal of the Geological Society of London*, **49**, 330–358.

Barrow, G. (1912). On the geology of the lower Dee-side and the southern Highland border. *Proceedings of the Geologist's Association*, **23**, 268–284.

Basaltic Volcanism Study Project (BVSP) (1981). *Basaltic Volcanism on the Terrestrial Planets*. New York: Pergamon Press, 1286 pp.

Bass, J. D. (1995). Elasticity of minerals, glasses, and melts. In *Global Earth Physics: A Handbook of Physical Constants*. ed. T. J. Ahrens, American Geophysical Union Reference Shelf 2, 45–63.

Baumgartner, L. P., and Ferry, J. M. (1991). A model for coupled fluid-flow and mixed-volatile mineral reactions with applications to regional metamorphism. *Contributions to Mineralogy and Petrology*, **106**, 273–285.

Baxter, E. F., and DePaolo, D. J. (2002a). Field measurement of high temperature bulk reaction rates: I. Theory and technique. *American Journal of Science*, **302**, 442–464.

Baxter, E. F., and DePaolo, D. J. (2002b). Field measurement of high temperature bulk reaction rates: II. Interpretation of results from a field site near Simplon Pass, Switzerland. *American Journal of Science*, **302**, 465–516.

Baxter, E. F., Ague, J. J., and DePaolo, D. J. (2002). Prograde temperature-time evolution in the Barrovian type-locality constrained by Sm/Nd garnet ages from Glen Clova, Scotland. *Journal of the Geological Society of London*, **159**, 71–82.

Bear, J. (1988). *Dynamics of Fluids in Porous Media*. New York: Dover.

Bear, J. and Verruijt, A. (1987). Modeling groundwater flow and pollution. Dordrecht: D. Reidel.

Beaumont, C., Jamieson, R. A., Nguyen, M. H., and Lee, B. (2001). Himalayan tectonics explained by extrusion of a low-viscosity crustal channel coupled to focused surface denudation. *Nature*, **414**, 738–742.

Bebout, G. E., and Barton, M. D. (1989). Fluid flow and metasomatism in a subduction zone hydrothermal system: Catalina Schist terrane, California. *Geology*, **17**, 976–980.

Bebout, G. E., and Barton, M. D. (2002). Tectonic and metasomatic mixing in a high-T, subduction-zone mélange–insights into the geochemical evolution of the slab-mantle interface. *Chemical Geology*, **187**, 79–106.

Becker, U., and Prieto, M. (ed.) (2006). Solid solutions: from theory to experiment. *Chemical Geology*, **225**, 173–175.

Bédard, J. H. J., Marsh, B. D., Hersum, T. G., Naslund, H. R., and Mukasa, S. B. (2007). Large-scale mechanical redistribution of orthopyroxene and plagioclase in the Basement Sill, Ferrar dolerites, McMurdo Dry Valleys, Antarctica: petrological, mineral-chemical and field evidence for channelized movement of crystals and melt. *Journal of Petrology*, **48**, 2289–2326.

Bell, T. H., Rubenach, M. J., and Fleming, P. D. (1986). Porphyroblast nucleation, growth, and dissolution in regional metamorphic as a function of deformation partitioning during foliation development. *Journal of Metamorphic Geology*, **4**, 37–67.

Ben-Jacob, E., and Garik, P. (1990). The formation of patterns in non-equilibrium growth. *Nature*, **343**, 523–530.

Bergantz, G. W., and Ni, J. (1999). A numerical study of sedimentation by dripping instabilities in viscous fluids. *International Journal of Multiphase Flow* **25**, 307 320.

Berman, R. G. (1988). Internally-consistent thermodynamic data for minerals in the system Na_2O–K_2O–CaO–MgO–FeO–Fe_2O_3–Al_2O_3–SiO_2–TiO_2–H_2O–CO_2. *Journal of Petrology*, **29**, 445–522.

Berman, R. G. (1990). Mixing properties of Ca–Mg–Fe–Mn garnets. *American Mineralogist*, **75**, 328–344.

Berman, R. G. (1991). Thermobarometry using multi-equilibrium calculations: a new technique, with petrological applications. *Canadian Mineralogist*, **29**, 833–855.

Berman, R. G. (2007). winTWQ (version 2.3): a software package for performing internally-consistent thermobarometric calculations. *Geological Survey of Canada, Open File 5462*, (edn. 2.34), 41 pp.

Berner, R. A. (1980). *Early Diagenesis: A Theoretical Approach*. Princeton, NJ: Princeton University Press.

Bickle, M. J. (1992). Transport mechanisms by fluid-flow in metamorphic rocks: oxygen and strontium decoupling in the Trois Seigneurs Massif – a consequence of kinetic dispersion? *American Journal of Science*, **292**, 289–316.

Bickle, M. J., and McKenzie, D. (1987). The transport of heat and matter by fluids during metamorphism. *Contributions to Mineralogy and Petrology*, **95**, 384–392.

Bickle, M. J., Chapman, H. J., Ferry, J. M., Rumble, D., III, Fallick, A. E. (1997). Fluid flow and diffusion in the Waterville Limestone, south-central Maine: constraints from strontium, oxygen, and carbon isotope profiles. *Journal of Petrology*, **38**, 1489–1512.

Bijwaard, H., and Spakman, W. (1999). Tomographic evidence for a narrow whole mantle plume below Iceland. *Earth and Planetary Science Letters*, **166**, 121–126.

Birch, F., Roy, R. F., and Decker, E. R. (1968). Heat flow and thermal history in New England and New York. In *Studies of Appalachian Geology, Northern and Maritime*, ed. E-an Zen, W. S. White, J. B. Hadley, and J. B. Thompson. New York: Interscience Publishers, 437–452.

Blacic, T. M., Ito, G., Canales, J. P., Detrick, R. S., and Sinton, J. (2004). Constructing the crust along the Galapagos spreading center 91.3°-95.5°: correlation of seismic layer 2A with axial magma lens and topographic characteristics. *Journal of Geophysical Research*, **109**, B10310, doi:10.1029/2004JB003066.

Blanckenburg, F. v., Villa, I. M., Baur, H., Morteani, G., and Steiger, R. H. (1989). Time calibration of a P-T-path from the western Tauern Window, eastern Alps: the problem of closure temperatures. *Contributions to Mineralogy and Petrology*, **101**, 1–11.

Boehler, R. (2000). High-pressure experiments and the phase diagram of lower mantle and core materials. *Reviews of Geophysics*, **38**, 221–245.

Bohlen, S. R. (1987). Pressure–temperature–time paths and a tectonic model for the evolution of granulites. *Journal of Geology*, **95**, 617–632.

Bohlen, S. R., and Mezger, K. (1989). Origin of granulite terranes and the formation of the lowermost continental crust. *Science*, **244**, 326–329.

Bohlen, S. R., Boettcher, A. L., and Wall, V. J. (1982). The system albite–H_2O–CO_2: a model for melting and activities of water at high pressures. *American Mineralogist*, **67**, 451–462.

Bohlen, S. R., Wall, V. J., and Boettcher, A. L. (1983). Experimental investigations and geological applications of equilibria in the system FeO–TiO_2–Al_2O_3–SiO_2–H_2O. *American Mineralogist*, **68**, 1049–1058.

Bohlen, S. R., Valley, J. W., and Essene, E. J. (1985). Metamorphism in the Adirondacks: I. Petrology, pressure, and temperature. *Journal of Petrology*, **26**, 971–992.

Bolton, E. W., Lasaga, A. C., and Rye, D. M. (1999). Long-term flow/chemistry feedback in a porous medium with heterogeneous permeability: kinetic control of dissolution and precipitation. *American Journal of Science*, **299**, 1–68.

Boschi, L., and Dziewonski, A. M. (2000). Whole Earth tomography from delay times of *P*, *PcP*, and *PKP* phases: lateral heterogeneities in the outer core or radial anisotropy in the mantle? *Journal of Geophysical Research*, **105**, 13 675–13 696.

Bottinga, Y. and Javoy, M. (1975). Oxygen isotope partitioning among minerals in igneous and metamorphic rocks. *Reviews of Geophysics and Space Physics*, **13**, 401–418.

Bottinga, Y., and Weill, D. F. (1970). Density of liquid silicate systems calculated from partial molar volumes of oxide components. *American Journal of Science*, **269**, 169–182.

Bottinga, Y., and Weill, D. F. (1972). The viscosity of magmatic silicate liquids. *American Journal of Science*, **272**, 438–475.

Bottinga, Y., Weill, D. F., and Richet, P. (1982). Density calculations for silicate liquids: I. Revised method for aluminosilicate compositions. *Geochimica et Cosmochimica Acta*, **46**, 909–919.

Boucher, D. F., and Alves, G. E. (1959). Dimensionless numbers for fluid mechanics, heat transfer, mass transfer, and chemical reaction. *Chemical Engineering Progress*, **55**, 55–64.

Boudreau, A. E. (1995). Crystal aging and the formation of fine-scale igneous layering. *Mineralogy and Petrology*, **54**, 55–69.

Boudreau, A. E. (1999). Fluid fluxing of cumulates: the J-M Reef and associated rocks of the Stillwater complex, Montana. *Journal of Petrology*, **40**, 755–772.

Boudreau, A. (2003). IRIDIUM: a program to model reaction of silicate liquid infiltrating a porous solid assemblage. *Computers and Geosciences*, **29**, 423–429.

Boudreau, A. E. (2004). PALLADIUM – A program to model the chromatographic separation of the platinum-group elements, base metals and sulfur in a solidifying igneous crystal pile. *Canadian Mineralogist*, **42**, 393–403.

Boudreau, A. E., and McBirney, A. E. (1997). The Skaergaard layered series: III. Non-dynamic layering. *Journal of Petrology*, **38**, 1003–1020.

Boudreau, A., and Philpotts, A. R. (2002). Quantitative modeling of compaction in the Holyoke flood basalt flow, Hartford Basin, Connecticut. *Contributions to Mineralogy and Petrology*, **144**, 176–184.

Bowden, P., and Turner, D. C. (1974). Peralkaline and associated ring complexes in the Nigeria–Niger province, West Africa. In *The Alkaline Rocks*, ed. H. Sørensen. Chichester, West Sussex: John Wiley, 330–351.

Bowen, N. L. (1913). The melting phenomena of the plagioclase feldspars. *American Journal of Science*, **34**, 577–599.

Bowen, N. L. (1915a). The crystallization of haplobasaltic, haplodioritic, and related magmas. *American Journal of Science*, **40**, 161–185.

Bowen, N. L. (1915b). Crystallization-differentiation in silicate liquids. *American Journal of Science*, **39**, 175–191.

Bowen, N. L. (1928). *The Evolution of the Igneous Rocks*. Princeton, NJ: Princeton University Press, 334 pp.

Bowen, N. L. (1940). Progressive metamorphism of siliceous limestone and dolomite. *Journal of Geology*, **48**, 225–274.

Bowen, N. L., and Anderson, O. (1914). The binary system MgO–SiO_2. *American Journal of Science*, **37**, 487–500.

Bowen, N. L., and Schairer, J. F. (1932). The system FeO–SiO_2. *American Journal of Science*, **24**, 177–213.

Bowen, N. L., and Schairer, J. F. (1935). The system MgO–FeO–SiO_2. *American Journal of Science*, **29**, 151–217.

Bowman, J. R., Willett, S. D., and Cook, S. J. (1994). Oxygen isotopic transport and exchange during fluid flow: one-dimensional models and applications. *American Journal of Science* **294**, 1–55.

Boyd, F. R. (1961). Welded tuffs and flows in the rhyolite plateau of Yellowstone Park, Wyoming. *Geological Society of America Bulletin*, **72**, 387–426.

Boyd, F. R., and Gurney, J. J. (1986). Diamonds and the African lithosphere. *Science*, **232**, 472–477.

Boyd, F. R., and Nixon, P. H. (1975). Origins of the ultramafic nodules from some kimberlites of northern Lesotho and the Monastery Mine, South Africa. In *Physics and Chemistry of the Earth*, vol. 9, ed. L. H. Ahrens, J. B. Dawson, A. R. Duncan, and A. J. Erlank. Oxford: Pergamon Press, 431–454.

Brady, J. B. (1983). Intergranular diffusion in metamorphic rocks. *American Journal of Science*, **283A**, 181–200.

Brady, J. B. (1988). The role of volatiles in the thermal history of metamorphic terranes. *Journal of Petrology*, **29**, 1187–1213.

Brady, J. B. (1995). Diffusion data for silicate minerals, glasses and liquids. *Mineral Physics and Crystallography: A Handbook of Physical Constants*, ed. T. J. Ahrens. American Geophysical Union Reference Shelf 2, 269–290.

Brady, J. B., and McCallister, R. H. (1982). Diffusion data for clinopyroxenes from homogenization and self-diffusion experiments. *American Mineralogist*, **68**, 95–105.

Brady, J. B., and Yund, R. A. (1983). Interdiffusion of K and Na in alkali feldspars: homogenization experiments. *American Mineralogist*, **68**, 106–111.

Brandeis, G., and Jaupart, C. (1986). On the interaction between convection and crystallization in cooling magma chambers. *Earth and Planetary Science Letters*, **77**, 345–361.

Brandeis, G., and Jaupart, C. (1987). Crystal sizes in intrusions of different dimensions: constraints on the cooling regime and the crystallization kinetics. In *Magmatic Processes: Physicochemical Principles*, ed. B. O. Mysen. *Geochemical Society Special Publication*, **1**, 307–318.

Brandeis, G., Jaupart, C., and Allègre, C. J. (1984). Nucleation, crystal growth and the thermal regime of cooling magmas. *Journal of Geophysical Research*, **89**, 10161–10177.

Braun, M. G., and Kelemen, P. B. (2002). Dunite distribution in the Oman ophiolite: implications for melt flux through porous dunite conduits. *Geochemistry Geophysics Geosystems*, **3**(11), 8603, doi:10.1029/2001GC000289.

Breeding, C. M., and Ague, J. J. (2002). Slab-derived fluids and quartz-vein formation in an accretionary prism, Otago Schist, New Zealand. *Geology*, **30**, 499–502.

Breeding, C. M., Ague, J. J., and Bröcker, M. (2004). Fluid-metasedimentary rock interactions and the chemical composition of arc magmas. *Geology*, **32**, 1041–1044.

Brey, G. P., and Köhler, T. (1990). Geothermobarometry in four-phase lherzolites: II. New thermobarometers, and practical assessment of existing thermobarometers. *Journal of Petrology*, **31**, 1353–1378.

Bridgwater, D., and Coe, K. (1970). The role of stoping in the emplacement of the giant dikes of Isortoq, South Greenland. In *Mechanism of Igneous Intrusion*, ed. G. Newall, and H. Rast. Liverpool, Lancashire: Liverpool Geological Society, *Geological Journal Special Issue*, **2**, 67–78.

Brimhall, G. H, Jr. (1977). Early fracture-controlled disseminated mineralization at Butte, Montana. *Economic Geology*, **72**, 37–59.

Brimhall, G. H, Jr. (1979). Lithologic determination of mass transfer mechanisms of multiple-stage porphyry copper mineralization at Butte, Montana: Vein formation by hypogene leaching and enrichment of potassium-silicate protore. *Economic Geology*, **74**, 556–589.

Brimhall, G. H, Lewis, C. J., Ague, J. J., *et al*. (1988). Metal enrichment in bauxites by deposition of chemically mature aeolian dust. *Nature*, **333**, 819–824.

Bröcker, M., and Enders, M. (2001). Unusual bulk-rock compositions in eclogite-facies rocks from Syros and Tinos (Cyclades, Greece): implications for U-Pb zircon geochronology. *Chemical Geology*, **175**, 581–603.

Brøgger, W. G. (1921). Die Eruptivegestein des Kristianiagebietes, IV. Das Fengebiet in Telemark, Norvegen. *Vid ensk. Skr. I Mat.-Naturv. Klasse*, **1920**(9), 150–167.

Brophy, J. G., Whittington, C. S., and Park, Y.-R. (1999). Sector-zoned augite megacrysts in Aleutian high alumina basalts: implications for the conditions of basalt crystallization and the generation of calc-alkaline series magmas. *Contributions to Mineralogy and Petrology*, **135**, 277–290.

Brown, G. C., and Mussett, A. E. (1981). *The Inaccessible Earth*. London: George Allen and Unwin.

Brown, J. M., and McQueen, R. G. (1986). Phase transitions, Grüneisen parameter, and elasticity for shocked iron between 77 GPa and 400 GPa. *Journal of Geophysical Research*, **91**, 7485–7494.

Brown, M. (2006). Duality of thermal regimes is the distinctive characteristic of plate tectonics since the Neoarchean. *Geology*, **34**, 961–964.

Brown, M. A., Brown, M., Carlson, W. D., and Denison, C. (1999). Topology of syntectonic melt-flow networks in the deep crust: inferences from three-dimensional images of leucosome geometry in migmatites. *American Mineralogist*, **84**, 1793–1818.

Brunelli, D., Seyler, M., Cipriani, A., Ottolini, L., and Bonatti, E. (2006). Discontinuous melt extraction and weak refertilization of mantle peridotites at the Vema lithospheric section (mid-Atlantic ridge). *Journal of Petrology*, **47**, 745–771.

Bryan, W. B. (1972). Morphology of quench crystals in submarine basalts. *Journal of Geophysical Research*, **77**, 5812–5819.

Buck, W. R., Carbotte, S. M., and Mutter, C. (1997). Controls on extrusion at mid-ocean ridges. *Geology*, **25**, 935–938.

Buddington, A. F. (1959). Granite emplacement with special reference to North America. *Geological Society of America Bulletin*, **70**, 671–748.

Buddington, A. F., and Lindsley, D. H. (1964). Iron-titanium oxide minerals and synthetic equivalents. *Journal of Petrology*, **5**, 310–357.

Bunch, T. E., Dence, M. R., and Cohen, A. J. (1967). Natural terrestrial maskelynite. *American Mineralogist*, **52**, 244–253.

Bundy, F. P., Bovenkerk, H. P., Strong, H. M., and Wentorf, R. H., Jr. (1961). Diamond graphite equilibrium line from growth and graphitization of diamond. *Journal of Chemical Physics*, **35**, 383.

Bunsen, R. (1851). Über die Processe der vulkanischen Gesteinsbildung Islands. *Annalen der Physik und Chemie*, **83** (6), 197–272.

Burg, J. P., and Vigneresse, J. L. (2002). Non-linear feedback loops in the rheology of cooling-crystallizing felsic magma and heating-melting felsic rock. In *Deformation Mechanisms, Rheology and Tectonics: Current Status and Future Perspectives*, ed. S. de Meer, M. R. Drury, J. H. P. de Bresser, and G. M. Pennock. London: Geological Society, Special Publications, **200**, 275–292.

Burlini, L., Vinciguerra, S., Di Toro, G., *et al*. (2007). Seismicity preceding volcanic eruptions: new experimental insights. *Geology*, **35**, 183–186.

Burnham, C. W. (1979). The importance of volatile constituents. In *The Evolution of the Igneous Rocks: Fiftieth Anniversary*

Perspectives, ed. H. S. Yoder, Jr. Princeton, NJ: Princeton University Press, 439–482.

Burnham, C. W., and Davis, N. F. (1974). The role of H_2O in silicate melts, II: thermodynamic and phase relations in the system $NaAlSiO_3O_8$–H_2O to 10 kilobars, 700 to 1100 °C. *American Journal of Science*, **274**, 902–940.

Burnham, C. W., Holloway, J. R., and Davis, N. F. (1969). Thermodynamic properties of water to 1000 °C and 10,000 bars. *Geological Society of America Special Paper*, **32**, 1–96.

Calas, G., Henderson, G. S., and Stebbins, J. F. (2006). Glasses and melts: linking geochemistry and materials science. *Elements*, **2**, 265–268.

Camacho, A., Lee, J. K. W., Hensen, B. J., and Braun, J. (2005). Short-lived orogenic cycles and the eclogitization of cold crust by spasmodic hot fluids. *Nature*, **435**, 1191–1196.

Cameron, E. N., Jahns, R. H., McNair, A. H., and Page, L. R. (1949). The internal structure of granitic pegmatites. *Economic Geology Monograph*, **2**, 115 pp.

Campbell, I. H. (1985). The difference between oceanic and continental tholeiites: a fluid dynamic explanation. *Contributions to Mineralogy and Petrology*, **91**, 37–43.

Campbell, I. H. (2007). Testing the plume theory. *Chemical Geology*, **241**, 153–176.

Campbell, I. H., and Turner, J. S. (1986). The influence of viscosity on fountains in magma chambers. *Journal of Petrology*, **27**, 1–30.

Candela, P. A. (2003). Ores in the Earth's crust. In *The Crust*, ed. R. L. Rudnick, vol. 3 of *Treatise on Geochemistry*, ed. H. D. Holland, and K. K. Turekian. Amsterdam: Elsevier, 411–431.

Canup, R. M., and Asphaug, E. (2001). Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature*, **412**, 708–712.

Carena, S., Suppe, J., and Kao, H. (2002). Active detachment of Taiwan illuminated by small earthquakes and its control of first-order topography. *Geology*, **30**, 935–938.

Carlson, W. D. (2002). Scales of disequilibrium and rates of equilibration during metamorphism. *American Mineralogist*, **87**, 185–204.

Carlson, W. D. (2006). Rates of Fe, Mg, Mn, and Ca diffusion in garnet. *American Mineralogist*, **91**, 1–11.

Carman, M. F., Jr., Cameron, M., Gunn, B., Cameron, K. L., and Butler, J. C. (1975). Petrology of Rattlesnake Mountain Sill, Big Bend National Park, Texas. *Geological Society of America Bulletin*, **86**, 177–193.

Carmichael, D. M. (1969). On the mechanism of prograde metamorphic reactions in quartz-bearing pelitic rocks. *Contributions to Mineralogy and Petrology*, **20**, 244–267.

Carmichael, D. M. (1978). Metamorphic bathozones and bathograds: a measure of post-metamorphic uplift and erosion on a regional scale. *American Journal of Science*, **278**, 769–797.

Carmichael, I. S. E., and Ghiorso, M. S. (1986). Oxidation-reduction relations in basic magma: a case for homogeneous equilibria. *Earth and Planetary Science Letters*, **78**, 200–210.

Carmichael, I. S. E., Nicholls, J., and Smith, A. L. (1970). Silica activity in igneous rocks. *American Mineralogist*, **55**, 246–263.

Caro, G., Bourdon, B., Birck, J.-L., and Moorbath, S. (2003). ^{146}Sm–^{142}Nd evidence from Isua metamorphosed sediments for early differentiation of the Earth's mantle. *Nature*, **423**, 428–432.

Carslaw, H. S., and Jaeger, J. C. (1959). *Conduction of Heat in Solids*, 2nd edn. Oxford: Oxford University Press, 510 pp.

Carson, C. J., Powell, R., and Clarke, G. L. (1999). Calculated mineral equilibria for eclogites in CaO-Na_2O-FeO-MgO-Al_2O_3-SiO_2-H_2O: application to the Pouébo terrane, Pam Peninsula, New Caledonia. *Journal of Metamorphic Geology* **17**, 9 24.

Carter, J. L. (1970). Mineralogy and chemistry of the Earth's upper mantle based on the partial fusion – partial crystallization model. *Geological Society of America Bulletin*, **81**, 2021–2034.

Carter, S. R., Evenson, N. M., Hamilton, P. J., and O'Nions, R. K. (1978). Neodymium and strontium isotope evidence for crustal contamination of continental volcanics. *Science*, **202**, 743–747.

Cashman, K. V. (1990). Textural constraints on the kinetics of crystallization of igneous rocks. In *Modern Methods of Igneous Petrology*, ed. J. Nicholls, and J. K. Russell, vol. 24 of *Reviews in Mineralogy*. Washington, DC: Mineralogical Society of America, 259–314.

Cashman, K. V., and Ferry, J. M. (1988). Crystal size distribution (CSD) in rocks and the kinetics and dynamics of crystallization: III. Metamorphic crystallization. *Contributions to Mineralogy and Petrology*, **99**, 401–415.

Cashman, K. V., and Marsh, B. D. (1988). Crystal size distribution (CSD) in rocks and the kinetics and dynamics of crystallization: II. Makaopuhi lava lake. *Contributions to Mineralogy and Petrology*, **99**, 292–305.

Castellan, G. W. (1983). *Physical Chemistry*, 3rd edn. Reading, MA: Addison-Wesley.

Catlos, E. J., and Sorensen, S. S. (2003). Phengite-based chronology of K- and Ba-rich fluid flow in two paleosubduction zones. *Science*, **299**, 92–95.

Cawood, P. A., Kroner, A., and Pisarevsky, S. (2006). Precambrian plate tectonics: criteria and evidence. *GSA Today*, **16**(7), 4–11.

Cawthorn, R. G. (ed.) (1996). *Layered Intrusions*. Amsterdam: Elsevier, 542 pp.

Chadwick, W. W., Jr., Geist, D. J., Jónsson, S., *et al*. (2006). A volcano bursting at the seams: inflation, faulting, and eruption at Sierra Negra volcano, Gálapagos. *Geology*, **34**, 1025–1028.

Chamberlain, C. P., and Rumble, D. (1988). Thermal anomalies in a regional metamorphic terrane: an isotopic study of the role of fluids. *Journal of Petrology*, **29**, 1215–1232.

Chamberlain, C. P., and Rumble, D. (1989). The influence of fluids on the thermal history of a metamorphic terrain: New Hampshire, USA. In *Evolution of Metamorphic Belts*, ed. J. S. Daly, R. A. Cliff, and B. W. D. Yardley. *Geological Society Special Publication*, **43**, 203–213.

Chapin, C. E., and Elston, W. E. (eds.) (1979). *Ash-flow Tuffs. Geological Society of America Special Paper*, **180**, 211 pp.

Chapman, C. A. (1962). Diabase-granite composite dikes, with pillow-like structure, Mount Desert Island, Maine. *Journal of Geology*, **70**, 539–564.

Chappell, B. W., and White, A. J. R. (1992). I- and S-type granites in the Lachlan fold belt. *Transactions of the Royal Society of Edinburgh, Earth Science*, **83**, 1–26.

Cheadle, M. J., Elliott, M. T., and McKenzie D. (2004). Percolation threshold and permeability of crystallizing igneous rocks: the importance of textural equilibrium. *Geology*, **32**, 757–760.

Cherniak, D. J. (2002). Ba diffusion in feldspar. *Geochimica et Cosmochimica Acta*, **66**, 1641–1650.

Cherniak, D. J., and Ryerson, F. J. (1993). A study of strontium diffusion in apatite using Rutherford backscattering spectroscopy and ion implantation. *Geochimica et Cosmochimica Acta*, **57**, 4653–4662.

Cherniak, D. J., Watson, E. B., and Wark, D. A. (2007). Ti diffusion in quartz. *Chemical Geology*, **236**, 65–74.

Chinner, G. A. (1961). The origin of sillimanite in Glen Clova, Angus. *Journal of Petrology* **2**, 312 323.

Chopin, C. (1984). Coesite and pure pyrope in high-grade blueschists of the Western Alps: a first record and some consequences. *Contributions to Mineralogy and Petrology*, **86**, 107–118.

Chopin, C. (2003). Ultrahigh-pressure metamorphism: tracing continental crust into the mantle. *Earth and Planetary Science Letters*, **212**, 1–14.

Christensen, J. N., Rosenfeld, J. L., and DePaolo, D. J. (1989). Rates of tectonometamorphic processes from rubidium and strontium isotopes in garnet. *Science*, **244**, 1465–1469.

Christiansen, R. L. (1979). Cooling units and composite sheets in relation to caldera structure. In *Ash-flow Tuffs*, ed. C. E. Chapin and W. E. Elston. *Geological Society of America Special Paper*, **180**, 29–42.

Christiansen, R. L., and Lipman, P. W. (1966). Emplacement and thermal history of a rhyolite lava flow near Fortymile Canyon, southern Nevada. *Geological Society of America Bulletin*, **77**, 671–684.

Clayton, R. N. (1977). Genetic relations among meteorites and planets. In *Comets, Asteroids, Meteorites: Interrelations, Evolution, and Origins*, ed. A. H. Delsemme. Toledo: University of Toledo Press, 545–550.

Clemens, J. D., and Wall, V. J. (1981). Origin and crystallization of some peraluminous (S-type) granitic magmas. *Canadian Mineralogist*, **19**, 111–131.

Clough, C. T., Maufe, H. B., and Bailey, E. B. (1909). The cauldron-subsidence of Glen-Coe, and the associated igneous phenomena. *Quarterly Journal of the Geological Society of London*, **65**, 611–678.

Coe, K. (1966). Intrusive tuff of West Cork, Ireland. *Quarterly Journal of the Geological Society of London*, **122**, 1–28.

Coffin, M. F., and Eldholm, O. (1994). Large igneous provinces: structure; crustal structure, dimensions and external consequences. *Reviews of Geophysics*, **32**, 1–36.

Coggon, R., and Holland, T. J. B. (2002). Mixing properties of phengitic micas and revised garnet-phengite thermobarometers. *Journal of Metamorphic Geology*, **20**, 683–696.

Coleman, R. G. (1977). *Ophiolites*. New York: Springer-Verlag, 229 pp.

Condie, K. C. (2005). *Earth as an Evolving Planetary System*. Boston: Elsevier Academic Press, 447 pp.

Connolly, J. A. D. (1997). Devolatilization-generated fluid pressure and deformation-propagated fluid flow during prograde regional metamorphism. *Journal of Geophysical Research*, **102**, 18149–18173.

Connolly, J. A. D., and Petrini, K. (2002). An automated strategy for calculation of phase diagram sections and retrieval of rock properties as a function of physical conditions. *Journal of Metamorphic Geology*, **20**, 697–708.

Conrad, W. K., and Kay, R. W. (1984). Ultramafic and mafic inclusions from Adak Island: crystallization history, and implications for the nature of primary magmas and crustal evolution in the Aleutian arc. *Journal of Petrology*, **25**, 88–125.

Cooper, R. F., and Kohlstedt, D. L. (1984). Solution-precipitation enhanced diffusional creep of partially molten olivine-basalt aggregates during hot-pressing. *Tectonophysics*, **107**, 207–233.

Cox, K. G. (1978). Kimberlite pipes. *Scientific American*, **238**, 120–130.

Cox, K. G. (1980). A model for flood basalt vulcanism. *Journal of Petrology*, **21**, 629–650.

Crank, J. (1975). *The Mathematics of Diffusion*, 2nd Edn. Oxford: Oxford University Press, 414 pp.

Crank, J., and Nicolson, P. (1947). A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Proceedings of the Cambridge Philosophical Society*, **43**, 50–67.

Crawford, M. L., and Hollister, L. S. (1986). Metamorphic fluids, the evidence from fluid inclusions. In *Fluid-rock Interactions During Metamorphism*, ed. J. V. Walther, and B. J. Wood. New York: Springer, 1–35.

Crough, S. T., and Jurdy, D. M. (1980). Subducted lithosphere, hot-spots, and the geoid. *Earth and Planetary Science Letters*, 48, 15–22.

Cruden, A. R. (1990). Flow and fabric development during the diapiric rise of magma. *Journal of Geology*, **98**, 681–698.

Curie, P. (1885). Sur la formation des cristaux et sur les constantes capillaires de leurs différentes faces. *Société Minéralogique de France Bulletin*, **8**, 145–150.

Cui, X., Nabelek, P. I., and Liu, M. (2001). Heat and fluid flow in contact metamorphic aureoles with layered and transient permeability, with application to the Notch Peak aureole, Utah. *Journal of Geophysical Research*, **106**, 6477–6492.

Dalrymple, G. B. (2004). *Ancient Earth, Ancient Skies*. Stanford, CA: Stanford University Press.

Dalrymple, G. B., Gromme, C. S., and White, R. W. (1975). Potassium-argon age and paleomagnetism of diabase dikes in Liberia: initiation of central Atlantic rifting. *Geological Society of America Bulletin*, **86**, 399–411.

Daly, R. A. (1933). *Igneous Rocks and the Depths of the Earth*. New York: Hafner, 598 pp. (reprinted, 1968).

Danckwerth, P. A., and Newton, R. C. (1978). Experimental determination of the spinel peridotite to garnet peridotite reaction in the system MgO–Al_2O_3–SiO_2 in the range 900°–1100 °C and Al_2O_3 isopleths of enstatite in the spinel field. *Contributions to Mineralogy and Petrology*, **66**, 189–201.

Danes, Z. F. (1972). Dynamics of lava flows. *Journal of Geophysical Research*, **77**, 1430–1432.

Darcy, H. P. G. (1856). *Les Fontaines Publiques de la Ville de Dijon*. Paris: Victor Dalmont, 647 pp.

Darken, L. S., and Gurry, R. W. (1945). The system iron – oxygen: I. The wüstite field and related equilibria. *Journal of the American Chemical Society*, **67**, 1398–1412.

Dasgupta, R., Hirschmann, M. M., and Stalker, K. (2006). Immiscible transition from carbonate-rich to silicate-rich melts in the 3 GPa melting interval of eclogite + CO_2 and genesis of silica-undersaturated ocean island lavas. *Journal of Petrology*, **47**, 647–671.

Davidson, J. P., Morgan, D. J., Charlier, B. L. A., Harlou, R., and Hora, J. M. (2007). Microsampling and isotopic analysis of igneous rocks: implications for the study of magmatic systems. *Annual Review of Earth and Planetary Sciences*, **35**, 273–311.

Davies, G. F. (1999). *Dynamic Earth Plates, Plumes and Mantle Convection*. Cambridge: Cambridge University Press, 458 pp.

Davis, A. M. (2003). Introduction to volume 1. In *Meteorites, Comets, and Planets*, ed. A. M. Davis, vol. 1 of *Treatise on Geochemistry*, ed. H. D. Holland, and K. K. Turekian. Amsterdam: Elsevier, xv–xvii.

Davis, B. T. C., and Schairer, J. F. (1965). Melting relations in the join diopside–forsterite–pyrope at 40 kilobars and at one atmosphere. *Carnegie Institution Washington Yearbook*, **64**, 123–126.

Davis, E., and Elderfield, H. (eds.) (2004). *Hydrogeology of the Oceanic Lithosphere*. Cambridge: Cambridge University Press, 640 pp.

Dawson, J. B. (1962). The geology of Oldoinyo Lengai. *Bulletin of Volcanology*, **24**, 349–387.

Dawson, J. B., and Hawthorne, J. B. (1973). Magmatic sedimentation and carbonatite differentiation in kimberlite sills at Benfontein, South Africa. *Journal of the Geological Society London*, **129**, 61–85.

de Boer, J. Z., and Sanders, D. S. (2004). *Volcanoes in Human History*. Princeton, NJ: Princeton University Press, 320 pp.

DeGraff, J. M., and Aydin, A. (1987). Surface morphology of columnar joints and its significance to mechanics and direction of joint growth. *Geological Society of America Bulletin*, **99**, 605–617.

Delaney, P. T. and Pollard, D. D. (1981). Deformation of host rocks and flow of magma during growth of minette dykes and breccia-bearing intrusions near Ship Rock, New Mexico. *U.S. Geological Survey Professional Paper*, **1202**, 61 pp.

Delaney, P. T., and Pollard, D. D. (1982). Solidification of basaltic magma during flow in a dike. *American Journal of Science*, **282**, 856–885.

Demouchy, S., Jacobsen, S. D., Gaillard, F., and Stern, C. R. (2006). Rapid magma ascent recorded by water diffusion profiles in mantle olivine. *Geology*, **34**, 429–432.

Denbigh, K. (1957). *The Principles of Chemical Equilibrium*. Cambridge: Cambridge University Press, 491 pp.

Dence, M. R. (1971). Impact melts. *Journal of Geophysical Research*, **76**, 5552–5565.

DePaolo, D. J. (1981a). Nd isotopic studies: some new perspectives on Earth structure and evolution. *EOS*, **62**, 137–140.

DePaolo, D. J. (1981b). Trace element and isotopic effects of combined wallrock assimilation and fractional crystallization. *Earth and Planetary Science Letters*, **53**, 189–202.

DePaolo, D. J. (1985). Isotopic studies of processes in mafic magma chambers, I: the Kiglapait Intrusion, Labrador. *Journal of Petrology*, **26**, 925–951.

DePaolo, D. J., and Johnson, R. W. (1979). Magma genesis in the New Britain island-arc: constraints from Nd and Sr isotopes and trace-element patterns. *Contributions to Mineralogy and Petrology*, **70**, 367–379.

DePaolo, D. J., and Wasserburg, G. J. (1979). Petrogenetic mixing models in Nd-Sr isotopic patterns. *Geochimica et Cosmochimica Acta*, **43**, 615–627.

Dickin, A. P. (1995). *Radiogenic Isotope Geology*. Cambridge: Cambridge University Press, 490 pp.

Dickson, L. D. (2006). Detailed textural analysis of the Palisades Sill, New Jersey. Unpublished Ph.D. thesis, University of Connecticut, 181 pp.

Diener, J. F. A., Powell, R., White, R. W., and Holland, T. J. B. (2007). A new thermodynamic model for clino- and orthoamphiboles in the system Na_2O-CaO-FeO-MgO-Al_2O_3-SiO_2-H_2O-O. *Journal of Metamorphic Geology*, **25**, 631–656.

Dietz, R. S. (1961). Vredefort ring structure: meteorite impact scar. *Journal of Geology*, **69**, 499–516.

Dietz, R. S. (1964). Sudbury structure as an astrobleme. *Journal of Geology*, **72**, 412–434.

Dingwell, D. B. (1995). Viscosity and anelasticity of melts. In *Mineral Physics and Crystallography: A Handbook of Physical Constants*, ed. T. J. Ahrens. American Geophysical Union Reference Shelf 2, 209–217.

Dingwell, D. B. (2006). Transport properties of magmas: diffusion and rheology. *Elements*, **2**, 281–286.

Dingwell, D. B., Hess, K.-U., Knoche, R. (1996). Granite and granitic pegmatite melts: volumes and viscosities. *Transactions of the Royal Society of Edinburgh: Earth Science* **87**, 65 72.

Dipple, G. M., and Ferry, J. M. (1992a). Metasomatism and fluid flow in ductile fault zones. *Contributions to Mineralogy and Petrology*, **112**, 149–164.

Dipple, G. M., and Ferry, J. M. (1992b). Fluid flow and stable isotopic alteration in rocks at elevated temperatures with applications to metamorphism. *Geochimica et Cosmochimica Acta*, **56**, 3539–3550.

Dixon, J. E. (1997). Degassing of alkalic basalt. *American Mineralogist*, **82**, 368–378.

Dixon, J. E., and Ridley, J. R. (1987). Syros. In *Chemical Transport in Metasomatic Processes*, ed. H. C. Helgeson. Dordrecht: D. Reidel, NATO ASI series and Berlin: Springer-Verlag, 489–501.

Dixon, J. E., Stolper, E. M., and Holloway, J. R. (1995). An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids: I. Calibration and solubility models. *Journal of Petrology*, **36**, 1607–1631.

Dixon, J. M. (1975). Finite strain and progressive deformation in models of diapiric structures. *Tectonophysics*, **28**, 89–124.

Dixon, S., and Rutherford, M. J. (1979). Plagiogranites as late stage immiscible liquids in ophiolite and mid-ocean ridge suites; an experimental study. *Earth and Planetary Science Letters*, **45**, 45–60.

Dobran, F., Neri, A., and Todesco, M. (1994). Assessing the pyroclastic flow hazard at Vesuvius. *Nature*, **367**, 551–554.

Dohmen, R., and Chakraborty, S. (2003). Mechanism and kinetics of element and isotopic exchange mediated by a fluid phase. *American Mineralogist*, **88**, 1251–1270.

Dowty, E. (1980). Crystal growth and nucleation theory and the numerical simulation of igneous crystallization. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 420–485.

Doyle, C. D. (1987). The relationship between activities of divalent cation oxides and the solution of sulfide in silicate and aluminosilicate liquids. In *Physical Chemistry of Magma*, ed. I. Kushiro and I. L. Perchuk. Advances in Physical Geochemistry Series, vol. 7, Chap. 9. New York: Springer-Verlag.

Duchesne, J.-C. (1984). Massif anorthosites: another partisan review. In *Feldspars and feldspathoids*, ed. W. L. Brown. Dordrecht: D. Reidel, 411–433.

Duffield, W. A. (1972). A naturally occurring model of global plate tectonics. *Journal of Geophysical Research*, **73**, 619–634.

Dziewonski, A. M. (1984). Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6. *Journal of Geophysical Research*, **89**, 5929–5952.

Eby, G. N. (1980). Minor and trace element partitioning between immiscible ocelli-matrix pairs from lamprophyric dikes and sills, Monteregian hills petrographic province, Quebec. *Contributions to Mineralogy and Petrology*, **75**, 269–278.

Eby, G. N. (1984). Geochronology of the Monteregian Hills alkaline igneous province, Quebec. *Geology*, **12**, 468–470.

Eckert, J. O., Newton, R. C., and Kleppa, O. J. (1991). The *H* of reaction and recalibration of garnet-pyroxene-plagioclase-quartz geobarometers in the CMAS system by solution calorimetry. *American Mineralogist*, **76**, 148–160.

Eggler, D. H. (1973). Role of CO_2 in melting processes in the mantle. *Carnegie Institution of Washington Yearbook*, **72**, 457–467.

Eggler, D. H. (1974). Effect of CO_2 on the melting of peridotite. *Carnegie Institution of Washington Yearbook*, **73**, 215–224.

Eggler, D. H. (1976). Composition of the partial melt of carbonated peridotite in the system CaO–MgO–SiO_2–CO_2. *Carnegie Institution of Washington Yearbook* **75**, 623 626.

Eggler, D. H., and Burnham, C. W. (1973). Crystallization and fractionation trends in the system andesite-H_2O-CO_2-O_2 at pressures to 10 Kb. *Geological Society of America Bulletin*, **84**, 2517–2532.

Eichelberger, J. C. (1975). Origin of andesite and dacite; evidence of mixing at Glass Mountain in California and at other circum-Pacific volcanoes. *Geological Society of America Bulletin*, **86**, 1381–1391.

Eiler, J. M., Valley, J. W., and Baumgartner, L. P. (1993). A new look at stable isotope thermometry. *Geochimica et Cosmochimica Acta*, **57**, 2571–2583.

Elliott, M. T., Cheadle M. J., and Jerram, D. A. (1997). On the identification of textural equilibrium in rocks using dihedral angle measurements. *Geology*, **25**, 355–358.

Ellis, D. J., and Green, D. H. (1979). An experimental study of the effect of Ca upon garnet-clinopyroxene Fe–Mg exchange equilibria. *Contributions to Mineralogy and Petrology*, **71**, 13–22.

Elo, S., and Korja, A. (1993). Geophysical interpretation of the crustal and upper mantle structure in the Wiborg rapakivi granite area, southeastern Finland. *Precambrian Research*, **64**, 273–288.

Emeleus, C. H. (1987). The Rhum layered complex, Inner Hebrides, Scotland. In *Origins of Igneous Layering*, ed. I. Parsons. Dordrecht: D. Reidel, 263–286.

Emeleus, C. H. (1997). Geology of Rum and the adjacent islands. *British Geological Survey Sheet Memoir*, **60** (Scotland), 170 pp.

Emeleus, C. H., Cheadle, M. J., Hunter, R. H., Upton, B. G. J., and Wadsworth, W. J. (1996). The Rum layered suite. In *Layered Intrusions*, ed. R. G. Cawthorn. Amsterdam: Elsevier, 403–440.

Emslie, R. F. (1985). Proterozoic anorthosite massifs. In *The Deep Proterozoic Crust in the North Atlantic Provinces*, ed. A. C. Tobi, and J. L. R. Touret. Dordrecht: D. Reidel, 39–60.

Engel, A. E. J., Itson, S. P., Engel, C. G., Stickney, D. M., and Cray, E. J., Jr. (1974). Crustal evolution and global tectonics: a petrogenetic view. *Geological Society of America Bulletin*, **85**, 843–858.

England, P. C., and Molnar, P. (1990). Surface uplift, uplift of rocks, and exhumation of rocks. *Geology*, **18**, 1173–1177.

England, P. C., and Richardson, S. W. (1977). The influence of erosion upon the mineral facies of rocks from different metamorphic environments. *Journal of the Geological Society of London*, **134**, 201–213.

England, P. C., and Thompson, A. B. (1984). Pressure-temperature-time paths of regional metamorphism: I. Heat transfer during the evolution of regions of thickened continental crust. *Journal of Petrology*, **25**, 894–928.

England, P. C., Molnar, P., and Richter, F. (2007). John Perry's neglected critique of Kelvin's age for the Earth: a missed opportunity in geodynamics. *GSA Today*, **17**, No. 1, 4–9.

Epp, D. (1984). Possible perturbations to hotspot traces and implications for the origin and structure of the Line Islands. *Journal of Geophysical Research*, **89**, 11273–11286.

Ernst, R. E., and Buchan, K. L. (2003). Recognizing mantle plumes in the geological record. *Annual Review of Earth and Planetary Sciences*, **31**, 469–523.

Ernst, R. E., Grosfils, E. B., and Mège, D. (2001). Giant dike swarms: Earth, Venus, and Mars. *Annual Review of Earth and Planetary Sciences*, **29**, 489–534.

Ernst, W. G. (1976). *Petrologic Phase Equilibria*. San Francisco: W. H. Freeman, 333 pp.

Ernst, W. G. (1988). Tectonic history of subduction zones inferred from retrograde blueschist P–T paths. *Geology*, **16**, 1081–1084.

Ernst, W. G., and Banno, S. (1991). Neoblastic jadeitic pyroxene in Franciscan metagraywackes from Pacheco Pass, central Diablo Range, California, and implications for the inferred metamorphic P–T trajectory. *New Zealand Journal of Geology and Geophysics*, **34**, 285–292.

Eskola, P. (1920). The mineral facies of rocks. *Norsk Geolologisk Tidsskrift*, **6**, 143–194.

Essene, E. J. (1982). Geologic thermometry and barometry. In *Characterization of Metamorphism through Mineral Equilibria*, ed. J. M. Ferry, vol. 10 of *Reviews in Mineralogy*. Washington, DC: Mineralogical Society of America, 153–206.

Etheridge, M. A., Wall, V. J., Cox, S. F., and Vernon, R. H. (1984). High fluid pressures during regional metamorphism and deformation: implications for mass transport and deformation mechanisms. *Journal of Geophysical Research*, **89**, 4344–4358.

Evans, B. W. (2007a). Metamorphism of limestone and the petrogenetic grid. In *Landmark Papers: Metamorphic Petrology*, ed. B. W. Evans, and B. J. Wood, (executive ed.). Twickenham, Middlesex: Mineralogical Society of Great Britain and Ireland, L63–L66.

Evans, B. W. (2007b). Thermal models of collision belts. In *Landmark Papers: Metamorphic Petrology*, ed. B. Evans, and B. J. Wood. Twickenham, Middlesex: Mineralogical Society of Great Britain and Ireland, L199–L203.

Evans, B. W., and Davidson, G. F. (1999). Kinetic control of metamorphic imprint during synplutonic loading of batholiths: an example from Mount Stuart, Washington. *Geology*, **27**, 415–418.

Evans, B. W., and Trommsdorff, V. (1974). Stability of enstatite + talc, CO_2-metasomatism of metaperidotite, Val d'Efra, Lepontine Alps. *American Journal of Science*, **274**, 274–296.

Evans, K. A., and Bickle, M. J. (1999). Determination of time-integrated metamorphic fluid fluxes from the reaction progress of multivariant assemblages. *Contributions to Mineralogy and Petrology*, **134**, 277–293.

Ewart, A. (1976). Mineralogy and chemistry of modern orogenic lavas – some statistics and implications. *Earth and Planetary Science Letters*, **31**, 417–432.

Falloon, T. J., Danyushevsky, L. V., Ariskin, A., Green D. H., and Ford, C. E. (2007). The application of olivine geothermometry to infer crystallization temperatures of parental liquids: implications for the temperature of MORB magmas. *Chemical Geology*, **241**, 207–233.

Farnetani, C. G., and Samuel, H. (2005). Beyond the thermal plume paradigm. *Geophysical Research Letters*, **32**, L07311, doi:10.10029/2005GL022360.

Faryad, S. W., and Chakraborty, S. (2005). Duration of Eo-Alpine metamorphic events obtained from multicomponent diffusion modeling of garnet: a case study from the eastern Alps. *Contributions to Mineralogy and Petrology*, **150**, 306–318.

Faure, G. (1986). *Principles of Isotope Geology*, 2nd edn. New York: John Wiley.

Faure, G., and Mensing, T. M. (2004). *Isotopes: Principles and Applications*, 3rd edn. New York: John Wiley, 464 pp.

Feehan, J. G., and Brandon, M. T. (1999). Contribution of ductile flow to exhumation of low-temperature, high-pressure metamorphic rocks: San Juan-Cascade nappes, NW Washington State. *Journal of Geophysical Research*, **104**, 10883–10902.

Feineman, M. D., and DePaolo, D. J. (2003). Steady-state $^{226}Ra/^{230}Th$ disequilibrium in mantle minerals: implications for melt transport rates in island arcs. *Earth and Planetary Science Letters*, **215**, 339–355.

Ferry, J. M. (1980). A comparative study of geothermometers and geobarometers in pelitic schists from south-central Maine. *American Mineralogist*, **65**, 720–732.

Ferry, J. M. (1983). On the control of temperature, fluid composition, and reaction progress during metamorphism. *American Journal of Science*, **283-A**, 201–232.

Ferry, J. M. (1992). Regional metamorphism of the Waits River Formation, eastern Vermont: delineation of a new type of giant metamorphic hydrothermal system. *Journal of Petrology*, **33**, 45–94.

Ferry, J. M., and Dipple, G. M. (1991). Fluid flow, mineral reactions, and metasomatism. *Geology*, **19**, 211–214.

Ferry, J. M., and Spear, F. S. (1978). Experimental calibration of the partitioning of Fe and Mg between biotite and garnet. *Contributions to Mineralogy and Petrology*, **66**, 113–117.

Fincham, C. J. B., and Richardson, F. D. (1954). The behaviour of sulfur in silicate and aluminate melts. *Royal Society of London Philosophical Transactions*, **A233**, 40–62.

Fisher, G. W. (1973). Nonequilibrium thermodynamics as a model for diffusion-controlled metamorphic processes. *American Journal of Science*, **273**, 897–924.

Fisher, G. W., and Schminke, H. V. (1984). *Pyroclastic Rocks*. New York: Springer-Verlag, 339 pp.

Fisler, D. K., and Cygan, R. T. (1999). Diffusion of Ca and Mg in calcite. *American Mineralogist*, **84**, 1392–1399.

Fleischer, R. L., and Price, R. B. (1964). Techniques for geological dating of minerals by chemical etching of fission fragment tracks. *Geochimica et Cosmochimica Acta*, **28**, 1705–1714.

Fleischer, R. L., Price, R. B., and Walker, R. M. (1975). *Nuclear Tracks in Solids: Principles and Applications*. Berkeley, CA: University of California Press, 605 pp.

Fletcher, R. C., and Hofmann, A. W. (1974). Simple models of diffusion and combined diffusion-infiltration metasomatism. In *Geochemical Transport and Kinetics*, ed. A. W. Hofmann, B. J. Giletti, H. S. Yoder, and R. A. Yund. Washington, DC: Carnegie Institution, 243–259.

Foland, K. A., and Faul, H. (1977). Ages of the White Mountain intrusives – New Hampshire, Vermont, and Maine. *American Journal of Science*, **277**, 888–904.

Foulger, G. R., Natland, J. H., and Anderson, D. L. (2005). A source for Icelandic magmas in remelted Iapetus crust. *Journal of Volcanology and Geothermal Research*, **141**, 23–44.

Francis, P., Horrocks, L., and Oppenheimer, C. (2000). Monitoring gases from andesitic volcanoes. *Philosophical Transactions of the Royal Society of London*, **A 358**, 1567–1584.

Freestone, I. C. (1978). Liquid immiscibility in alkali-rich magmas. *Chemical Geology*, **23**, 115–123.

French, B. M. (1972). Shock-metamorphism features in the Sudbury structure: a review. In *New Developments in Sudbury Geology*, ed. J. Guy-Bray. *Geological Association of Canada Special Paper*, **10**, 19–28.

French, B. M. (1998). *Traces of Catastrophe: A Handbook of Shock-Metamorphic Effects in Terrestrial Meteorite Impact Structures*. Contribution No. 954. Houston: Lunar and Planetary Institute, 120 pp.

Frost, B. R., and Frost, C. D. (1987). CO_2, melts and granulite metamorphism. *Nature*, **327**, 503–506.

Fuhrman, M. L., and Lindsley, D. H. (1988). Ternary feldspar modeling and thermometry. *American Mineralogist*, **73**, 201 215.

Fujii, N., and Osamura, K. (1986). Effect of water saturation on the distribution of partial melt in the olivine–pyroxene–plagioclase system. *Journal of Geophysical Research*, **91**, 9253–9259.

Fyfe, W. S., Price, N. J., and Thompson, A. B. (1978). *Fluids in the Earth's Crust*. Amsterdam: Elsevier.

Gallagher, K., Brown, R., Johnson, C. (1998). Fission track analysis and its application to geological problems. *Annual Review of Earth and Planetary Sciences*, **26**, 519–572.

Ganguly, J. (2005). Adiabatic decompression and melting of mantle rocks: an irreversible thermodynamic analysis. *Geophysical Research Letters*, **32**, L06312, doi:1029/2005GL022365.

Ganguly, J., Cheng, W., and Chakraborty, S. (1998). Cation diffusion in aluminosilicate garnets: experimental determinations in pyrope-almandine diffusion couples. *Contributions to Mineralogy and Petrology*, **131**, 171–180.

Garcia, M. O., Pietruszka, A. J., Rhodes, J. M., and Swanson, K. (2000). Magmatic processes during the prolonged Pu'u 'O'o eruption of Kilauea volcano, Hawaii. *Journal of Petrology*, **41**, 967–990.

Gardner, J. E., Thomas, R. M. E., Jaupart, C., and Tait, S. (1996). Fragmentation of magma during Plinian volcanic eruptions. *Bulletin of Volcanology*, **58**, 144–162.

Garnero, E. J. (2000). Heterogeneity of the lowermost mantle. *Annual Review of Earth and Planetary Sciences*, **28**, 509–537.

Garven, G., and Freeze, A. R. (1984). Theoretical analysis of the role of groundwater flow in the genesis of stratabound ore deposits: 2. Quantitative results. *American Journal of Science*, **284**, 1085–1174.

Gass, I. G. (1970). Tectonic and magmatic evolution of the Afro-Arabian dome. In *African Magmatism and Tectonics*, ed. T. N. Clifford, and I. G. Gass. Edinburgh: Oliver & Boyd, 285–300.

Gebauer, D., Schertl, H. P., Brix, M., and Schreyer, W. (1997). 35 Ma old ultrahigh-pressure metamorphism and evidence for very rapid exhumation in the Dora Maira massif, Western Alps. *Lithos*, **41**, 5–24.

Geiser, P. A., and Sansone, S. (1981). Joints, microfractures, and the formation of solution cleavage in limestone. *Geology*, **9**, 280–285.

Ghent, E. D. (1976). Plagioclase–garnet–Al_2SiO_5–quartz: a potential geobarometer/geothermometer. *American Mineralogist*, **61**, 710–714.

Ghent, E. D., and Simony, P. S. (2005). Geometry of isogradic, isothermal, and isobaric surfaces: interpretation and application. *The Canadian Mineralogist*, **43**, 295–310.

Ghent, E. D., and Stout, M. Z. (1981). Geobarometry and geothermometry of plagioclase-biotite-garnet-muscovite assemblages. *Contributions to Mineralogy and Petrology*, **76**, 92–97.

Ghent, E. D., Robins, D. B., and Stout, M. Z. (1979). Geothermometry, geobarometry, and fluid compositions of metamorphosed calcsilicates and pelites, Mica Creek, British Columbia. *American Mineralogist*, **64**, 874–885.

Ghiorso, M. S. (1997). Thermodynamic models of igneous processes. *Annual Review of Earth and Planetary Sciences*, **25**, 221–241.

Ghiorso, M. S., and Sack, R. O. (1995). Chemical mass transfer in magmatic processes: IV. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid–solid equilibria in magmatic systems at elevated temperatures and pressures. *Contributions to Mineralogy and Petrology*, **119**, 197–212.

Ghiorso, M. S., Hirschmann, M. M., Reiners, P. W., and Kress, V. C., III (2002). A revision of MELTS aimed at improving calculation of phase relations and major element partitioning involved in partial melting of the mantle at pressures up to 3 GPa. *Geochemistry, Geophysics, Geosystems*, **3**(5), 10.1029/2001GC000217.

Gibb, F. G. F., and Henderson, C. M. B. (1992). Convection and crystal settling in sills. *Contributions to Mineralogy and Petrology*, **109**, 538–545.

Gibb, F. G. F., and Henderson, C. M. B. (2006). Chemistry of the Shiant Isles Main Sill, NW Scotland, and wider implications for the petrogenesis of mafic sills. *Journal of Petrology*, **47**, 191–230.

Gibbs, J. W. (1875). On the equilibrium of heterogeneous substances. *Connecticut Academy of Arts and Sciences Transactions*, **3**, 108–248.

Gilbert, G. K. (1877). *Report on the Geology of the Henry Mountains*. Washington, DC: U.S. Geographical and Geological Survey of the Rocky Mountain Region (Powell), 160 pp.

Gill, J. B. (1981). *Orogenic Andesites and Plate Tectonics*. New York: Springer-Verlag.

Goff, F. (1996). Vesicle cylinders in vapor-differentiated basalt flows. *Journal of Volcanology and Geothermal Research*, **71**, 167–185.

Goldschmidt, V. M. (1911). Die Kontaktmetamorphose im Kristianiagebiet. *Norske Videnskabers Selskabs Skrifter I, Mat.-Naturv. Klasse*, no. 1.

Goranson, R. W. (1938). Silicate–water systems: phase equilibria in the $NaAlSi_3O_8$–H_2O and $KAlSi_3O_8$–H_2O systems at high temperatures and pressures. *American Journal of Science*, **35A**, 71–91.

Graham, C. M., and England, P. C. (1976). Thermal regimes and regional metamorphism in the vicinity of overthrust faults: an example of shear heating and inverted metamorphic zonation from southern California. *Earth and Planetary Science Letters*, **31**, 142–152.

Grant, J. A. (1986). The isocon diagram: a simple solution to Gresens' equation for metasomatic alteration. *Economic Geology*, **81**, 1976–1982.

Gray, C. M., Cliff, R. A., and Goode, A. D. T. (1981). Neodymium-strontium isotopic evidence for extreme contamination in a layered basic intrusion. *Earth and Planetary Science Letters*, **56**, 189–198.

Gray, N. H. (1971). A parabolic hourglass structure in titanaugite. *American Mineralogist*, **56**, 952–958.

Gray, N. H. (1973). Estimation of parameters in petrologic materials balance equations. *Mathematical Geology*, **5**, 225–236.

Gray, N. H. (1978). Crystal growth and nucleation in flash-injected diabase dikes. *Canadian Journal of Earth Science*, **15**, 1904–1923.

Gray, N. H., Philpotts, A. R., and Dickson, L. D. (2003). Quantitative measures of textural anisotropy resulting from magmatic compaction illustrated by a sample from the Palisades sill, New Jersey. *Journal of Volcanology and Geothermal Research*, **121**, 293–312.

Great Plume Debate (2007). A series of papers based on the 2005 Chapman Conference on the "Great Plume Debate." *Chemical Geology*, **241**, 149–374.

Green, D. H. (1972). Archean greenstone belts may include terrestrial equivalents of lunar maria? *Earth and Planetary Science Letters*, **15**, 263–270.

Green, D. H. (1975). Genesis of Archean peridotitic magmas and constraints on Archean geothermal gradients and tectonics. *Geology* **3**, 15 18.

Greenwood, H. J. (1967a). Mineral equilibria in the system MgO-SiO_2-H_2O-CO_2. In *Researches in Geochemistry*, vol. 2, ed. P. H. Abelson. New York: John Wiley, 542–567.

Greenwood, H. J. (1967b). Wollastonite: stability in H_2O-CO_2 mixtures and occurrence in a contact metamorphic aureole near Salmo, British Columbia, Canada. *American Mineralogist*, **52**, 1669–1680.

Greenwood, H. J. (1975). Buffering of pore fluids by metamorphic reactions. *American Journal of Science*, **275**, 573–593.

Greenwood, H. J. (1976). Metamorphism at moderate temperatures and pressures. In *The Evolution of the Crystalline Rocks*, ed. D. K. Bailey, and R. MacDonald. London: Academic Press, 187–259.

Greig, J. W., and Barth, T. F. W. (1938). The system $Na_2O{\cdot}Al_2O_3{\cdot}5SiO_2$ (nephelite, carnegieite) – $Na_2O{\cdot}Al_2O_3{\cdot}6SiO_2$ (albite). *American Journal of Science*, **35A**, 93–112.

Gresens, R. L. (1967). Composition-volume relations of metasomatism. *Chemical Geology*, **2**, 47–65.

Grieve, R., and Therriault, A. (2000). Vredefort, Sudbury, Chicxulub: three of a kind? *Annual Review of Earth and Planetary Sciences*, **28**, 305–338.

Griffin, W. L., Wass, S. Y., and Hollis, J. D. (1984). Ultramafic xenoliths from Bullenmerri and Gnotuk maars, Victoria, Australia: petrology of a subcontinental crust–mantle transition. *Journal of Petrology*, **25**, 53–87.

Grove, T. L., Donnelly-Nolan, J. M., and Housh, T. B. (1997). Magmatic processes that generated the rhyolite of Glass Mountain, Medicine Lake Volcano, N. California. *Contributions to Mineralogy and Petrology*, **127**, 205–223.

Grover, J. E. (1980). Thermodynamics of pyroxenes. In *Pyroxenes*, ed. C. T. Prewitt, vol. 7 of *Reviews in Mineralogy*. Washington, DC: Mineralogical Society of America, 341–418.

Gudmundsson, A. (2000). Dynamics of volcanic systems in Iceland: example of tectonism and volcanism at juxtaposed hot spot and mid-ocean ridge systems. *Annual Review of Earth and Planetary Sciences*, **28**, 107–140.

Gudmundsson, S., Gudmundsson, M. T., Björnsson, H., *et al*. (2002). Three-dimensional glacier surface motion maps at the Gja'lp eruption site, Iceland, inferred from combining InSAR and other ice-displacement data. *Annals of Glaciology*, **34**, 315–322.

Guéguen, Y., and Palciauskas, V. (1994). *Introduction to the Physics of Rocks*. Princeton, NJ: Princeton University Press, 294 pp.

Guilbaud, M. N., Self, S., Thordarson, T., and Blake, S. (2005). Morphology, surface structures and emplacement of lavas produced by Laki, A. D. 1783–1784. *Geological Society of America, Special Paper*, **396**, 81–102.

Guy-Bray, J., and Geological Staff, International Nickel Co. (1966). Shatter cones at Sudbury. *Journal of Geology*, **74**, 243–245.

Haapala, I., and Rämö, O. T. (1999). Rapakivi granites and related rocks: an introduction. *Precambrian Research*, **95**, 1–7.

Haase, R. (1990). *Thermodynamics of Irreversible Processes*. New York: Dover.

Hacker, B. R., Kirby, S. H., and Bohlen, S. R. (1992). Time and metamorphic petrology: calcite to aragonite experiments. *Science*, **258**, 110–112.

Hacker, B. R., Peacock, S. M., Abers, G. A., and Holloway, S. D. (2003). Subduction factory: 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *Journal of Geophysical Research*, **108**, no. B1, doi:10.1029/2001JB001129.

Hakli, T. A., and Wright, T. L. (1967). The fractionation of nickel between olivine and augite as a geothermometer. *Geochimica et Cosmochimica Acta*, **31**, 877–884.

Hall, J. (1805). Experiments on whinstone and lava. *Transactions of the Royal Society of Edinburgh*, **5**, 43–74.

Hamilton, W. B. (2003). An alternative Earth. *GSA Today*, **13**(11), 4–12.

Handley, H. K., Macpherson, C. G., Davidson, J. P., Berlo, K., and Lowry, D. (2007). Constraining fluid sediment contributions to subduction-related magmatism in Indonesia: Ijen volcanic complex. *Journal of Petrology*, **48**, 1155–1183.

Hanson, R. B. (1992). Effects of fluid production on fluid flow during regional and contact metamorphism. *Journal of Metamorphic Geology*, **10**, 87–97.

Hanson, R. B. (1997). Hydrodynamics of regional metamorphism due to continental collision. *Economic Geology*, **92**, 880–891.

Hargraves, R. B., Johnson, D., Chan, C. Y. (1991). Distribution anisotropy: the cause of AMS in igneous rocks? *Geophysical Research Letters*, **18**, 2193–2196.

Harker, R. I., and Tuttle, O. F. (1956). Experimental data on the P_{CO2}-T curve for the reaction calcite + quartz = wollastonite + $\mathrm{CO_2}$. *American Mineralogist*, **265**, 239–256.

Harley, S. L. (1989). The origins of granulites – a metamorphic perspective. *Geological Magazine*, **126**, 215–247.

Harley, S. L. (2004). Extending our understanding of ultrahigh temperature crustal metamorphism. *Journal of Mineralogical and Petrological Sciences*, **99**, 140–158, doi: 10.2465/jmps.99.140.

Harlov, D. E., Wirth, R., and Forster, H. J. (2005). An experimental study of dissolution-reprecipitation in fluorapatite: fluid infiltration and the formation of monazite. *Contributions to Mineralogy and Petrology*, **150**, 268–286.

Harrison, T. M., Lovera, O. M., and Grove, M. (1997). New insights into the origin of two contrasting Himalayan granite belts. *Geology*, **25**, 899–902.

Hart, S. R., and Allègre, C. J. (1980). Trace-element constraints on magma genesis. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 121–151.

Hart, S. R., Hauri, E. H., Oschmann, L. A., Whitehead, J. A. (1992). Mantle plumes and entrainment: isotopic evidence. *Science*, **256**, 517–520.

Harte, B., and Hudson, N. F. C. (1979). Pelite facies series and temperatures and pressures of Dalradian metamorphism in Eastern Scotland. In *The Caledonides of the British Isles – Reviewed*, ed. A. L. Harris, C. H. Holland, and B. E. Leake. Geological Society of London, Special Publication, 8. Edinburgh: Scottish Academic Press, 323–337.

Hatherton, T., and Dickinson, W. R. (1969). The relationship between andesitic volcanism and seismicity in Indonesia, the Lesser Antilles and other island arcs. *Journal of Geophysical Research*, **74**, 5301–5310.

Haughton, D. R., Roeder, P. L., and Skinner, B. J. (1974). Solubility of sulfur in mafic magmas. *Economic Geology*, **69**, 451–467.

Hauzenberger, C. A., Baumgartner, L. P., and Pak, T. M. (2001). Experimental study on the solubility of the "model"-pelite mineral assemblage albite + K-feldspar + andalusite + quartz in supercritical chloride-rich aqueous solutions at 0.2 GPa and 600 °C. *Geochimica et Cosmochimica Acta*, **65**, 4493–4507.

Hawkesworth, C. J., and Kemp, A. I. S. (2006). Using hafnium and oxygen isotopes in zircons to unravel the record of crustal evolution. *Chemical Geology* **226**, 144 162.

Hawkesworth, C. J., Norry, M. J., Roddick, J. C., *et al*. (1979). ^{143}Nd/^{144}Nd, ^{87}Sr/^{86}Sr, and incompatible element variations in calc-alkaline andesites and plateau lavas from South America. *Earth and Planetary Science Letters*, **42**, 45–57.

Hawkesworth, C. J., Erlank, A. J., Marsh, J. S., Menzies, M. A., and van Calsteren, P. (1983). Evolution of the continental lithosphere: evidence from volcanics and xenoliths in southern Africa. In *Continental Basalts and Mantle Xenoliths*, ed. C. J. Hawkesworth, and M. J. Norry. Nantwich, Cheshire: Shiva Publishing, 111–138.

Hawkesworth, C. J., Blake, S., Evans, P. *et al*. (2000). Time scales of crystal fractionation in magma chambers – integrating physical, isotopic and geochemical perspectives. *Journal of Petrology*, **41**, 991–1006.

Hay, R. S., and Evans, B. (1988). Intergranular distribution of pore fluid and the nature of high-angle grain boundaries in limestone and marble. *Journal of Geophysical Research*, **93**, B8, 8959–8974.

Head, J. W., and Wilson, L. (1992). Lunar mare volcanism: stratigraphy, eruption conditions, and the evolution of secondary crusts. *Geochimica et Cosmochimica Acta*, **56**, 2155–2175.

Heald, E. F., Naughton, J. J., and Barnes, I. L., Jr. (1963). The chemistry of volcanic gases. 2. Use of equilibrium calculations in the interpretation of volcanic gas samples. *Journal of Geophysical Research*, **68**, 545–557.

Heath, E., Turner, S. P., Macdonald, R., Hawkesworth, C. J., and van Calsteren, P. (1998). Long magma residence times at an island arc volcano (Soufrière, St. Vincent) in the Lesser Antilles: evidence from ^{238}U–^{230}Th isochron dating. *Earth and Planetary Science Letters*, **160**, 49–63.

Heiken, G., Goff, F., Gardner, J. N., *et al*. (1990). The Valles/Toledo caldera complex, Jemez volcanic field, New Mexico. *Annual Review of Earth and Planetary Sciences*, **18**, 27–53.

Heinrich, E. W. (1966). *The Geology of Carbonatites*. Chicago: Rand-McNally.

Hekinian, R. (1982). *Petrology of the Ocean Floor*. Amsterdam: Elsevier Science.

Helgason, J. (1999). Formation of Olympus Mons and the aureole-escarpment problem on Mars. *Geology*, **27**, 231–234.

Helgeson, H. C. (1967). Solution chemistry and metamorphism. In *Researches in Geochemistry*, ed. P. H. Abelson. New York: John Wiley, 362–404.

Helgeson, H. C., Delany, J. M., Nesbitt, H. W., and Bird, D. K. (1978). Summary and critique of the thermodynamic properties of rock forming minerals. *American Journal of Science*, **278A**, 1–229.

Henderson, P. (1982). *Inorganic Geochemistry*. Oxford: Pergamon Press, 353 pp.

Hergt, J. M., Peate, D. W., and Hawkesworth, C. J. (1991). The petrogenesis of Mesozoic Gondwana low-Ti flood basalts. *Earth and Planetary Science Letters*, **105**, 134–148.

Hersum, T., Hilpert, M., and Marsh B. (2005). Permeability and melt flow in simulated and natural partially molten basaltic magmas. *Earth and Planetary Science Letters*, **237**, 798–814.

Herzberg, C. T. (1987). Magma density at high pressure. 2: A test of the olivine flotation hypothesis. In *Magmatic Processes: Physicochemical Principles*, ed., B. O. Mysen. Washington, DC: Geochemical Society Special Publication, 1, 47–58.

Herzberg, C. T., Fyfe, W. S., and Carr, M. J. (1983). Density constraints on the formation of the continental Moho and crust. *Contributions to Mineralogy and Petrology*, **84**, 1–5.

Hess, H. H. (1960). Stillwater igneous complex, Montana. *Geological Society of America Memoir* **80**, 230 pp.

Hess, P. C. (1969). The metamorphic paragenesis of cordierite in pelitic rocks. *Contributions to Mineralogy and Petrology*, **24**, 191–207.

Hess, P. C. (1980). Polymerization model for silicate melts. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 3–48.

Hewitt, D. A. (1973). The metamorphism of micaceous limestones from south-central Connecticut. *American Journal of Science*, **273-A**, 444–469.

Higgins, M. D. (1991). The origin of laminated and massive anorthosite, Sept Iles layered intrusion, Quebec, Canada. *Contributions to Mineralogy and Petrology*, **106**, 340–354.

Higgins, M. D. (1996). Magma dynamics beneath Kameni volcano, Thera, Greece, as revealed by crystal size and shape measurements. *Journal of Volcanology and Geothermal Research*, **70**, 37–48.

Higgins, M. D. (1998). Origin of anorthosite by textural coarsening: quantitative measurements of a natural sequence of textural development. *Journal of Petrology*, **39**, 1307–1323.

Higgins, M. D. (2000). Measurement of crystal size distributions. *American Mineralogist*, **85**, 1105–1116.

Higgins, M. D. (2006). *Quantitative Textural Measurements in Igneous and Metamorphic Rocks*. Cambridge: Cambridge University Press, 276 pp.

Higgins, M. W. (1971). Cataclastic rocks. *United States Geological Survey, Professional Paper*, **687**.

Hildreth, W. (1981). Gradients in silicic magma chambers: implications for lithospheric magmatism. *Journal of Geophysical Research*, **86**, 10153–10192.

Hildreth, W., and Wilson, C. J. N. (2007). Compositional zoning of the Bishop Tuff. *Journal of Petrology*, **48**, 951–999.

Hiraga, T., Osamu, N., Nagase, T., and Akizuki, M. (2001). Morphology of intergranular pores and wetting angles in pelitic schists studied by transmission electron microscopy. *Contributions to Mineralogy and Petrology*, **141**, 613–622.

Ho, A. M., and Cashman, K. V. (1997). Temperature constraints on the Ginko flow of the Columbia River Basalt Group. *Geology*, **25**, 403–406.

Hobbs, B. E., Means, W. D., and Williams, P. F. (1976). *An Outline of Structural Geology*. New York: John Wiley.

Hodges, K. V., and Crowley, P. (1985). Error estimation and empirical geothermobarometry for pelitic systems. *American Mineralogist*, **70**, 702–709.

Hodges, K. V., and Royden, L. (1984). Geologic thermobarometry of retrograded metamorphic rocks: an indication of the uplift trajectory of a portion of the northern Scandinavian Caledonides. *Journal of Geophysical Research*, **89**, 7077–7090.

Hodges, K. V., and Spear, F. S. (1982). Geothermometry, geobarometry, and the Al_2SiO_5 triple point at Mt. Moosilauke, New Hampshire. *American Mineralogist*, **67**, 1118–1134.

Hofmann, A. W. (1980). Diffusion in natural silicate melts: a critical review. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 385–417.

Hofmann, A. W. (1997). Mantle geochemistry: the message from oceanic volcanism. *Nature*, **385**, 219–229.

Hofmann, A. W., Giletti, B. J., Yoder, H. S., Jr., and Yund, R. A. (eds.) (1974). *Geochemical Transport and Kinetics*. London: Academic Press.

Hofmeister, A. M., and Criss, R. E. (2005). Earth's heat flux revised and linked to chemistry. *Tectonophysics*, **395**, 159–177.

Hoisch, T. D. (1987). Heat transport by fluids during Late Cretaceous regional metamorphism in the Big Maria Mountains, southeastern California. *Geological Society of America Bulletin*, **98**, 549–553.

Holdaway, M. J. (1971). Stability of andalusite and the aluminum silicate phase diagram. *American Journal of Science*, **271**, 97–131.

Holister, L. S. (1966). Garnet zoning: an interpretation based on the Rayleigh fractionation model. *Science*, **154**, 1647–1651.

Holister, L. S. (1970). Origin, mechanism and consequences of compositional sector zoning in staurolite. *American Mineralogist*, **55**, 742–766.

Holland, T. J. B. (1980). The reaction albite = jadeite + quartz determined experimentally in the range 600–1200 °C. *American Mineralogist*, **65**, 129–134.

Holland, T. J. B., and Powell, R. (1998). An internally consistent thermodynamic data set for phases of petrological interest. *Journal of Metamorphic Geology*, **16**, 309–343.

Holness, M. B. (1993). Temperature and pressure dependence of quartz-aqueous fluid dihedral angles: the control of adsorbed H_2O on the permeability of quartzites. *Earth and Planetary Science Letters*, **117**, 363–377.

Holness, M. B. (2006). Melt-solid dihedral angles of common minerals in natural rocks. *Journal of Petrology*, **47**, 791–800.

Holness, M. B., Cheadle, M. J., and McKenzie, D. (2005). On the use of changes in dihedral angle to decode late-stage textural evolution in cumulates. *Journal of Petrology*, **46**, 1565–1583.

Holtz, F., Behrens, H., Dingwell, D. B., and Johannes W. (1995). H_2O solubility in haplogranitic melts: compositional, pressure, and temperature dependence. *American Mineralogist*, **80**, 94–108.

Holtzman, B. K., and Kohlstedt, D. L. (2007). Stress-driven melt segregation and strain partitioning in partially molten rocks: effects of stress and strain. *Journal of Petrology*, **48**, 2379–2406.

Holtzman, B. K., Groebner, N. J., Zimmerman, M. E., Ginsberg, S. B., and Kohlstedt, D. L. (2003). Stress-driven melt segregation in partially molten rocks. *Geochemistry Geophysics Geosystems*, **4**(5), 8607, doi:10.1029/2001GC000258.

Holtzman, B. K., Kohlstedt, D. L., and Phipps Morgan, J. (2005). Viscous energy dissipation and strain partitioning in partially molten rocks. *Journal of Petrology*, **46**, 2569–2592.

Hort, M., Marsh, B. D., Resmini, R. G., and Smith, M. K. (1999). Convection and crystallization in a liquid cooled from above: an experimental and theoretical study. *Journal of Petrology*, **40**, 1271–1300.

Hoskuldsson, A., and Sparks, R. S. J. (1997). Thermodynamics and fluid dynamics of effusive subglacial eruptions. *Bulletin of Volcanology*, **59**, 219–230.

Huang, W. L., and Wyllie, P. J. (1973). Melting of muscovite-granite to 35 kbar as a model for fusion of metamorphosed subducted oceanic sediments. *Contributions to Mineralogy and Petrology*, **42**, 1–14.

Hudon, P., Yung, I., and Baker, D. R. (2005). Experimental investigation and optimization of thermodynamic properties and phase diagrams in the system $CaO–SiO_2$, $MgO–SiO_2$, $CaMgSi_2O_6–SiO_2$ and $CaMgSi_2O_6–Mg_2SiO_4$. *Journal of Petrology*, **46**, 1859–1880.

Huppert, H. E., and Sparks, R. S. J. (1980). The fluid dynamics of a basaltic magma chamber replenished by influx of hot, dense ultrabasic magma. *Contributions to Mineralogy and Petrology*, **75**, 279–289.

Huppert, H. E., and Sparks, R. S. J. (1985). Cooling and contamination of mafic and ultramafic magmas during ascent through continental crust. *Earth and Planetary Science Letters*, **74**, 371 386.

Huppert, H. E., and Sparks, R. S. J. (1988). The generation of granitic magmas by intrusion of basalt into continental crust. *Journal of Petrology*, **29**, 599–624.

Huppert, H. E., Sparks, R. S. J., Turner, J. S., and Arndt, N. T. (1984). Emplacement and cooling of komatiite lavas. *Nature*, **309**, 19–22.

Huppert, H. E., Sparks, R. S. J., Wilson, J. R., and Hallworth, M. A. (1986). Cooling and crystallization at an inclined plane. *Earth and Planetary Science Letters*, **79**, 319–328.

Huppert, H. E., Sparks, R. S. J., Wilson, J. R., Hallworth, M. A., and Leitch, A. M. (1987). Laboratory experiments with aqueous solutions modelling magma chamber processes: II. Cooling and crystallization along inclined planes. In *Origins of Igneous Layering*, ed. I. Parsons. Dordrecht: D. Reidel, 539–568.

Hussenoeder, S. A., Collins, J. A., Kent, G.M., Detrick, R. S., and the TERA Group (1996). Seismic analysis of the axial magma chamber reflector along the southern East Pacific Rise from conventional reflection profiling. *Journal of Geophysical Research*, **101**, 22087–22105.

Ildefonse, J.-P., and Gabis, V. (1976). Experimental study of silica diffusion during metasomatic reactions in the presence of water at 550 °C and 1000 bars. *Geochimica et Cosmochimica Acta*, **40**, 297–303.

Ingebritsen, S. E., and Manning, C. E. (1999). Geological implications of a permeability-depth curve for the continental crust. *Geology*, **27**, 1107–1110.

Ingersoll, L. R., Zobel, O. J., and Ingersoll, A. C. (1954). *Heat Conduction*. University of Wisconsin Press, 325 pp.

Irvine, T. N. (1974). Petrology of the Duke Island Ultramafic Complex, southeastern Alaska. *Geological Society of America Memoir*, **138**.

Irvine, T. N. (1977). Chromite crystallization in the join Mg_2SiO_4–$CaMgSi_2O_6$–$CaAl_2Si_2O_8$–$MgCr_2O_4$–SiO_2. *Carnegie Institution of Washington Yearbook*, **76**, 465–472.

Irvine, T. N. (1979). Rocks whose composition is determined by crystal accumulation and sorting. In *The Evolution of the Igneous Rocks: Fiftieth Anniversary Perspectives*, ed. H. S. Yoder, Jr. Princeton, NJ: Princeton University Press, 245–306.

Irvine, T. N. (1980). Magmatic infiltration metasomatism, double-diffusive fractional crystallization, and adcumulus growth in the Muskox Intrusion and other layered intrusions. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 325–385.

Irvine, T. N. (1982). Terminology for layered intrusions. *Journal of Petrology*, **23**, 127–162.

Irvine, T. N. (1987). Layering and related structures in the Duke Island and Skaergaard Intrusions: similarities, differences, and origins. In *Origins of Igneous Layering*, ed. I. Parsons. Dordrecht: D. Reidel, 185–246.

Irvine, T. N., and Baragar, W. R. A. (1971). A guide to the chemical classification of the common volcanic rocks. *Canadian Journal of Earth Science*, **8**, 523–548.

Irvine, T. N., and Smith, C. H. (1967). The ultramafic rocks of the Muskox Intrusion, Northwest Territories, Canada. In *Ultramafic and Related Rocks*, ed. P. J. Wyllie. New York: John Wiley, 38–49.

Irvine, T. N., Keith, D. W., and Todd, S. G. (1983). The J-M platinum reef of the Stillwater Complex, Montana: II. Origin by double-diffusive convective magma mixing and implications for the Bushveld Complex. *Economic Geology*, **78**, 1287–1334.

Irvine, T. N., Andersen, J. C.Ø., and Brooks, C. K. (1998). Included blocks (and blocks within blocks) in the Skaergaard Intrusion: geological relations and the origins of rhythmic modally graded layers. *Geological Society of America Bulletin*, **110**, 1398–1447.

Isachsen, Y. W. (ed.) (1968). *Origin of Anorthosites and Related Rocks*. *New York State Museum and Science Service Memoir*, **18**.

Jackson, E. D. (1967). Ultramafic cumulates in the Stillwater, Great Dyke, and Bushveld Intrusion. In *Ultramafic and Related Rocks*, ed. P. J. Wyllie. New York: John Wiley, 20–38.

Jacobsen, S. B. (2003). How old is planet Earth? *Science*, **300**, 1513–1514.

Jacobsen, S. B. (2005). The Hf–W isotopic system and the origin of the Earth and Moon. *Annual Review of Earth and Planetary Sciences*, **33**, 531–570.

Jaeger, J. C. (1968). Cooling and solidification of igneous rocks. In *Basalts*, vol. 2, ed. H. H. Hess and A. Poldervaart. New York: John Wiley, 503–536.

Jager, E., and Hunziker, J. C., (ed.) (1979). *Lectures in Isotope Geology*. Berlin: Springer-Verlag, 329 pp.

Jahns, R. H., and Burnham, C. W. (1969). Experimental studies of pegmatite genesis, I: a model for the derivation and crystallization of granitic pegmatites. *Economic Geology*, **64**, 843–864.

Jakobsen, J. D., Veksler, I. V., Tegner, C., and Brooks, C. K. (2005). Immiscible iron- and silica-rich melt in basalt petrogenesis documented in the Skaergaard intrusion. *Geology*, **33**, 885–888.

Jamtveit, B., Bucher-Nurminen, K., and Austrheim, H. (1989). Fluid controlled eclogitization of granulites in deep crustal shear zones, Bergen Arcs, Western Norway. *Contributions to Mineralogy and Petrology*, **104**, 184–193.

Jaupart, C., and Tait, S. (1995). Dynamics of differentiation in magma reservoirs. *Journal of Geophysical Research*, **100**, 17615–17636.

Javoy, M., Fourcade, S., and Allègre, C. J. (1970). Graphical method for examination of $^{18}O/^{16}O$ fractionations in silicate rocks. *Earth and Planetary Science Letters*, **10**, 12–16.

Jerram, D. A., and Higgins, M. D. (2007). 3D analysis of rock textures: quantifying igneous rock textures. *Elements*, **3**, 239–245.

Joesten, R. L. (1974). Local equilibrium and metasomatic growth of zoned calc-silicate nodules in a contact aureole, Christmas Mountains, Big Bend Region, Texas. *American Journal of Science*, **274**, 876–901.

Joesten, R. (1976). High-temperature contact metamorphism of carbonate rocks in a shallow crustal environment, Christmas Mountains, Big Bend Region, Texas. *American Mineralogist*, **61**, 776–781.

Joesten, R. (1977). Evolution of mineral assemblage zoning in diffusion metasomatism. *Geochimica et Cosmochimica Acta*, **41**, 649–670.

Joesten, R. (1979). Kinetics of diffusion-controlled mineral growth in the Christmas Mountains, Texas, contact aureole. *Mineralogical Society of London Bulletin*, **42**, 3.

Joesten, R. L. (1983). Grain growth and grain-boundary diffusion in quartz from the Christmas Mountains (Texas) contact aureole. *American Journal of Science*, **283-A**, 233–254.

Joesten, R. L., and Fisher, G. (1988). Kinetics of diffusion-controlled mineral growth in the Christmas Mountains (Texas) contact aureole. *Geological Society of America Bulletin*, **100**, 714–732.

Johannsen, A. (1931). *A Descriptive Petrography of the Igneous Rocks*: vol. I. *Introduction, Textures, Classification and Glossary*. Chicago: University of Chicago Press, 267 pp.

Johnson, A. M. (1970). *Physical Processes in Geology*. San Francisco, CA: Freeman, Cooper, 577 pp.

Johnson, A. M., and Pollard, D. D. (1973). Mechanics of growth of some laccolithic intrusions in the Henry Mountains, Utah: I. *Tectonophysics*, **18**, 261–309.

Johnson, M. C., and Plank, T. (1999). Dehydration and melting experiments constrain the fate of subducted sediments. *Geochemistry, Geophysics and Geosystems*, **1**, paper number 1999GC000014.

Jónsson, S., Zebker, H., and Amelung, F. (2005). On trapdoor faulting at Sierra Negra volcano, Galápagos. *Journal of Volcanology and Geothermal Research*, **144**, 59–71.

Josephsen, K. (2003). Magmakammerprocesser i Fongen-Hyllingen Intrusionen, Norge – et detaljestudie af modal lagdeling. (Magma chamber process in the Fongen-Hyllingen Intrusion, Norway – a detailed study of modal layering). Unpublished MSc thesis, University of Aarhus, 120 pp.

Jurewicz, S. R., and Watson, E. B. (1985). The distribution of partial melt in a granitic system: the application of liquid phase sintering theory. *Geochimica et Cosmochimica Acta*, **49**, 1109–1121.

Kalsbeek, F., and Jepsen, H. F. (1984). The late Proterozoic Zig-Zag Dal basalt formation of eastern North Greenland. *Journal of Petrology*, **25**, 644–664.

Karabinos, P., and Ketcham, R. (1988). Thermal structure of active thrust belts. *Journal of Metamorphic Geology*, **6**, 559–570.

Kattenhorn, S. A., and Watkeys, M. K. (1995). Blunt-ended dyke segments. *Journal of Structural Geology*, **17**, 1535–1542.

Katz, R. F., Spiegelman, M., and Holtzman, B. (2006). The dynamics of melt and shear localization in partially molten aggregates. *Nature*, **442**, 676–679.

Kay, R. W., and Kay, S. M. (1993). Delamination and delamination magmatism. *Tectonophysics*, **219**, 177–189.

Kelemen, P. B., Whitehead, J. A., Aharonov, E., and Jordahl, K. A. (1995). Experiments on flow focusing in partially soluble porous media, with applications to melt extraction from the mantle. *Journal of Geophysical Research*, **100**, 475–496.

Kelemen, P. B., Hirth, G., Shimizu, N., Spiegelman, N., and Dick, H. J. B. (1997). A review of melt migration processes in the adiabatically upwelling mantle beneath oceanic spreading ridges. *Philosophical Transactions of the Royal Society of London*, **A 355**, 1–35.

Kelemen, P. B., Hanghøj, K. and Greene, A. R. (2003). One view of the geochemistry of subduction-related magmatic arcs, with an emphasis on primitive andesite and lower crust. In *The Crust*, ed. R. L. Rudnick, vol. 3 of *Treatise on Geochemistry*, ed. H. D. Holland, and K. K. Turekian. Amsterdam: Elsevier, 593–659.

Kelsey, C. H. (1965). Calculation of the CIPW norm. *Mineralogical Magazine*, **34**, 276–282.

Kelsey, D. E. (2008). On ultrahigh-temperature crustal metamorphism. *Gondwana Research*, **13**, 1–29.

Kennedy, C. S., and Kennedy, G. C. (1976). The equilibrium boundary between graphite and diamond. *Journal of Geophysical Research*, **81**, 2467–2470.

Kepezhinskas, K. B., and Khlestov, V. V. (1977). The petrogenetic grid and subfacies for middle-temperature metapelites. *Journal of Petrolology*, **18**, 114–143.

Kern, R, and Weisbrod, A. (1967). *Thermodynamics for Geologists*. San Francisco: Freeman, Cooper, 304 pp.

Kerr, A. C., and Mahoney, J. J. (2007). Oceanic plateaus: problematic plumes, potential paradigms. *Chemical Geology*, **241**, 332–353.

Kerrick, D. M. (1972). Experimental determination of the muscovite + quartz stability with $P_{H_2O}<$ P$_{total}$. *American Journal of Science* **272**, 946 958.

Kerrick, D. M., and Connolly, J. A. D. (2001a). Metamorphic devolatilization of subducted marine sediments and the transport of volatiles into the Earth's mantle. *Nature*, **411**, 293–296.

Kerrick, D. M., and Connolly, J. A. D. (2001b). Metamorphic devolatilization of subducted oceanic metabasalts: implications for seismicity, arc magmatism and volatile recycling. *Earth and Planetary Science Letters*, **189**, 19–29.

Kerrick, D. M., and Jacobs, G. K. (1981). A modified Redlich-Kwong equation for H_2O, CO_2, and H_2O–CO_2 mixtures at elevated pressures and temperatures. *American Journal of Science*, **281**, 735–767.

Kieffer, S. W. (1981). Blast dynamics at Mount St Helens on 18 May 1980. *Nature*, **291**, 568–570.

King, R. L., Kohn, M. J., and Eiler, J. M. (2003). Constraints on the petrologic structure of the slab-mantle interface from Franciscan Complex exotic ultramafic blocks. *Geological Society of America Bulletin*, **115**, 1097–1109.

King, S. D. (1995). Models of mantle viscosity. In *Mineral Physics and Crystallography: A Handbook of Physical Constants*, ed. T. J. Ahrens. American Geophysical Union Reference Shelf, 2, 227–236.

Kingery, W. D. (1960). *Introduction to Ceramics*, New York: John Wiley, 781 pp.

Kingsley, L. (1931). Cauldron subsidence of the Ossipee Mountains. *American Journal of Science*, **222**, 139–168.

Kirkpatrick, R. J. (1974). Kinetics of crystal growth in the system $CaMgSi_2O_6$–$CaAl_2SiO_6$. *American Journal of Science*, **274**, 215–242.

Kirkpatrick, R. J. (1975). Crystal growth from the melt: a review. *American Mineralogist*, **60**, 798–814.

Kirkpatrick, R. J. (1977). Nucleation and growth of plagioclase, Makaopuhi and Alae lava lakes, Kilauea. *Geological Society of America Bulletin*, **88**, 78–84.

Kirkpatrick, R. J. (1981). Kinetics of crystallization of igneous rocks. In *Kinetics of Geochemical Processes*, ed. A. C. Lasaga and R. J. Kirkpatrick, vol. 8 of *Reviews in Mineralogy.* Washington, DC: Mineralogical Society of America, 321–398.

Kirkpatrick, R. J., Klein, L., Uhlmann, D. R., and Hays, J. F. (1979). Rates and processes of crystal growth in the system anorthite–albite. *Journal of Geophysical Research*, **84**, 3671–3676.

Kjarsgaard, B. A., and Hamilton, D. L. (1988). Liquid immiscibility and the origin of alkali-poor carbonatites. *Mineralogical Magazine*, **52**, 43–55.

Klein, C. (2005). Some Precambrian banded iron-formations (BIFs) from around the world: their age, geologic setting, mineralogy, metamorphism, geochemistry, and origins. *American Mineralogist*, **90**, 1473–1499.

Klein, F. W. (1984). Eruption forecasting at Kilauea Volcano, Hawaii. *Journal of Geophysical Research*, **89**, 3059–3073.

Kohn, M. J. (2003). Geochemical zoning in metamorphic minerals. In *The Crust*, ed. R. L. Rudnick, vol. 3 of *Treatise on Geochemistry*, ed. H. D. Holland, and K. K. Turekian. Amsterdam: Elsevier, 229–261.

Kohn, M. J., and Spear, F. S. (1990). Two new geobarometers for garnet amphibolites, with application to southeastern Vermont. *American Mineralogist*, **75**, 89–96.

Kohn, M. J., and Spear, F. S. (2000). Retrograde net transfer insurance for pressure-temperature estimates. *Geology*, **28**, 1127–1130.

Komar, P. D. (1972). Flow differentiation in igneous dikes and sills: profiles of velocity and phenocryst concentration. *Geological Society of America Bulletin* **83**, 3443 3448.

Korenaga, J. (2007). Thermal cracking and the deep hydration of oceanic lithosphere: a key to the generation of plate tectonics? *Journal of Geophysical Research*, **112**, B05408, doi:10.1029/2006JB004502.

Korja, A., Heikkinen, P., and Aaro, S. (2001). Crustal structure of the northern Baltic Sea paleorift. *Tectonophysics*, **331**, 341–358.

Korzhinskii, D. S. (1970). *Theory of Metasomatic Zoning*. Oxford: Clarendon Press.

Koster Van Groos, A. F., and Wyllie, P. J. (1966). Liquid immiscibility in the system Na_2O–Al_2O_3–SiO_2–CO_2 at pressures to 1 kilobar. *American Journal of Science*, **264**, 234–255.

Kretz, R. (1966). Interpretation of the shape of mineral grains in metamorphic rocks. *Journal of Petrology*, **7**, 68–94.

Kretz, R. (1973). Kinetics of the crystallization of garnet at two localities near Yellowknife. *Canadian Mineralogist*, **12**, 1–20.

Kuno, H. (1960). High alumina basalt. *Journal of Petrology*, **1**, 125–145.

Kurszlaukis, S, Büttner, R., Zimanowski, B., and Lorenz, V. (1998). On the first experimental phreatomagmatic explosion of a kimberlite melt. *Journal of Volcanology and Geothermal Research*, **80**, 323–326.

Kushiro, I. (1969). The system forsterite–diopside–silica with and without water at high pressures. *American Journal of Science*, **267-A**, 269–294.

Kushiro, I. (1970). Stability of amphibole and phlogopite in the upper mantle. *Carnegie Institution of Washington Yearbook*, **68**, 245–247.

Kushiro, I. (1972). Effect of water on the composition of magmas formed at high pressures. *Journal of Petrology*, **13**, 311–334.

Kushiro, I. (1980). Viscosity, density, and structure of silicate melts at high pressures, and their petrological applications. In *Physics of Magmatic Processes*, ed. R.B. Hargraves. Princeton: Princeton University Press, 93–120.

Kushiro, I., and Yoder, H. S., Jr. (1966). Anorthite–forsterite and anorthite–enstatite reactions and their bearing on the basalt–eclogite transformation. *Journal of Petrology*, **7**, 337–362.

Kushiro, I., Syong, Y., and Akimoto, S. (1968). Melting of a peridotite nodule at high pressures and high water pressures. *Journal of Geophysical Research*, **73**, 6023–6029.

Lachenbruch, A. H. (1968). Preliminary geothermal model of the Sierra Nevada. *Journal of Geophysical Research*, **73**, 6977–6990.

Lachenbruch, A. H., and Sass, J. H. (1977). Heat flow in the United States and the thermal regime of the crust. In *The Earth's Crust*, ed. J. G. Heacock. American Geophysical Union Monograph, 20. Washington, DC: AGU, 626–675.

Landtwing, M. R., Dillenbeck, E. D., Leake, M. H., and Heinrich, C. A. (2002). Evolution of breccia-hosted porphyry Cu-Mo-Au deposit at Agua Rica, Argentina: progressive unroofing of a magmatic hydrothermal system. *Economic Geology*, **97**, 1273–1292.

Langmuir, C. H., Klein, E. M. and Plank, T. (1992). Petrological systematics of mid-ocean ridge basalts: constraints on melt generation beneath ocean ridges. In *Mantle Flow and Melt Generation at Mid-ocean Ridges*, ed. J. Phipps Morgan *et al*. Geophysical Monograph, 71. Washington, DC: American Geophysical Union, 183–280.

Lassey, K. R., and Blattner, P. (1988). Kinetically controlled oxygen istotope exchange between fluid and rock in one-dimensional advective flow. *Geochimica et Cosmochimica Acta*, **52**, 2169 2175.

Latypov, R. M. (2003a). The origin of marginal compositional reversals in basic–ultrabasic sills and layered intrusions by Soret fractionation. *Journal of Petrology*, **44**, 1579–1618.

Latypov, R. M. (2003b). The origin of basic–ultrabasic sills with S-, D-, and I-shaped compositional profiles by in situ crystallization of a single input of phenocryst-poor parental magma. *Journal of Petrology*, **44**, 1619–1656.

Lay, T., Garnero, E. J., and Williams, Q. (2004). Partial melting in a thermo-chemical boundary layer at the base of the mantle. *Physics of the Earth and Planetary Interiors*, **146**, 441–467.

Le Bas, M. J., Le Maitre, R. W., Streckeisen, A., and Zanettin, B. (1986). A chemical classification of volcanic rocks based on the total alkali-silica diagram. *Journal of Petrology*, **27**, 745–750.

Lee, K. K. M., and Jeanloz, R. (2003). High-pressure alloying of potassium and iron: radioactivity in the Earth's core. *Geophysical Research Letters*, **30**(23), 2212, doi:10.1029/2003GL018515.

Léger, A., and Ferry, J. M. (1993). Fluid infiltration and regional metamorphism of the Waits River Formation, northeast Vermont, USA. *Journal of Metamorphic Geology*, **11**, 3–29.

Lesher, C. E., and Walker, D. (1991). Thermal diffusion in petrology. In *Diffusion, Atomic Ordering, and Mass Transport: Selected Topics in Geochemistry*, ed. J. Ganguly. Advances in Physical Chemistry, 8. New York: Springer, 396–451.

Levin, E. M., Robbins, C. R., and McMurdie, H. F. (1964). *Phase Diagrams for Ceramists*. Columbus, OH: American Ceramic Society.

Lindsley, D. H. (1983). Pyroxene thermometry. *American Mineralogist*, **68**, 477–493.

Lister, G. S., and Baldwin, S. L. (1993). Plutonism and the origin of metamorphic core complexes. *Geology*, **21**, 607–610.

Lister, J. R., and Kerr, R. C. (1991). Fluid-mechanical models of crack propagation and their application to magma transport in dykes. *Journal of Geophysical Research*, **96**, 10049–10077.

Liu, X., O'Neill, H.C, and Berry, A. J. (2006). The effects of small amounts of H_2O, CO_2 and Na_2O on the partial melting of spinel lherzolite in the system CaO–MgO–Al_2O_3–$SiO_2 \pm H_2O \pm CO_2 \pm Na_2O$ at 1.1 GPa. *Journal of Petrology*, **47**, 409–434.

Lofgren, G. (1974). An experimental study of plagioclase crystal morphology: isothermal crystallization. *American Journal of Science*, **274**, 243–273.

Lofgren, G. (1980). Experimental studies on the dynamic crystallization of silicate melts. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 487–551.

London, D. (1987). Internal differentiation of rare-element pegmatites: effects of boron, phosphorous, and fluorine. *Geochimica et Cosmochimica Acta*, **51**, 403–420.

London, D. (2005). Granitic pegmatites: an assessment of current concepts and directions for the future. *Lithos*, **80**, 281–303.

Long, P. E., and Wood, B. J. (1986). Structures, textures, and cooling histories of Columbia River basalt flows. *Geological Society of America Bulletin*, **97**, 1144–1155.

Lonsdale, P. (1985). Nontransform offsets of the Pacific-Cocos plate boundary and their traces on the rise flank. *Geological Society of America Bulletin*, **96**, 313–327.

Lorenz, V. (1975). Formation of phreatomagmatic maar-diatreme volcanoes and its relevance to kimberlite diatremes. *Physics and Chemistry of the Earth*, **9**, 17–27.

Luth, W. C., Jahns, R. H., and Tuttle, O. F. (1964). The granite system at pressures of 4 to 10 kilobars. *Journal of Geophysical Research*, **69**, 759 773.

Lüttge, A., Bolton, E. W., and Rye, D. M. (2004). A kinetic model of metamorphism: an application to siliceous dolomites. *Contributions to Mineralogy and Petrology*, **146**, 546–565.

Lux, D. R., DeYoreo, J. J., and Guidotti, C. V., and Decker, E. R. (1986). Role of plutonism in low-pressure metamorphic belt evolution. *Nature*, **323**, 794–797.

Lyle, P. (2000). The eruption environment of multi-tiered columnar basalt lava flows. *Journal of the Geological Society of London*, **157**, 715–722.

Lynn, H. B., Hale, L. D., and Thompson, G. A. (1981). Seismic reflections from the basal contacts of batholiths. *Journal of Geophysical Research*, **86**, 10633–10638.

Maaløe, S. (2003). Melt dynamics of a partially molten mantle with randomly oriented fractures. *Journal of Petrology*, **44**, 1193–1210.

Maaløe, S., and Printzlau, I. (1979). Natural partial melting of spinel lherzolite. *Journal of Petrology*, **20**, 727–741.

MacDonald, G. A. (1972). *Volcanoes*. Englewood Cliffs, NJ: Prentice-Hall, 510 pp.

Macgregor, A. M. (1951). Some milestones in the Precambrian of southern Rhodesia. *Proceedings of the Geological Society of South Africa Transactions*, **54**, 27–71.

Mackwell, S. J., and Kohlstedt, D. L. (1990). Diffusion of hydrogen in olivine: implications for water in the mantle. *Journal of Geophysical Research*, **95**, 5079–5088.

MacLean, W. H. (1969). Liquidus phase relations in the FeS–FeO–Fe_3O_4–SiO_2 system, and their application in geology. *Economic Geology*, **64**, 865–884.

Mahon, K. I., Harrison, T. M., and Drew, D. A. (1988). Ascent of a granitoid diapir in a temperature varying medium. *Journal of Geophysical Research*, **93**, 1174–1188.

Manning, C. E. (1994). The solubility of quartz in the lower crust and upper mantle. *Geochimica et Cosmochimica Acta*, **58**, 4831–4839.

Manning, C. E. (1997). Coupled reaction and flow in subduction zones: silica metasomatism in the mantle wedge. In *Fluid Flow and Transport in Rocks*, ed. B. Jamtveit, and B. W. D. Yardley. London: Chapman-Hall, 139–148.

Manning, C. E. (2001). Experimental studies of fluid-rock interaction at high-pressure: the role of polymerization and depolymerization of solutes. Hot Springs, VA, 11th Annual V. M. Goldschmidt Conference. Washington, DC: Mineralogical Society of America.

Manning, C. E. (2007). Solubility of corundum + kyanite in H_2O at 700 °C and 10 kbar: evidence for Al-Si complexing at high pressure and temperature. *Geofluids*, **7**, 258–269.

Manning, D. A. C., and Pichavant, M. (1983). The role of fluorine and boron in the generation of granitic melts. In *Migmatites, Melting, and Metamorphism*, ed. M. P. Atherton, and C. D. Gribble. Cambridge, MA: Birkhauser Boston, 94–109.

Markl, G., Ferry, J., and Bucher, K. (1998). Formation of saline brines and salt in the lower crust by hydration reactions in partially retrogressed granulites from the Lofoten Islands, Norway. *American Journal of Science*, **298**, 705–757.

Marsh, B. D. (1979). Island arc volcanism. *American Scientist*, **67**, 161–172.

Marsh, B. D. (1981). On the crystallinity, probability of occurrence, and rheology of lava and magma. *Contributions to Mineralogy and Petrology*, **78**, 85–98.

Marsh, B. D. (1982). On the mechanism of igneous diapirism, stoping, and zone melting. *American Journal of Science*, **282**, 808 855.

Marsh, B. D. (1988a). Crystal size distribution (CSD) in rocks and the kinetics and dynamics of crystallization: I. Theory. *Contributions to Mineralogy and Petrology*, **99**, 277–291.

Marsh, B. D. (1988b). Crystal capture, sorting, and retention in convecting magma. *Geological Society of America Bulletin*, **100**, 1720–1737.

Marsh, B. D. (1989). On convective style and vigor in sheet-like magma chambers. *Journal of Petrology*, **30**, 479–530.

Marsh, B. D. (1998). On the interpretation of crystal size distributions in magmatic systems. *Journal of Petrology*, **39**, 553–599.

Marsh, B. D., and Carmichael, I. S. E. (1974). Benioff zone magmatism. *Journal of Geophysical Research*, **79**, 1196–1206.

Marsh, B. D., and Kantha, L. H. (1978). On the heat and mass transfer from an ascending magma. *Earth and Planetary Science Letters*, **39**, 435–443.

Marsh, B. D., and Maxey, M. R. (1985). On the distribution and separation of crystals in convecting magma. *Journal of Volcanology and Geothermal Research*, **24**, 95–150.

Martel, C., Ali, A. R., Poussineau, S., Gourgaud, A., and Pichavant, M. (2006). Basalt-inherited microlites in silicic magmas: evidence from Mount Pelée (Martinique, French West Indies). *Geology*, **34**, 905–908.

Martin, D., Griffiths, W., and Campbell, I. H. (1987). Compositional and thermal convection in magma chambers. *Contributions to Mineralogy and Petrology*, **96**, 465–475.

Martin, H. (1987). Petrogenesis of Archaean trondhjemites, tonalites, and granodiorites from eastern Finland: major and trace element geochemistry. *Journal of Petrology*, **28**, 921–953.

Martini, J. E. J. (1991). The nature, distribution and genesis of the coesite and stishovite associated with the pseudotachylite of the Vredefort Dome, South Africa. *Earth and Planetary Science Letters*, **103**, 285–300.

Maruyama, S., and Liou, J. G. (1998). Initiation of ultrahigh-pressure metamorphism and its significance on the Proterozoic-Phanerozoic boundary. *The Island Arc*, **7**, 6–35.

Masaitis, V. L., Shafranovsky, G. I., Grieve, R. A. F., *et al*. (1999). Impact diamonds in the suevitic breccias of the Black Member of the Onaping Formation, Sudbury Structure, Ontario, Canada. In *Large Meteorite Impacts and Planetary Evolution: II*, ed. B. O. Dressler, and V. L. Sharpton. *Geological Society of America Special Paper*, **339**, 317–321.

Massone, H. (2001). First find of coesite in the ultrahigh-pressure metamorphic area of the central Erzgebirge, Germany. *European Journal of Mineralogy*, **13**, 565–570.

Masters, R. L., and Ague, J. J. (2005). Regional-scale fluid flow and element mobility in Barrow's metamorphic zones, Stonehaven, Scotland. *Contributions to Mineralogy and Petrology*, **150**, 1–18.

Masters, T. G., and Shearer, P. M. (1995). Seismic models of the Earth: elastic and anelastic. In *Global Earth Physics: A Handbook of Physical Constants*, ed. T. J. Ahrens. American Geophysical Union Reference Shelf, 1, 188–103.

Mathias, M. (1974). Alkaline rocks of southern Africa. In *The Alkaline Rocks*, ed. H. Sørensen. NY: Wiley-Interscience, 189–202.

McBirney, A. R. (1975). Differentiation of the Skaergaard intrusion. *Nature*, **253**, 691–694.

McBirney, A. R. (1985). Further considerations of double-diffusive stratification and layering in the Skaergaard Intrusion. *Journal of Petrology*, **26**, 993–1001.

McBirney, A. R., and Nicolas, A. (1997). The Skaergaard layered series: II. Magmatic flow and dynamic layering. *Journal of Petrology* **38**, 569 580.

McBirney, A. R., and Noyes, R. M. (1979). Crystallization and layering of the Skaergaard intrusion. *Journal of Petrology*, **20**, 487–554.

McBirney, A. R., Taylor, H. P., and Armstrong, R. L. (1987). Paricutin re-examined: a classic example of crustal assimilation in calc-alkaline magma. *Contributions to Mineralogy and Petrology*, **95**, 4–20.

McCaffrey, K. J. W., and Petford, N. (1997). Are granitic intrusions scale invariant? *Journal of the Geological Society of London*, **154**, 1–4.

McDonough, W. F., and Sun, S.–s. (1995). The composition of the Earth. *Chemical Geology*, **120**, 223–253.

McGee, K. A., and Jefferson Sutton, A. (1994). Eruptive activity at Mount St Helens, Washington, USA, 1984–1988: a gas geochemistry perspective. *Bulletin of Volcanology*, **56**, 435–446.

McGetchin, T. R., and Ullrich, G. W. (1973). Xenoliths and maars and diatremes with inferences for the Moon, Mars, and Venus. *Journal of Geophysical Research*, **78**, 1833–1853.

McGuire, A. V., and Bohannon, R. G. (1989). Timing of mantle upwelling: evidence for a passive origin for the Red Sea rift. *Journal of Geophysical Research*, **94**, 1677–1682.

McKenzie, D. (1967). Some remarks on heat flow and gravity anomalies. *Journal of Geophysical Research*, **72**, 6261–6273.

McKenzie, D. (1984). The generation and compaction of partially molten rock. *Journal of Petrology*, **25**, 713–765.

McKenzie, D., and Bickle, M. J. (1988). The volume and composition of melt generated by extension of the lithosphere. *Journal of Petrology*, **29**, 625–679.

McKenzie, D., and Brune, J. N. (1972). Melting on fault planes during large earthquakes. *Royal Astronomical Society Geophysical Journal*, **29**, 65–78.

McKenzie, D., and Weiss, N. (1975). Speculations on the thermal and tectonic history of the Earth. *Geophysics Journal of the Royal Astronomical Society*, **42**, 131–174.

McMullin, D. W. A., Berman, R. G., and Greenwood, H. J. (1991). Calibration of the SGAM thermobarometer for pelitic rocks using data from phase-equilibrium experiments and natural assemblages. *The Canadian Mineralogist*, **29**, 889–908.

McNutt, S. R. (2005). Volcanic seismology. *Annual Review of Earth and Planetary Sciences*, **32**, 461–491.

McTaggart, K. C. (1960). The mobility of nuées ardentes. *American Journal of Science*, **258**, 369–382.

Melnik, O., and Sparks, R. S. J. (2005). Controls on conduit magma flow dynamics during lava dome building eruptions. *Journal of Geophysical Research*, **110**, B02209, doi:10.1029/2004JB003183.

Mel'nik, Y. P. (1972). Thermodynamic parameters of compressed gases and metamorphic reactions involving water and carbon dioxide. *Geochemistry International*, **9**, 419–426.

MELT Seismic Team (1998). Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment. *Science*, **280**, 1215–1218.

Menneken, M., Nemchin, A. A., Geisler, T., Pidgeon, R. T., and Wilde, S. A. (2007). Hadean diamonds in zircon from Jack Hills, Western Australia. *Nature*, **448**, 917–920.

Menzies, M. A., and Hawkesworth, C. J., eds. (1987). *Mantle Metasomatism*. London: Academic Press, 477 pp.

Meurer, W. P., and Boudreau, A. E. (1998). Compaction of igneous cumulates: part II. Compaction and the development of igneous foliations. *Journal of Geology* **106**, 293 304.

Meyer, H. O. A. (1979). Kimberlites and the mantle. *Reviews of Geophysics and Space Physics*, **17**, 776–788.

Mitchell, R. H. (1986). *Kimberlites: Mineralogy, Geochemistry, and Petrology*, New York: Plenum Press, 460 pp.

Miyashiro, A. (1961). Evolution of metamorphic belts. *Journal of Petrology*, **2**, 277–311.

Miyashiro, A. (1973). *Metamorphism and Metamorphic Belts*. New York: John Wiley.

Moecher, D. P., Essene, E. J., and Anovitz, L. M. (1988). Calculation and application of clinopyroxene-garnet-plagioclase-quartz geobarometers. *Contributions to Mineralogy and Petrology*, **100**, 92–106.

Molyneux, S. J., and Hutton, D. H. W. (2000). Evidence for significant granite space creation by the ballooning mechanism: the example of the Ardara pluton, Ireland. *Geological Society of America Bulletin*, **112**, 1543–1558.

Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R., and Hung, S.-H. (2004). Finite-frequency tomography reveals a variety of plumes in the mantle. *Science*, **303**, 338–343.

Moorbath, S. (1977). Ages, isotopes and evolution of Precambrian continental crust. *Chemical Geology*, **20**, 151–187.

Moore, G., Vennemann, T., and Carmichael, I. S. E. (1995). Solubility of water in magmas to 2 kbar. *Geology*, **23**, 1099–1102.

Moore, G., Vennemann, T., and Carmichael, I. S. E. (1998). An empirical model for the solubility of water in magmas to 3 kilobars. *American Mineralogist*, **83**, 36–42.

Morgan, D. J., Blake, S., Rogers, N. W., *et al*. (2006). Magma chamber recharge at Vesuvius in the century prior to the eruption of A.D. 79. *Geology*, **34**, 845–848.

Morgan, W. J. (1971). Convective plumes in the lower mantle. *Nature*, **230**, 42–43.

Morris, J. D., Leeman, W. P., and Tera, F. (1990). The subducted component in island arc lavas: constraints from Be isotopes and B-Be systematics. *Nature*, **344**, 31–36.

Morse, S. A. (1969). The Kiglapait layered intrusion, Labrador. *Geological Society of America Memoir*, **112**.

Morse, S. A. (1970). Alkali feldspars with water at 5 kb pressure. *Journal of Petrology*, **11**, 221–253.

Morse, S. A. (1980). *Basalts and Phase Diagrams*. New York: Springer-Verlag.

Morse, S. A. (1982). A partisan review of Proterozoic anorthosites. *American Mineralogist*, **67**, 1087–1100.

Morse, S. A. (1986). Convection in aid of adcumulus growth. *Journal of Petrology*, **27**, 1183–1214.

Muan, A. (1955). Phase equilibria in the system FeO–Fe_2O_3–SiO_2. *Transactions of the American Institute of Mining, Metallurgical and Petroleum Engineers*, **203**, 965–976.

Muan, A. (1958). Phase equilibria at high temperatures in oxide systems involving changes in oxidation states. *American Journal of Science*, **256**, 171–207.

Müller, T., Baumgartner, L. P., Foster, C. T., Jr., and Vennemann, T. W. (2004). Metastable prograde reactions in contact aureoles. *Geology*, **32**, 821–824.

Murakami, M., Hirose, K., Kawamura, K., Sata, N., and Ohishi, Y. (2004). Post-perovskite phase transition in $MgSiO_3$. *Science*, **304**, 855–858.

Myers, J. S. (1975). Cauldron subsidence and fluidizaton: mechanisms of intrusion of the coastal batholith of Peru into its own volcanic ejecta. *Geological Society of America Bulletin*, **86**, 1209–1220.

Myers, J. S. (1993). Precambrian history of the West Australian Craton and adjacent orogens. *Annual Review of Earth and Planetary Sciences* **21**, 453 485.

Nabelek, P. I. (2007). Fluid evolution and kinetics of metamorphic reactions in calc-silicate contact aureoles – from H_2O to CO_2 and back. *Geology*, **35**, 927–930.

Naldrett, A. J. (1969). A portion of the system Fe–S–O between 900 and 1080 °C and its application to sulfide ore magmas. *Journal of Petrology*, **10**, 171–201.

Naldrett, A. J. (1989). *Magmatic Sulfide Deposits*. New York: Oxford University Press, 186 pp.

Naldrett, A. J. (2003). From impact to riches: evolution of geological understanding as seen at Sudbury, Canada. *GSA Today*, **13**, 4–9.

Naslund, H. R., and McBirney, A. R. (1996). Mechanisms of formation of igneous layering. In *Layered Intrusions*, ed. R. G. Cawthorn. Amsterdam: Elsevier, 1–43.

Nataf, H.-C. (2000). Seismic imaging of mantle plumes. *Annual Review of Earth and Planetary Sciences*, **28**, 391–417.

Navrotsky, A., Hon, R., Weill, D. F., and Henry, D. J. (1980). Thermochemistry of glasses and liquids in the systems $CaMgSi_2O_6$–$CaAl_2Si_2O_8$–$NaAlSi_3O_8$, SiO_2–$CaAl_2Si_2O_8$–$NaAlSi_3O8$ and SiO_2–Al_2O_3–CaO–Na_2O. *Geochimica et Cosmochimica Acta*, **44**, 1409–1423.

Nelson, K. D. (1981). A simple thermal mechanical model for mid-ocean ridge topographic variation. *Geophysical Journal of the Royal Astronomical Society*, **65**, 19–30.

Nesbitt, B. E. (1988). Gold deposit continuum: a genetic model for lode Au mineralization in the continental crust. *Geology*, **16**, 1044–1048.

Newman, S., and Lowenstern, J. B. (2002). VolatileCalc: a silicate melt–H_2O–CO_2 solution model written in Visual Basic for Excel. *Computers & Geosciences*, **28**, 597–604.

Newton, R. C. (1995). Simple-system mineral reactions and high-grade metamorphic fluids. *European Journal of Mineralogy*, **7**, 861–881.

Newton, R. C., and Manning, C. E. (2000). Quartz solubility in H_2O-NaCl and H_2O-CO_2 solutions at deep crust-upper mantle pressures and temperatures: 2–15 kbar and 500–900 °C. *Geochimica et Cosmochimica Acta*, **64**, 2993–3005.

Newton, R. C., Smith, J. V., and Windley, B. (1980). Carbonic metamorphism, granulites and crustal growth. *Nature*, **288**, 45–50.

Nicolas, A. (1986). A melt extraction model based on structural studies in mantle peridotites. *Journal of Petrology*, **27**, 999–1022.

Nicolas, A., and Jackson, M. (1982). High temperature dikes in peridotites: origin by hydraulic fracturing. *Journal of Petrology*, **23**, 568–582.

Nicolas, A., and Prinzhofer, A. (1983). Cumulative or residual origin for the transition zone in ophiolites: structural evidence. *Journal of Petrology*, **24**, 188–206.

Nielsen, T. F. D. (2004). The shape and volume of the Skaergaard Intrusion, Greenland: implications for mass balance and bulk composition. *Journal of Petrology*, **45**, 507–530.

Nolet, G., Allen, R., and Zhao, D. (2007). Mantle plume tomography. *Chemical Geology*, **241**, 248–263.

Nordlie, B. E. (1971). The composition of the magmatic gas of Kilauea and its behavior in the near-surface environment. *American Journal of Science*, **271**, 417–463.

Norry, M. J., and Fitton, J. G. (1983). Compositional differences between oceanic and continental basic lavas and their significance. In *Continental Basalts and Mantle Xenoliths*, ed. C. J. Hawkesworth, and M. J. Norry. Nantwich, Cheshire: Shiva Publishing, 5–19.

Norton, D. L., and Dutrow, B. L. (2001). Complex behavior of magma-hydrothermal processes: role of supercritical fluid. *Geochimica et Cosmochimica Acta* **65**, 4009 4017.

Norton, D., and Knapp, R. (1977). Transport phenomena in hydrothermal systems: the nature of porosity. *American Journal of Science*, **277**, 913–936.

Norton, D., and Taylor, H. P. (1979). Quantitative simulation of the hydrothermal systems of crystallizing magmas on the basis of transport theory and oxygen isotope data: an analysis of the Skaergaard intrusion. *Journal of Petrology*, **20**, 421–486.

Obayashi, M., and Fukao, Y. (1997). P and PcP travel time tomography for the core-mantle boundary. *Journal of Geophysical Research*, **102**, 17825–17841.

O'Driscoll, B., Donaldson, C. H., Troll, V. R., Jerram, D. A., and Emeleus, C. H. (2007). An origin for harrisitic and granular olivine in the Rum layered suite, NW Scotland: a crystal size distribution study. *Journal of Petrology*, **48**, 253–270.

Oftedahl, C. (1960). Permian rocks and structures of the Oslo region. In *Geology of Norway*, ed. O. Holtedahl. *Norges Geologiske Undersøkelse*, **208**, 298–343.

Ogata, A., and Banks, R. B. (1961). A solution of the differential equations of longitudinal dispersion in porous media. *U.S. Geological Survey Professional Paper*, **411-A**.

Ohtani, E., and Maeda, M. (2001). Density of basaltic melt at high pressure and stability of melt at the base of the lower mantle. *Earth and Planetary Science Letters*, **193**, 69–75.

Okubo, P. G., Benz, H. M., and Chouet, B. A. (1997). Imaging the crustal magma sources beneath Mauna Loa and Kilauea volcanoes, Hawaii. *Geology*, **25**, 867–870.

Oliver, G. J. H., Chen, F., Buchwaldt, R., and Hegner, E. (2000). Fast tectonometmorphism and exhumation in the type area of the Barrovian and Buchan zones. *Geology*, **28**, 459–462.

Oliver, N. H. S., Dipple, G. M., Cartwright, I., and Schiller, J. (1998). Fluid flow and metasomatism in the genesis of the amphibolite-facies, pelite-hosted Kanmantoo copper deposit, south Australia. *American Journal of Science*, **298**, 181–218.

O'Neill, H. St. C. (1981). The transition between spinel lherzolite and garnet lherzolite, and its use as a geobarometer. *Contributions to Mineralogy and Petrology*, **77**, 185–194.

O'Nions, R. K. (1984). Isotopic abundances relevant to the identification of magma sources. *Philosophical Transactions of the Royal Society of London*, **A310**, 591–603.

O'Nions, R. K., Hamilton, P. J., and Evensen, N. M. (1977). Variations in $^{143}Nd/^{144}Nd$ and $^{87}Sr/^{86}Sr$ ratios in oceanic basalts. *Earth and Planetary Science Letters*, **39**, 13–22.

O'Nions, R. K., Hamilton, P. J., and Evensen, N. M. (1980). The chemical evolution of the Earth's mantle. *Scientific American*, **242**, 120–133.

Orville, P. M. (1962). Alkali metasomatism and the feldspars. *Norsk Geologisk Tidsskrift*, **42**, 283–316.

Osberg, P. H. (1971). An equilibrium model for Buchan-type metamorphic rocks, south-central Maine. *American Mineralogist*, **56**, 570–586.

Osborn, E. F., and Schairer, J. F. (1941). The ternary system pseudowollastonite–akermanite–gehlenite. *American Journal of Science*, **239**, 715–763.

Oxburgh, E. R. (1980). Heat flow and magma genesis. In *Physics of Magmatic Processes*, ed. R. B. Hargraves, Princeton, NJ: Princeton University Press, 161–200.

Oxburgh, E. R., and McRae, T. (1984). Physical constraints on magma contamination in the continental crust: an example, the Adamello complex. *Philosophical Transactions of the Royal Society of London* **A 310**, 457 472.

Oxburgh, E. R., and Turcotte, D. L. (1971). Origin of paired metamorphic belts and crustal dilation in island arc regions. *Journal of Geophysical Research*, **76**, 1315–1327.

Oxtoby, D. W., Gillis, H. P., and Nachtrieb, N. H. (1999). *Principles of Modern Chemistry*, 4th edn. New York: Saunders College.

Page, F. Z., Armstrong, L. S., Essene, E. J., and Mukasa, S. B. (2007). Prograde and retrograde history of the Junction School eclogite, California, and an evaluation of garnet–phengite–clinopyroxene thermobarometry. *Contributions to Mineralogy and Petrology*, **153**, 533–555.

Palin, J. M. (1992). Stable isotope studies of regional metamorphism in the Wepawaug Schist, Connecticut. Ph.D. thesis. New Haven, CT: Yale University.

Papale, P., Moretti, R., and Barbato, D. (2006). The compositional dependence of the saturation surface of H_2O+CO_2 fluids in silicate melts. *Chemical Geology*, **229**, 78–95.

Parsons, B. A. (1982). Causes and consequences of the relation between area and age of the ocean floor. *Journal of Geophysical Research*, **87**, 289–302.

Parsons, B., and McKenzie, D. (1978). Mantle convection and the thermal structure of the plates. *Journal of Geophysical Research*, **83**, 4485–4496.

Parsons, B., and Sclater, J. G. (1977). An analysis of the variation of ocean floor bathymetry and heat flow with age. *Journal of Geophysical Research*, **82**, 803–827.

Passchier, C. W., and Trouw, R. A. J. (1996). *Microtectonics*. Berlin: Springer-Verlag.

Patterson, E. M. (1951). A petrochemical study of the Tertiary lavas of northeast Ireland. *Geochimica et Cosmochimica Acta*, **2**, 283–299.

Pattison, D. R. M. (2001). Instability of Al_2SiO_5 "triple-point" assemblages in muscovite + biotite + quartz-bearing metapelites, with implications. *American Mineralogist*, **86**, 1414–1422.

Pattison, D. R. M., and Newton, R. C. (1989). Reversed experimental calibration of the garnet-clinopyroxene Fe–Mg exchange thermometer. *Contributions to Mineralogy and Petrology*, **101**, 87–103.

Peacock, S. A. (1990). Fluid processes in subduction zones. *Science*, **248**, 329–337.

Peacock, S. M. (1987). Thermal effects of metamorphic fluids in subduction zones. *Geology*, **15**, 1057–1060.

Pearce, J. A., and Cann, J. R. (1973). Tectonic setting of basic volcanic rocks determined using trace element analyses. *Earth and Planetary Science Letters*, **19**, 920–300.

Pearce, J. A., and Parkinson, I. J. (1993). Trace element models for mantle melting: application to volcanic arc petrogenesis. *Geological Society of London Special Publication*, **76**, 373–403.

Pearce, J. A., and Peate, D. W. (1995). Tectonic implications of the composition of volcanic arc magmas. *Annual Review of Earth and Planetary Sciences*, **23**, 251–285.

Peck, D. L., Moore, J. G., and Kojima, G. (1964). Temperatures in the crust and melt of Alae lava lake, Hawaii, after the August 1963 eruption of Kilauea volcano – a preliminary report. *U.S. Geological Survey Professional Paper* **501D**, 1–7.

Peck, W. H., King, E. M., and Valley, J. W. (2000). Oxygen isotope perspective on Precambrian crustal growth and maturation. *Geology*, **28**, 363–366.

Peltier, W. R., and Andrews, J. T. (1976), Glacial isostatic adjustment, I: the forward problem. *Geophysics Journal of the Royal Astronomical Society*, **46**, 605–646.

Penniston-Dorland, S. C., and Ferry, J. M. (2006). Development of spatial variations in reaction progress during regional metamorphism of micaceous carbonate rocks, Northern New England. *American Journal of Science*, **306**, 475–524.

Peppard, B. T., Steele, I. M., Davis, A. M., Wallace, P. J., and Anderson, A. T. (2001). Zoned quartz phenocrysts from the rhyolitic Bishop Tuff. *American Mineralogist*, **86**, 1034–1052.

Perkins, M. E., and Nash, B. P. (2002). Explosive volcanism of the Yellowstone hotspot: the ash fall tuff record. *Geological Society of America Bulletin*, **114**, 367–381.

Petford, N. (2003). Rheology of granitic magmas during ascent and emplacement. *Annual Reviews of Earth and Planetary Science*, **31**, 399–427.

Petford, N., and Clemens, J. D. (2000). Granites are not diapiric! *Geology Today*, Sept.-Oct., 180–184.

Philbrick, S. S. (1936). The contact metamorphism of the Onawa pluton, Piscataquis County, Maine. *American Journal of Science*, **231**, 1–40.

Philpotts, A. R. (1964). Origin of pseudotachylites. *American Journal of Science*, **262**, 1008–1035.

Philpotts, A. R. (1968). Igneous structures and mechanism of emplacement of Mount Johnson, a Monteregian intrusion, Quebec. *Canadian Journal of Earth Science*, **5**, 1131–1137.

Philpotts, A. R. (1972). Density, surface tension, and viscosity of the immiscible phase in a basic alkaline magma. *Lithos*, **5**, 1–18.

Philpotts, A. R. (1974). The Monteregian Province. In *The Alkaline Rocks*, ed. H. Sørensen. New York: Wiley-Interscience, 293–310.

Philpotts, A. R. (1976). Silicate liquid immiscibility: its probable extent and petrogenetic significance. *American Journal of Science*, **276**, 1147–1177.

Philpotts, A. R. (1981). A model for the generation of massif-type anorthosites. *Canadian Mineralogist*, **19**, 233–253.

Philpotts, A. R. (1982). Compositions of immiscible liquids in volcanic rocks. *Contributions to Mineralogy and Petrology*, **80**, 201–218.

Philpotts, A. R. (1989). *Petrography of Igneous and Metamorphic Rocks*. Prospect Heights, IL: Waveland Press, 178 pp.

Philpotts, A. R. (1998). Nature of a flood-basalt-magma reservoir based on the compositional variation in a single flood-basalt flow and its feeder dike in the Mesozoic Hartford Basin, Connecticut. *Contributions to Mineralogy and Petrology*, **133**, 69–82.

Philpotts, A. R., and Asher, P. M. (1993). Wallrock melting and reaction effects along the Higganum diabase dike in Connecticut: contamination of a continental flood basalt feeder. *Journal of Petrology*, **34**, 1029–1058.

Philpotts, A. R., and Asher, P. M. (1994). Magmatic flow-direction indicators in a giant diabase feeder dike, Connecticut. *Geology*, **22**, 363–366.

Philpotts, A. R., and Doyle, C. D. (1983). Effect of magma oxidation state on the extent of silicate liquid immiscibility in a tholeiitic basalt. *American Journal of Science*, **283**, 967–986.

Philpotts, A. R., and Lewis, C. L. (1987). Pipe vesicles – an alternate model for their origin. *Geology*, **15**, 971–974.

Philpotts, A. R., and Martello, A. (1986). Diabase feeder dikes for the Mesozoic basalts in southern New England. *American Journal of Science*, **286**, 105–126.

Philpotts, A. R., and Philpotts, D. E. (2005). Crystal-mush compaction in the Cohassett flood-basalt flow, Hanford, Washington. *Journal of Volcanology and Geothermal Research*, **145**, 2005.

Philpotts, A. R., and Philpotts, D. E. (2007). Upward and downward flow in a camptonite dike as recorded by deformed vesicles and the anisotropy of magnetic susceptibility (AMS). *Journal of Volcanology and Geothermal Research* **161**, 81 94.

Philpotts, A. R., Shi, J., and Brustman, C. M. (1998). Role of plagioclase crystal chains in the differentiation of partly crystallized basaltic magma. *Nature*, **395**, 343–346.

Philpotts, A. R., Brustman, C. M., Shi, J., Carlson, W. D., and Denison, C. (1999). Plagioclase-chain networks in slowly cooled basaltic magma. *American Mineralogist*, **84**, 1819–1829.

Pierce, K. L., and Morgan, L. A. (1992). The rack of the Yellowstone hot spot: volcanism, faulting, and uplift. In *Regional Geology of Eastern and Western Wyoming*, ed P. K. Link, M. A. Kuntz, and L. B. Platt. *Geological Society of America Memoir*, **179**, 1–53.

Pitcher, W. S. (1997). *The Nature and Origin of Granite*. New York: Chapman & Hall.

Plank, T., and Langmuir, C. H. (1988). An evaluation of the global variations in the major element chemistry of arc basalts. *Earth and Planetary Science Letters*, **90**, 290–300.

Plank, T., and Langmuir, C. H. (1993). Tracing trace elements from sediment input to volcanic output at subduction zones. *Nature*, **362**, 739–743.

Platt, J. P. (1986). Dynamics of orogenic wedges and the uplift of high-pressure metamorphic rocks. *Geological Society of America Bulletin*, **97**, 1037–1053.

Poe, B. T., Romano, C., Liebske, C., *et al*. (2006). High-temperature viscosity measurements of hydrous albite liquid using in-situ falling-sphere viscometry at 2.5 GPa. *Chemical Geology*, **229**, 2–9.

Pollack, H. N., and Chapman, D. S. (1977). On the regional variation of heat flow, geotherms, and lithospheric thickness. *Tectonophysics*, **38**, 279–296.

Pollack, H. N., Hurter, S. J., and Johnson, J. R. (1993). Heat flow from the Earth's interior: analysis of the global data set. *Reviews of Geophysics*, **31**, 267–280.

Pollard, D. D. (1973). Derivation and evaluation of a mechanical model for sheet intrusions. *Tectonophysics*, **19**, 233–269.

Pollard, D. D., and Johnson, A. M. (1973). Mechanics of growth of some laccolithic intrusions in the Henry Mountains, Utah: II. *Tectonophysics*, **18**, 311–354.

Pollard, D. D., Muller, O. H., and Dockstader, D. R. (1975). The form and growth of fingered sheet intrusions. *Geological Society of America Bulletin*, **86**, 351–363.

Polyak, B. G., and Smirnov, Y. B. (1968). Relationship between terrestrial heat flow and the tectonics of continents. *Geotectonics*, **4**, 205–213.

Powell, R., Holland, T. J. B., and Worley, B. (1998). Calculating phase diagrams involving solid solutions via non-linear equations, with examples using THERMOCALC. *Journal of Metamorphic Geology*, **16**, 577–588.

Presnall, D. C., and Bateman, P. C. (1973). Fusion relations in the system $NaAlSi_3O_8–CaAl_2Si_2O_8–KAlSi_3O_8–SiO_2–H_2O$ and generation of granitic magmas in the Sierra Nevada batholith. *Geological Society of America Bulletin*, **84**, 3181–3202.

Presnall, D. C., Dixon, S. A., Dixon, J. R., *et al*. (1978). Liquidus phase relations on the join diopside-forsterite-anorthite from 1 atm to 20 kbar: their bearing on the generation and crystallization of basaltic magmas. *Contributions to Mineralogy and Petrology*, **66**, 203–220.

Presnall, D. C., Dixon, J. R., O'Donnell, T. H., and Dixon, S. A. (1979). Generation of mid-ocean ridge tholeiites. *Journal of Petrology*, **20**, 3–35.

Putirka, K. (2005). Mantle potential temperatures at Hawaii, Iceland and the mid-ocean ridge system, as inferred from olivine phenocrysts: evidence for thermally-driven mantle plumes. *Geochemistry, Geophysics, Geosystems*. Doi:10.1029/2005GCGC000915.

Putnis, A. and Holland, T. J. B. (1986). Sector trilling in cordierite and equilibrium overstepping in metamorphism. *Contributions to Mineralogy and Petrology*, **93**, 265–272.

Pyke, D. R., Naldrett, A. J., and Eckstrand, O. R. (1973). Archean ultramafic flows in Munro Township, Ontario. *Geological Society of America Bulletin*, **84**, 955–978.

Rabinowicz, M., and Vigneresse, J. L. (2004). Melt segregation under compaction and shear channeling: application to granitic magma segregation in a continental crust. *Journal of Geophysical Research*, **109**, 4407.

Ramberg, H. (1981). *Gravity, Deformation and the Earth's Crust*, 2nd edn. London: Academic Press, 452 pp.

Ramberg, I. B. (1976). Gravity interpretation of the Oslo Graben and associated igneous rocks. *Norges Geologiske Undersøkelse*, **325**, 194 pp.

Ramsay, J. (1980). The crack-seal mechanism of rock deformation. *Nature*, **284**, 135–139.

Randolph, A. D., and Larson, M. A. (1971). *Theory of Particulate Processes*. New York: Academic Press, 251 pp.

Rankin, A. H., and Le Bas, M. J. (1974). Liquid immiscibility between silicate and carbonate melts in naturally occurring ijolite magma. *Nature*, **250**, 206–209.

Ravna, E. J. K. (2000). The garnet-clinopyroxene Fe^{2+}-Mg geothermometer: an updated calibration. *Journal of Metamorphic Geology*, **18**, 211–219.

Ravna, E. J. K., and Terry, M. P. (2004). Geothermobarometry of UHP and HP eclogites and schists – an evaluation of equilibria among garnet–clinopyroxene–kyanite–phengite–coesite/quartz. *Journal of Metamorphic Geology*, **22**, 579–592.

Read, H. H. (1923). Petrology of the Arnage district. *Journal of the Geological Society London*, **79**, 447–486.

Read, H. H. (1957). *The Granite Controversy*. London: Murby.

Reagan, M. K., Sims, K. W. W., Erich, J., *et al*. (2003). Time-scales of differentiation from mafic parents to rhyolite in the North American continental arcs. *Journal of Petrology*, **44**, 1703–1726.

Reches, Z., and Fink, J. (1988). The mechanism of intrusion of the Inyo dike, Long Valley caldera, California. *Journal of Geophysical Research*, **93**, 4321–4334.

Reidel, S. P., Tolan, T. L., and Beeson, M. H. (1994). Factors that influenced the eruptive and emplacement history of flood basalt flows: a field guide to selected vents and flows of the Columbia River Basalt Group. In *Geologic Field Trips in the Pacific Northwest*, ed. D. A. Swanson and R. A. Haugerud. Seattle, WA: University of Washington, **1**, 1–18.

Reiners, P. W., Nelson, B. K., and Ghiorso, M. S. (1995). Assimilation of felsic crust by basaltic magma: thermal limits and extents of crustal contamination of mantle-derived magmas. *Geology*, **23**, 563–566.

Reynolds, D. L. (1954). Fluidization as a geological process and its bearing on the problem of intrusive granites. *American Journal of Science*, **252**, 577–613.

Rhodes, J. M., Dungan, M. A., Blanchard, D. P., and Long, P. E. (1979). Magma mixing at mid-ocean ridges: evidence from basalts drilled near 22°N on the Mid-Atlantic Ridge. *Tectonophysics*, **55**, 35–62.

Rhodes, R. C. (1975). New evidence for impact origin of the Bushveld Complex, South Africa. *Geology*, **3**, 549–554.

Rice, A. H. N., and Mitchell, J. I. (1991). Porphyroblast textural sector-zoning and matrix displacement. *Mineralogical Magazine*, **55**, 379–396.

Rice, J. M., and Ferry, J. M. (1982). Buffering, infiltration, and the control of intensive variables during metamorphism. In

Characterization of Metamorphism through Mineral Equilibria, ed. J. M. Ferry. vol. 10 of *Reviews in Mineralogy*. Washington, DC: Mineralogical Society of America 263–326.

Richard, G., Bercovici, D., and Karato, S. (2006). Slab dehydration in the Earth's mantle transition zone. *Earth and Planetary Science Letters*, **251**, 156–167.

Richardson, S. W., Gilbert, M. C., and Bell, P. M. (1969). Experimental determination of kyanite-andalusite and andalusite-sillimanite equilibria: the aluminum silicate triple point. *American Journal of Science*, **267**, 259–272.

Richter, D. H., Eaton, J. P., Murata, K. J., Ault, W. U., and Krivoy, H. L. (1970). Chronological narrative of the 1959–60 eruption of Kilauea Volcano, Hawaii. *U.S. Geological Survey Professional Paper*, **537-E**, 73 pp.

Richter, F. M., and McKenzie, D. (1981). On some consequences and possible causes of layered mantle convection. *Journal of Geophysical Research*, **86**, 6133–6142.

Ridley, J. and Thompson, A. B. (1986). The role of mineral kinetics in the development of metamorphic microtextures. In *Fluid-rock Interactions During Metamorphism*, ed. J. V. Walther, and B. J. Wood. New York: Springer, 154–193.

Righter, K. (2003). Metal-silicate partitioning of siderophile elements and core formation in the early Earth. *Annual Review of Earth and Planetary Sciences*, **31**, 135–174.

Ringwood, A. E. (1975). *Composition and Petrology of the Earth's Mantle*. New York: McGraw-Hill, 618 pp.

Roberts, J. L. (1970). The intrusion of magma into brittle rocks. In *Mechanisms of Igneous Intrusion*, ed. G. Newall, and H. Rast. *Geological Journal* Special Issue, 2. Liverpool, Lancashire: Liverpool Geological Society, 287–338.

Robertson, P. B., Dence, M. R., and Vos, M. A. (1968). Deformation in rock-forming minerals from Canadian craters. In *Shock Metamorphism of Natural Materials*, ed. B. M. French, and N. M. Short. Baltimore, MD: Mono Book Corp., 433–452.

Robie, R. A., Hemingway, B. S., and Fisher, J. R. (1978). Thermodynamic properties of minerals and related substances at 298.15 K and 1 bar (10^5 Pascals) pressure and at high temperatures. *U.S. Geological Survey Bulletin*, **1452**, 456 pp.

Robinson, P., Jaffe, W., Ross, M., and Klein, C., Jr. (1971). Orientation of exsolution lamellae in clinopyroxenes and clinoamphiboles: consideration of optimal phase boundaries. *American Mineralogist*, **56**, 68–94.

Robinson, P., Hollocher, K. T., Tracy, R. J., and Dietsch, C. W. (1982). High grade Acadian regional metamorphism in south-central Massachusetts. In *Guidebook for Fieldtrips in Connecticut and South Central Massachusetts*, ed. R. J. Joesten, and S. S. Quarrier. New England Intercollegiate Geological Conference, 74th Annual Meeting. Hartford, CT: State Geological and Natural History Survey of Connecticut, Guidebook No. 5, 289–339.

Roedder, E. (1951). Low temperature liquid immiscibility in the system K_2O–FeO–Al_2O_3–SiO_2. *American Mineralogist*, **36**, 282–286.

Roedder, E. (1979). Silicate liquid immiscibility in magmas. In *The Evolution of the Igneous Rocks, Fiftieth Anniversary Perspectives*, ed. Yoder, H. S., Jr. Princeton, NJ: Princeton University Press, 15–57.

Roeder, P. L., and Emslie, R. F. (1970). Olivine liquid equilibrium. *Contributions to Mineralogy and Petrology*, **29**, 275–289.

Roman, D. C., and Cashman, K. V. (2006). The origin of volcano-tectonic earthquake swarms. *Geology*, **34**, 457–460.

Romanowicz, B. (2003). Global mantle tomography. *Annual Review of Earth and Planetary Sciences* **31**, 303 328.

Rosenfeld, J. L. (1968). Garnet rotations due to the major Paleozoic deformations in southeast Vermont. In *Studies in Appalachian Geology, Northern and Maritime*, ed. E-an Zen, W. S. White, J. B. Hadley, and J. B. Thompson, Jr. New York: Wiley-Interscience, 185–202.

Rosenfeld, J. L. (1970). Rotated garnets in metamorphic rocks. *Geological Society of America, Special Paper*, **129**, 105 pp.

Ross, M. E. (1986). Flow differentiation, phenocryst alignment, and compositional trends within a dolerite dike at Rockport, Massachusetts. *Geological Society of America Bulletin*, **97**, 232–240.

Rowley, D. B. (2002). Rate of plate creation and destruction: 180 Ma to present. *Geological Society of America Bulletin*, **114**, 927–933.

Rubatto, D., and Hermann, J. (2001). Exhumation as fast as subduction? *Geology*, **29**, 3–6.

Rubie, D. C. (1998). Disequilibrium during metamorphism: the role of nucleation kinetics. In *What Drives Metamorphism and Metamorphic Reactions?*, ed. P. J. Treloar, and P. J. O'Brien. Geological Society, London, Special Publications, 138. 199–214.

Rubin, A. M. (1995). Propagation of magma-filled cracks. *Annual Review of Earth and Planetary Sciences*, **23**, 287–336.

Rudnick, R. L., and Fountain, D. M. 1995. Nature and composition of the continental crust: a lower crustal perspective. *Reviews of Geophysics*, **33**, 267–309.

Rumble, D., Ferry, J. M., Hoering, T. C., and Boucot, A. J. (1982). Fluid flow during metamorphism at the Beaver Brook fossil locality, New Hampshire. *American Journal of Science*, **282**, 886–919.

Ryan, M. P. (1993). Neutral buoyancy and the structure of mid-ocean ridge magma reservoirs. *Journal of Geophysical Research*, **98**, 321–328.

Ryan, M. P. (1994). Neutral-buoyancy controlled magma transport and storage in mid-ocean ridge magma reservoirs and their sheeted-dike complex: a summary of basic relationships. In *Magmatic Systems*, ed. M. P. Ryan. San Diego, CA: Academic Press, 97–138.

Ryan, M. P., Koyanagi, R. Y., and Fiske, R. S. (1981). Modeling the three dimensional structure of macroscopic magma transport systems: application to Kilauea Volcano, Hawaii. *Journal of Geophysical Research*, **86**, 7111–7129.

Rye, R. O., Schuiling, R. D., Rye, D. M., and Jansen, J. B. H. (1976). Carbon, hydrogen, and oxygen isotope studies of the regional metamorphic complex on Naxos, Greece. *Geochimica et Cosmochimica Acta*, **40**, 1031–1049.

Ryerson, F. J., and Hess, P. C. (1978). Implication of liquid-liquid distribution coefficients to mineral-liquid partitioning. *Geochimica et Cosmochimica Acta*, **42**, 921–932.

Sanchez-Navas, A. (1999). Sequential kinetics of a muscovite-out reaction: a natural example. *American Mineralogist*, **84**, 1270–1286.

Sasaki, S., and Nakazawa, K. (1986). Metal silicate fractionation in the growing Earth: energy source for the terrestrial magma ocean. *Journal of Geophysical Research*, **91**, 9231–9238.

Saunders, A. D. (2005). Large igneous provinces: origin and environmental consequences. *Elements*, **1**, 259–263.

Sautter, V., Jaoul, O., and Abel, F. (1988). Aluminum diffusion in diopside using the $^{27}Al(p, \gamma)^{28}Si$ nuclear reaction: preliminary results. *Earth and Planetary Science Letters*, **89**,109–114.

Saxena, S. K. (1973). *Thermodynamics of rock-forming crystalline solutions*. New York: Springer-Verlag, 188 pp.

Scarfe, C. M., Luth, W. C., and Tuttle, O. F. (1966). An experimental study bearing on the absence of leucite in plutonic rocks. *American Mineralogist*, **51**, 726–735.

Schairer, J. F., and Bowen, N. L. (1955). The system K_2O–Al_2O_3–SiO_2. *American Journal of Science*, **253**, 681–746.

Schairer, J. F., and Bowen, N. L. (1956). The system Na_2O–Al_2O_3–SiO_2. *American Journal of Science*, **254**, 129–195.

Schairer, J. F., and Yoder, H. S., Jr. (1960). The nature of residual liquids from crystallization, with data on the system nepheline–diopside–silica. *American Journal of Science*, **258A**, 273–283.

Schairer, J. F., and Yoder, H. S., Jr. (1961). Crystallization in the system nepheline–forsterite–silica at one atmosphere pressure. *Carnegie Institution of Washington Yearbook*, **60**, 141–144.

Schairer, J. F., and Yoder, H. S., Jr. (1964). Crystal and liquid trends in simplified alkali basalts. *Carnegie Institution of Washington Yearbook*, **63**, 65–74.

Schmidt, M. W. (2003). Generation of mobile components during subduction of oceanic crust. In *The Crust*, ed R. L. Rudnick, vol. 3 of *Treatise on Geochemistry*, ed. H. D. Holland, and K. K. Turekian. Amsterdam: Elsevier, 567–591.

Schramke, J. A., Kerrick, D. M., and Lasaga, A. C. (1987). The reaction muscovite + quartz = andalusite + K-feldspar + water. Part I. Growth kinetics and mechanism. *American Journal of Science*, **287**, 517–559.

Sclater, J. G., Jaupart, C., and Galson, D. (1980). The heat flow through oceanic and continental crust and the heat loss of the Earth. *Reviews of Geophysics and Space Physics*, **18**, 269–311.

Sclater, J. G., Parsons, B., and Jaupart, C. (1981). Oceans and continents: similarities and differences in the mechanisms of heat loss. *Journal of Geophysical Research*, **86**, 11535–11552.

Scott, D. R., and Stevenson, D. J. (1986). Magma ascent by porous flow. *Journal of Geophysical Research*, **91**, 9283–9296.

Self, S., Thordarson, T., Keszthelyi, L., *et al*. (1996). A new model for the emplacement of Columbia River basalts as large, inflated pahoehoe lava flow fields. *Geophysical Research Letters*, **23**, 2689–2692.

Self, S., Thordarson, T., and Widdowson, M. (2005). Gas flux from flood basalt eruptions. *Elements*, **1**, 283–287.

Selig, F. (1965). A theoretical prediction of salt-dome patterns. *Geophysics*, **30**, 633–643.

Selverstone, J. (1985). Petrologic constraints on imbrication, metamorphism, and uplift in the SW Tauern Window, eastern Alps. *Tectonics*, **4**, 687–704.

Selverstone, J. and Spear, F. S. (1985). Metamorphic *P–T* paths from pelitic schists and greenstones from the south-west Tauern Window, eastern Alps. *Journal of Metamorphic Geology*, **3**, 439–465.

Selverstone, J., Spear, F. S., Franz, G., and Morteani, G. (1984). High-pressure metamorphism in the SW Tauern Window, Austria: *P–T* paths from hornblende-kyanite-staurolite schists. *Journal of Petrology*, **25**, 501–531.

Shand, S. J. (1917). The pseudotachylyte of Parijs (Orange Free State), and its relation to "trap-shotten-gneiss" and "flinty-crush-rock". *Geological Society of London Quarterly Journal*, **72**, 198–221.

Sharp, Z. D. (1995). Oxygen isotope geochemistry of the Al_2SiO_5 polymorphs. *American Journal of Science*, **295**, 1058–1076.

Sharp, Z. D. (2006). *Stable Isotope Geochemistry*. Englewood Cliffs, NJ: Prentice-Hall.

Sharp, Z. D., Essene, E. J., and Hunziker, J. C. (1993). Stable isotope geochemistry and phase equilibria of coesite-bearing whiteschists, Dora Maira Massif, western Alps. *Contributions to Mineralogy and Petrology*, **114**, 1–12.

Shaw, H. R. (1965). Comments on viscosity, crystal settling, and convection in granitic magmas. *American Journal of Science*, **263**, 120–152.

Shaw, H. R. (1969). Rheology of basalt in the melting range. *Journal of Petrology*, **10**, 510–535.

Shaw, H. R. (1980). The fracture mechanism of magma transport from the mantle to the surface. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 201–264.

Shaw, H. R., and Swanson, D. A. (1970). Eruption and flow rates of flood basalts. In *Proceedings, 2nd Columbia River Basalts Symposium*, ed. E. H. Gilmour and D. Stradling. Cheney, WA: Eastern Washington State College Press, 271–299.

Shaw, H. R., Hamilton, M. S., and Peck, D. L. (1977). Numerical analysis of lava lake cooling models: I. Description of the method. *American Journal of Science*, **277**, 384–414.

Shaw, J. E., Baker, J. A., Menzies, M. A., Thirlwall, M. F., and Ibrahim, K. M. (2003). Petrogenesis of the largest intraplate volcanic field on the Arabian plate (Jordan): a mixed lithosphere–asthenosphere source activated by lithospheric extension. *Journal of Petrology*, **44**, 1657–1679.

Sheridan, M. F. (1979). Emplacement of pyroclastic flows: a review. In *Ash-Flow Tuffs*, ed. C. E. Chapin and W. E. Elston. *Geological Society of America Special Paper*, **180**, 125–138.

Sheridan, M. F., and Wohletz, K. H. (1983). Hydrovolcanism: basic considerations and review. *Journal of Volcanology and Geothermal Research*, **17**, 1–29.

Shi, P., and Saxena, S. K. (1992). Thermodynamic modeling of the C-H-O-S fluid system. *American Mineralogist*, **77**, 1038–1049.

Shirey, S. B., and Walker, R. J. (1998). The Re–Os isotope system in cosmochemistry and high-temperature geochemistry. *Annual Review of Earth and Planetary Sciences*, **26**, 423–500.

Shirley, D. N. (1987). Differentiation and compaction in the Palisades Sill, New Jersey. *Journal of Petrology*, **28**, 835–865.

Short, N. M. (1966). Effects of shock pressures from a nuclear explosion on mechanical and optical properties of granodiorite. *Journal of Geophysical Research*, **71**, 1195–1215.

Sibson, R. H., McMoore, J., and Rankin, R. H. (1975). Seismic pumping – a hydrothermal fluid transport mechanism. *Journal of the Geological Society, London*, **131**, 653–659.

Sigurdsson, H., and Sparks, S. (1978). Lateral magma flow within rifted Icelandic crust. *Nature*, **274**, 126–130.

Sigurdsson, H., Cashdollar, S. and Sparks, S. R. J. (1982). The eruption of Vesuvius in A.D. 79: reconstruction from historical and volcanological evidence. *American Journal of Archaeology*, **86**, 39–51.

Silver, L., and Stolper, E. M. (1989). Water in albitic glass. *Journal of Petrology*, **30**, 667–709.

Silver, P. G., Carlson, W., and Olson, P. (1988). Deep slabs, geochemical heterogeneity, and the large-scale structure of mantle convection: investigation of an enduring paradox. *Annual Review of Earth and Planetary Sciences*, **16**, 477–541.

Simpson, C. (1985). Deformation of granitic rocks across the brittle-ductile transition. *Journal of Structural Geology*, **7**, 503–511.

Sinton, J. M., and Detrick, R. S. (1992). Mid-ocean ridge magma chambers. *Journal of Geophysical Research*, **97**, 197–216.

Sisson, V. B., and Hollister, L. S. (1988). Low-pressure facies series metamorphism in an accretionary sedimentary prism, southern Alaska. *Geology* **16**, 358 361.

Skelton, A. D. L., Graham, C. M., and Bickle, M. J. (1995). Lithological and structural controls on regional 3-D fluid flow patterns during greenschist facies metamorphism of the Dalradian of the SW Scottish Highlands. *Journal of Petrology*, **36**, 563–586.

Skinner, E. M. W., and Marsh, J. S. (2004). Distinct kimberlite classes with contrasting eruption processes. *Lithos*, **76**, 183–200.

Skippen, G. B. (1974). An experimental model for low pressure metamorphism of siliceous dolomitic marble. *American Journal of Science*, **274**, 487–509.

Sleep, N. H. (1992). Hotspot volcanism and mantle plumes. *Annual Review of Earth and Planetary Sciences*, **20**, 119–143.

Smith, R. B., and Christiansen, R. L. (1980). Yellowstone Park as a window on the Earth's interior. *Scientific American*, **242**, 84–95.

Smith, R. L. (1979). Ash-flow magmatism. In *Ash-flow Tuffs*, ed. C. E. Chapin and W. E. Elston. *Geological Society of America Special Paper*, **180**, 5–27.

Smith, R. L., and Bailey, R. A. (1968). Resurgent cauldrons. *Geological Society of America Memoir*, **116**, 613–663.

Snoeyenbos, D. R., Williams, M. L., and Hanmer, S. (1995). Archean high-pressure metamorphism in the western Canadian Shield. *European Journal of Mineralogy*, **7**, 1251–1272.

Solomatov, V. S. (2000). Fluid dynamics of terrestrial magma ocean. In *Origin of the Earth and Moon*, ed. R. M. Canup, and K. Righter. Tucson, AZ: University of Arizona Press, 323–338.

Solomon, S. C. (1979). Formation, history and energetics of cores in the terrestrial planets. *Physics of the Earth and Planetary Interiors*, **19**, 168–182.

Sørensen, H. (ed.) (1974a). *The Alkaline Rocks*. London: John Wiley, 622 pp.

Sørensen, H. (1974b). Alkali syenites, feldspathoidal syenites and related lavas. In *The Alkaline Rocks*, ed. H. Sørensen. London: John Wiley, 22–52.

Sorensen, S. S., and Grossman, J. N. (1989). Enrichment of trace elements in garnet-amphibolites from a paleo-subduction zone: Catalina Schist, Southern California. *Geochimica et Cosmochimica Acta*, **53**, 3155–3177.

Snyder, D., and Tait, S. (1995). Replenishment of magma chambers: comparison of fluid-mechanic experiments with field relations. *Contributions to Mineralogy and Petrology*, **122**, 230–240.

Spandler, C., Mavrogenes, J., and Hermann, J. (2007). Experimental constraints on element mobility from subducted sediments using high-P synthetic fluid/melt inclusions. *Chemical Geology*, **239**, 228–249.

Sparks, R. S. J. (1978). The dynamics of bubble formation and growth in magmas: a review and analysis. *Journal of Volcanology and Geothermal Research*, **3**, 1–37.

Sparks, R. S. J. (1986). The role of crustal contamination in magma evolution through geological time. *Earth and Planetary Science Letters*, **78**, 211–223.

Sparks, R. S. J. (1990). Crystal capture, sorting, and retention in convecting magma: discussion and reply. *Geological Society of America Bulletin*, **102**, 847–850.

Sparks, R. S. J., and Marshall, L. (1986). Thermal and mechanical constraints on mixing between mafic and silicic magmas. *Journal of Volcanology and Geothermal Research*, **29**, 99–124.

Sparks, R. S. J., and Young, S. R. (2002). The eruption of Soufrière Hills volcano, Montserrat (1995–1999): overview of scientific results. In *The Eruption of the Soufrière Hills Volcano, Montserrat From 1995 to 1999*, ed. T. H. Druitt and B. P. Kokelaar. Geological Society, London, Memoirs, **21**, 45 69.

Sparks, R. S. J., Huppert, H. E., and Turner, J. S. (1984). The fluid dynamics of evolving magma chambers. *Philosophical Transactions of the Royal Society of London*, **A 310**, 511–534.

Sparks, R. S. J., Huppert, H. E., Kerr, R. C., McKenzie, D. P., and Tait, S. R. (1985). Postcumulus processes in layered intrusions. *Geological Magazine*, **122**, 555–568.

Sparks, R. S. J., Baker, L., Brown, R. J., *et al*. (2006). Dynamical constraints on kimberlite volcanism. *Journal of Volcanology and Geothermal Research*, **155**, 18–48.

Spear, F. S. (1993). *Metamorphic Phase Equilibria and Pressure-temperature-time Paths*. Monograph no. 1. Washington, DC: Mineralogical Society of America.

Spear, F. S., and Cheney, J. T. (1989). A petrogenetic grid for pelitic schists in the system SiO_2–$A1_2O_3$–FeO–MgO–K_2O–H_2O. *Contributions to Mineralogy and Petrology*, **101**, 149–164.

Spear, F. S., and Selverstone, J. (1983). Quantitative *P–T* paths from zoned minerals: theory and tectonic applications. *Contributions to Mineralogy and Petrology*, **83**, 348–357.

Spear, F. S., Selverstone, J., Hickmott, D., Crowley, P., and Hodges, K. V. (1984). *P–T* paths from garnet zoning: a new technique for deciphering tectonic processes in crystalline terranes. *Geology*, **12**, 87–90.

Spear, F. S., Peacock, S. M., Kohn, M. J., Florence, F. P. and Menard, T. (1991). Computer programs for petrologic *P-T-t* path calculations. *American Mineralogist*, **76**, 2009–2012.

Spera, F. J. (1980). Aspects of magma transport. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 265–323.

Spera, F. J., Yuen, D. A., and Kirschvink, S. J. (1982). Thermal boundary layer convection in silicic magma chambers: effects of temperature-dependent rheology and implications for thermogravitational chemical fractionation. *Journal of Geophysical Research*, **87**, 8755–8767.

Spiegelman, M., Kelemen, P. B., and Aharonov, E. (2001). Causes and consequences of flow organization during melt transport: the reaction infiltration instability in compactible media. *Journal of Geophysical Research*, **106**, 2061–2078.

Spry, A. (1969). *Metamorphic Textures*. Oxford: Pergamon Press.

Staudigel, H., Hart, S., Schmincke, H., and Smith, B. (1989). Cretaceous ocean crust at DSDP sites 417 and 418: carbon uptake from weathering versus loss by magmatic outgassing. *Geochimica et Cosmochimica Acta*, **53**, 3091–3094.

Steefel, C. I., and Lasaga, A. C. (1994). A coupled model for transport of multiple chemical species and kinetic precipitation/dissolution reactions with application to reactive flow in single phase hydrothermal systems. *American Journal of Science*, **294**, 529–592.

Stein, C. A., and Stein, S. (1992). A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature*, **359**, 123–129.

Stein, C. A., and Stein, S. (1994). Constraints on hydrothermal heat flux through the oceanic lithosphere from global heat flow. *Journal of Geophysical Research*, **99**, 3081–3095.

Stevenson, D. J. (1989). Spontaneous small-scale melt segregation in partial melts undergoing deformation. *Geophysical Research Letters*, **16**, 1067–1070.

Stixrude, L., and Karki, B. (2005). Structure and freezing of $MgSiO_3$ liquid in Earth's lower mantle. *Science*, **310**, 297–299.

Stolper E., and Asimow, P. (2007). Insights into mantle melting from graphical analysis of one-component systems. *American Journal of Science* **307**, 1051 1139.

Stolper, E., and Walker, D. (1980). Melt density and the average composition of basalt. *Contributions to Mineralogy and Petrology*, **74**, 7–12.

Streckeisen, A. (1976). To each plutonic rock its proper name. *Earth Science Reviews*, **12**, 1–33.

Streckeisen, A. (1979). Classification and nomenclature of volcanic rocks, lamrophyres, carbonatites, andmelilitic rocks: recommendations and suggestions of the IUGS Subcommission on the Systematics of Igneous Rocks. *Geology*, **7**, 331–335.

Strens, R. G. J. (1968). Stability of the Al_2SiO_5 solid solutions. *Mineralogical Magazine*. **36**, 839–849.

Stüwe, K. (2002). *Geodynamics of the Lithosphere*. Berlin: Springer-Verlag.

Sunagawa, I. (1992). In situ investigation of nucleation, growth, and dissolution of silicate crystals at high temperatures. *Annual Review of Earth and Planetary Sciences*, **20**, 113–142.

Suppe, J. (1985). *Principles of Structural Geology*. Englewood Cliffs, NJ: Prentice-Hall, 537 pp.

Suzuki, A., and Ohtani, E. (2003). Density of peridotite melts at pressure. *Physics and Chemistry of Minerals*, **30**, 449–456.

Swanson, S. E. (1977). Relation of nucleation and crystal-growth rate to the development of granitic textures. *American Mineralogist*, **62**, 966–978.

Swanson, S. E., and Fenn, P. M. (1992). The effect of F and Cl on albite crystallization: a model for granitic pegmatites? *Canadian Mineralogist*, **30**, 549–559.

Symmes, G. H., and Ferry, J. M. (1995). Metamorphism, fluid flow, and partial melting in pelitic rocks from the Onawa contact aureole, central Maine, USA. *Journal of Petrology*, **36**, 587–612.

Tait, S., and Jaupart, C. (1992). Compositional convection in a reactive crystalline mush and melt differentiation. *Journal of Geophysical Research*, **97**, 6735–6756.

Takahashi, E. (1986). Melting of a dry peridotite KLB-1 up to 14 GPa: implications on the origin of peridotitic upper mantle. *Journal of Geophysical Research*, **91**, 9367–9382.

Takahashi, E., and Scarfe, C. M. (1985). Melting of peridotite to 14 GPa and the genesis of komatiite. *Nature*, **315**, 566–568.

Tarduno, J. A. (2007). On the motion of Hawaii and other mantle plumes. *Chemical Geology*, **241**, 234–247.

Tarling, D. H. (1980). Lithosphere evolution and changing tectonic regimes. *Geological Society of London Quarterly Journal*, **137**, 459–467.

Tarling, D. H., and Hrouda, F. (1993). *The Magnetic Anisotropy of Rocks*. London: Chapman & Hall, 217 pp.

Taylor, H. P., Jr. (1990). Oxygen and hydrogen isotope constraints on the deep circulation of surface waters into zones of hydrothermal metamorphism and melting. In *The Role of Fluids in Crustal Processes*, ed. J. D. Bredehoeft, and D. L. Norton. Studies in Geophysics. Washington, DC: National Academy Press, 72–95.

Tegner, C., Wilson, J. R., and Robins, B. (2005). Crustal assimilation in basalt and jotunite: constraints from layered intrusions. *Lithos*, **83**, 299–316.

Thayer, T. P. (1964). Principal features and origin of podiform chromite deposits and some observations on the Guleman-Soridag district, Turkey. *Economic Geology*, **59**, 1497–1524.

Thompson, A. B. (1976a). Mineral reactions in pelitic rocks: I. Predictions of *P–T–X*(Fe-Mg) phase relations. *American Journal of Science*, **276**, 401–424.

Thompson, A. B. (1976b). Mineral reactions in pelitic rocks: II. Calculation of some *P–T–X*(Fe-Mg) phase relations. *American Journal of Science* **276**, 425 454.

Thompson, A. B., and England, P. C. (1984). Pressure-temperature-time paths of regional metamorphism: II. Their inference and interpretation using mineral assemblages in metamorphic rocks. *Journal of Petrology*, **25**, 929–955.

Thompson, J. B., Jr. (1957). The graphical analysis of mineral assemblages in pelitic schists. *American Mineralogist*, **42**, 842–858.

Thompson, J. B., Jr. (1959). Local equilibrium in metasomatic processes. In *Researches in Geochemistry*, vol. 1, ed. P. H. Abelson. New York: John Wiley, 427–457.

Thompson, J. B., Jr. (1967). Thermodynamic properties of simple solutions. In *Researches in Geochemistry*, vol. 2, ed. P. H. Abelson. New York: John Wiley, 340–361.

Thompson, R. N., Morrison, M. A., Dickin, A. P., and Hendry, G. L. (1983). Continental flood basalts . . . arachnids rule OK? In *Continental Basalts and Mantle Xenoliths*, ed. C. J. Hawkesworth, and M. J. Norry. Nantwich, Cheshire: Shiva Publishing, 158–185.

Thomson, W. (Lord Kelvin) (1863). The secular cooling of the Earth. *Transactions of the Royal Society of Edinburgh*, **23**, 157–170.

Thorarinsson, S. (1968). On the rate of lava- and tephra-production and the upward migration of magma in four Icelandic eruptions. *International Journal of Earth Sciences*, **57**, 705–718.

Thorarinsson, S. (1981). Tephra studies and tephrochronology: a historical review with special reference to Iceland. In *Tephra Studies*, ed. S. Self, and R. S. J. Sparks. Dordrecht: D. Reidel, 1–12.

Thornton, C. P., and Tuttle, O. F. (1960). Chemistry of igneous rocks, I: differentiation index. *American Journal of Science*, **258**, 664–668.

Thorpe, R. S. (ed.) (1982). *Andesites: Orogenic Andesites and Related Rocks*. Chichester, West Sussex: John Wiley.

Tilley, C. E. (1924). The facies classification of metamorphic rocks. *Geological Magazine*, **61**, 167–171.

Tilley, C. E. (1948). Earlier stages in the metamorphism of siliceous dolomite. *Mineralogical Magazine*, **28**, 272–276.

Tilley, C. E. (1952). Some trends of basaltic magma in limestone syntexis. *American Journal of Science*, Bowen Volume, 529–545.

Tilling, R. I., and Dvorak, J. J. (1993). Anatomy of a basaltic volcano. *Nature*, **363**, 125–133.

Toksoz, M. N., and Hsui, A. T. (1978). Numerical studies of back-arc convection and the formation of marginal basins. *Tectonophysics*, **50**, 177–196.

Tolstikhin, I. N., Kramers, J. D., and Hofmann, A. W. (2006). A chemical Earth model with whole mantle convection: the importance of a core–mantle boundary layer (D'') and its early formation. *Chemical Geology*, **226**, 79–99.

Toplis, M. J. (2005). The thermodynamics of iron and magnesium partitioning between olivine and liquid: criteria for assessing and predicting equilibrium in natural and experimental systems. *Contributions to Mineralogy and Petrology*, **149**, 22–49.

Touret, J. L. R. (1985). Fluid regime in southern Norway: the record of fluid inclusions. In *The Deep Proterozoic Crust in the North Atlantic Provinces*, ed. A. C. Tobi, and J. L. R. Touret. Dordrecht: D. Reidel, 517–549.

Tracy, R. J. (1982). Compositional zoning and inclusions in metamorphic minerals. In *Characterization of Metamorphism Through Mineral Equilibria*, ed. J. M. Ferry, vol. 2 of *Reviews in Mineralogy*. Washington, DC: Mineralogical Society of America, 355–397.

Tracy, R. J., and Robinson, P. (1980). Evolution of metamorphic belts: information from detailed petrologic studies. In *The Caledonides in the U.S.A.*, ed. D. J. Wones. Virginia Polytechnic Institute and State University, Memoir 2, 189 195.

Tracy, R. J., Robinson, P., and Thompson, A. B. (1976). Garnet composition and zoning in the determination of temperature and pressure of metamorphism, central Massachusetts. *American Mineralogist*, **61**, 762–775.

Tracy, R. J., Rye, D. M., Hewitt, D. A., and Schiffries, C. M. (1983). Petrologic and stable-isotopic studies of fluid-rock interactions, south-central Connecticut, I: the role of infiltration in producing reaction assemblages in impure marbles. *American Journal of Science*, **283-A**, 589–616.

Tritton, D. J. (1977). *Physical Fluid Dynamics*. Wokingham: Van Nostrand (UK), 362 pp.

Tullis, J., and Yund, R. A. (1987). The brittle–ductile transition in feldspathic rocks. *Eos Transactions, American Geophysical Union*, **68**, 1464.

Turcotte, D. L., and Schubert, G. (1982a). Fluid mechanics. In *Geodynamics, Applications of Continuum Physics to Geological Problems*. New York: John Wiley, Chap. 6, 226–291.

Turcotte, D. L., and Schubert, G. (1982b). Heat transfer. In *Geodynamics, Applications of Continuum Physics to Geological Problems*. New York: John Wiley, Chap. 4, 134–197.

Turner, D. C. (1963). Ring-structures in the Sara-Fier complex, northern Nigeria. *Quarterly Journal of the Geological Society of London*, **119**, 345–366.

Turner, F. J., and Weiss, L. E. (1963). *Structural Analysis of Metamorphic Tectonites*. New York: McGraw-Hill.

Turner, J. S. (1973). *Buoyancy Effects in Fluids*. Cambridge: Cambridge University Press, 368 pp.

Turner, S. P., George, M. M., Evans, P. J., Hawkesworth, C. J., and Zellmer, G. F. (2000). Time-scales of magma formation, ascent and storage beneath subduction-zone volcanoes. *Philosophical Transactions of the Royal Society of London*, **358**, 1443–1464.

Turner, S. P., Evans, P. J., and Hawkesworth, C. J. (2001). Ultrafast source-to-surface movement of melt at island arcs from ^{226}Ra-^{230}Th systematics. *Science*, **292**, 1363–1366.

Tuttle, O. F. and Bowen, N. L. (1958). Origin of granite in the light of experimental studies in the system $NaAlSi_3O_8$–$KalSi_3O_8$–SiO_2–H_2O. *Geological Society of America Memoir* **74**, 153 pp.

Tuttle, O. F., and Gittins, J. (ed.) (1966). *Carbonatites*. New York: John Wiley, 591 pp.

Ulmer, P. (2001). Partial melting in the mantle wedge: the role of H_2O in the genesis of mantle-derived "arc-related" magmas. *Physics of the Earth and Planetary Interiors*, **127**, 215–232.

Valley, J. W. (2001). Stable isotope thermometry at high temperatures. In *Stable Isotope Geochemistry*, ed. J. W. Valley, and D. R. Cole, vol. 43 of *Reviews in Mineralogy and Geochemistry.* Washington DC: Mineralogical Society of America, 365–414.

Valley, J. W., Bohlen, S. W., Essene, E. J., and Lamb, W. (1990). Metamorphism in the Adirondacks, II: the role of fluids. *Journal of Petrology*, **31**, 555–596.

van der Hilst, R. D. and Kárason, H. (1999). Compositional heterogeneity in the bottom 1000 kilometers of Earth's mantle: towards a hybrid convection model. *Science*, **283**, 1885–1888.

van Genabeek, O., and Rothman, D. H. (1996). Macroscopic manifestations of microscopic flows through porous media: phenomenology from simulation. *Annual Review of Earth and Planetary Sciences*, **24**, 63–87.

van Haren, J. L. M., Ague, J. J., and Rye, D. M. (1996). Oxygen isotope record of fluid infiltration and mass transfer during regional metamorphism of pelitic schist, Connecticut, USA. *Geochimica et Cosmochimica Acta*, **60**, 3487–3504.

Veksler, I. V., Dorfman, A. M., Danyushevsky, L. V., Jakobsen, J. K., and Dingwell, D. B. (2006). Immiscible silicate liquid partition coefficients: implications for crystal-melt element partitioning and basalt petrogenesis. *Contributions to Mineralogy and Petrology*, **152**, 685–702.

Velde, B. (1965). Phegite micas: synthesis, stability, and natural occurrence. *American Journal of Science*, **263**, 886–913.

Vernon, R. H., and Clarke, G. L. (2008). *Principles of Metamorphic Petrology.* Cambridge: Cambridge University Press.

Vidale, R. J. (1974). Vein assemblages and metamorphism in Dutchess County, New York. *Geological Society of America Bulletin*, **85**, 303–306.

Viljoen, M. J., and Viljoen, R. P. (1969). Evidence for the existence of a mobile extrusive peridotitic magma from the Komati Formation of the Onverwacht group. *Geological Society of South Africa Special Publication*, **2**, *Upper Mantle Project*, 87–112.

Vityk, M. O., and Bodnar, R. J. (1995). Do fluid inclusions in high-grade metamorphic terranes preserve peak metamorphic density during retrograde decompression? *American Mineralogist*, **80**, 641–644.

Vogel, T. A., and Wilband, J. T. (1978). Coexisting acidic and basic melts: geochemistry of a composite dike. *Journal of Geology*, **86**, 353–371.

von Bargen, N., and Waff, H. S. (1986). Permeabilites, interfacial areas and curvatures of partially molten systems: results of numerical computations of equilibrium microstructures. *Journal of Geophysical Research*, **91**, 9261–9276.

Von Damm, K. L. (1990). Seafloor hydrothermal activity: black smoker chemistry and chimneys. *Annual Review of Earth and Planetary Sciences*, **18**, 173–204.

Waff, H. S. (1986). Introduction to special section on partial melting phenomena in Earth and planetary evolution. *Journal of Geophysical Research*, **91**, 9217–9221.

Wager, L. R., and Brown, G. M. (1967). *Layered Igneous Rocks*. Edinburgh: Oliver & Boyd, 588 pp.

Wager, L. R., and Deer, W. A. (1939). Geological investigations in East Greenland: III. The petrology of the Skaergaard Intrusion, Kangerdluqssuaq, East Greenland. *Meddelelser om Grønland*, **105**, 1–352.

Wager, L. R., Brown, G. M., and Wadsworth, W. J. (1960). Types of igneous cumulates. *Journal of Petrology*, **1**, 73–85.

Waldbaum, D. R., and Thompson, J. B., Jr. (1969). Mixing properties of sanidine crystalline solutions. *American Mineralogist*, **54**, 1274–1298.

Walder, J., and Nur, A. (1984). Porosity reduction and crustal pore pressure development. *Journal of Geophysical Research*, **89**, 11 539–11 548.

Walker, D. (1986). Melting equilibria in multicomponent systems and liquidus/solidus convergence in mantle peridotite. *Contributions to Mineralogy and Petrology*, **92**, 303–307.

Walker, D., and DeLong, S. E. (1982). Soret separation of mid-ocean ridge basalt magma. *Contributions to Mineralogy and Petrology*, **79**, 231–240.

Walker, G. P. L. (1965). Evidence of crustal drift from Icelandic geology. *Philosophical Transactions of the Royal Society of London*, **A 258**, 199–204.

Walker, G. P. L. (1989). Gravitational (density) controls on volcanism, magma chambers and intrusions. *Australian Journal of Earth Science*, **36**, 149–165.

Wall, F., and Zaitsev, A. N. (ed.) (2004). *Phoscorites and Carbonatites from Mantle to Mine: The Key Example of the Kola Alkaline Province*. Mineralogical Society Series, 10. London: Mineralogical Society, 503 pp.

Wallace, P. and Carmichael, I. S. E. (1992). Sulfur in basaltic magmas. *Geochimica et Cosmochimca Acta*, **56**, 1863–1874.

Walther, J. V. (1990). Fluid dynamics during progressive regional metamorphism. In *The Role of Fluids in Crustal Processes*, ed. J. D. Bredehoeft, and D. L. Norton. Studies in Geophysics. Washington, DC: National Academy Press 64–70.

Walther, J. V., and Orville, P. M. (1982). Volatile production and transport in regional metamorphism. *Contributions to Mineralogy and Petrology*, **79**, 252–257.

Walther, J. V., and Wood, B. J. (1984). Rate and mechanism in prograde metamorphism. *Contributions to Mineralogy and Petrology*, **88**, 246–259.

Wark, D. A., and Watson, E. B. (2004). Interdiffusion of H_2O and CO_2 in metamorphic fluids at ~490 to 690 degrees C and 1 GPa. *Geochimica et Cosmochimica Acta*, **68**, 2693–2698.

Wark, D. A., and Watson, E. B. (2006). TitaniQ: a titanium-in-quartz geothermometer. *Contributions to Mineralogy and Petrology*, **152**, 743–754.

Wark, D. A., Williams, C. A., Watson, E. B., and Price, J. D. (2003). Reassessment of pore shapes in microstructurally equilibrated rocks, with implications for permeability of the upper mantle. *Journal of Geophysical Research*, **108**, doi:10.1029/2001JB001575.

Wark, D. A., Hildreth, W., Spear, F. S., Cherniak, D. J., and Watson, E. B. (2007). Pre-eruption recharge of the Bishop magma system. *Geology*, **35**, 235–238.

Waters, A. C. (1955). Volcanic rocks and the tectonic cycle. *Geological Society of America, Special Paper*, **62**, 703–722.

Watson, E. B. (1976). Two-liquid partition coefficients: experimental data and geochemical implications. *Contributions to Mineralogy and Petrology*, **56**, 119–134.

Watson, E. B. (1982). Melt infiltration and magma evolution. *Geology*, **10**, 236–240.

Watson, E. B. (1999). Lithologic partitioning of fluids and melts. *American Mineralogist*, **84**, 1693–1710.

Watson, E. B., and Brenan, J. M. (1987). Fluids in the lithosphere, 1: experimentally-determined wetting characteristics of CO_2–H_2O fluids and their implications for fluid transport, host-rock physical properties, and fluid inclusion formation. *Earth and Planetary Science Letters*, **85**, 497–515.

Watson, E. B., Harrison, T. M., Ryerson, F. J. (1985). Diffusion of Sm, Sr, and Pb in fluorapatite. *Geochimica et Cosmochimica Acta*, **49**, 1813–1823.

Watson, E. B., Wark, D. A., and Thomas, J. B. (2006). Crystallization thermometers for zircon and rutile. *Contributions to Mineralogy and Petrology*, **151**, 413–433.

Weigand, P. W., and Ragland, P. C. (1970). Geochemistry of Mesozoic dolerite dikes from eastern North America. *Contributions to Mineralogy and Petrology*, **29**, 195–214.

Weill, D. F., Hon, R., and Navrotsky, A. (1980). The igneous system $CaMgSi_2O_6$–$CaAl_2Si_2O_8$–$NaAlSi_3O_8$: variations on a classic theme by Bowen. In *Physics of Magmatic Processes*, ed. R. B. Hargraves. Princeton, NJ: Princeton University Press, 49–92.

Wetherill, G. W. (1990). Formation of the Earth. *Annual Review of Earth and Planetary Sciences*, **18**, 205–256.

White, A. J. R., and Chappell, B. W. (1983). Granitoid types and their distribution in the Lachlan Fold Belt, southeastern Australia. *Geological Society of America Memoir*, **159**, 21–34.

White, N. (1989). Nature of lithospheric extension in the North Sea. *Geology*, **17**, 111–114.

White, R. S., and McKenzie, D. (1989). Magmatism at rift zones: the generation of volcanic continental margins and flood basalts. *Journal of Geophysical Research* **94**, 7685 7729.

White, W. M., Dupre, B., and Vidal, P. (1985). Isotope and trace element geochemistry of sediments from the Barbados Ridge–Demerara Plain region, Atlantic Ocean. *Geochimica et Cosmochimica Acta*, **49**, 1875–1886.

Whitney, D. L., Mechum, T. A., Dilek, Y., and Kuehner, S. M. (1996). Modification of garnet by fluid infiltration during regional metamorphism in garnet through sillimanite zone rocks, Dutchess County, New York. *American Mineralogist*, **81**, 696–705.

Whitney, J. A. (1988). The origin of granite: the role and source of water in the evolution of granitic magmas. *Geological Society of America Bulletin*, **100**, 1886–1897.

Wickham, S. M., and Oxburgh, E. R. (1985). Continental rifts as a setting for regional metamorphism. *Nature*, **318**, 330–333.

Wicks, C., Jr., Thatcher, W., and Dzurisin, D. (1998). Migration of fluids beneath Yellowstone caldera inferred from satellite radar interferometry. *Science*, **282**, 458–462.

Wicks, C., Jr., Thatcher, W., Dzurisin, D., and Svarc, J. (2006). Uplift, thermal unrest and magma intrusion at Yellowstone caldera. *Nature*, **440**, 72–75.

Wiebe, R. A. (1993). The Pleasant Bay layered gabbro–diorite, coastal Maine: ponding and crystallization of basaltic injections into a silicic magma chamber. *Journal of Petrology*, **34**, 461–489.

Wiebe, R. A., Manon, M. R., Hawkins, D. P., and McDonough, W. F. (2004). Late-stage mafic injection and thermal rejuvination of the Vinalhaven granite, coastal Maine. *Journal of Petrology*, **45**, 2133–2153.

Wieczorek, M. A., Zuber, M. T., and Phillips, R. J. (2001). The role of magma buoyancy on the eruption of lunar basalts. *Earth and Planetary Science Letters*, **185**, 71–83.

Wignall, P. (2005). The link between large igneous province eruptions and mass extinctions. *Elements*, **1**, 293–267.

Wijbrans, J. R., and McDougall, I. (1986). $^{40}Ar/^{39}Ar$ dating of white micas from an Alpine high-pressure metamorphic belt on Naxos (Greece): the resetting of the argon isotopic system. *Contributions to Mineralogy and Petrology*, **93**, 187–194.

Wijbrans, J. R., and McDougall, I. (1988). Metamorphic evolution of the Attic Cycladic Metamorphic Belt on Naxos (Cyclades, Greece) utilizing $^{40}Ar/^{39}Ar$ age spectrum measurements. *Journal of Metamorphic Geology*, **6**, 571–594.

Wilbur, D. E., and Ague, J. J. (2006). Chemical disequilibrium during garnet growth: Monte Carlo simulations of natural crystal morphologies. *Geology*, **34**, 689–692.

Wilbur, D. E., and Ague, J. J. (2007). Chemical disequilibrium during garnet growth: Monte Carlo simulations of natural crystal morphologies – Reply. *Geology*, **35**, doi: 10.1130/G23787Y, e125.

Wilcox, R. E. (1954). Petrology of Paricutin volcano, Mexico. *U.S. Geological Survey Bulletin*, **965-C**, 281–353.

Wilde, S. A., Valley, J. W., Peck, W. H., and Graham, C. M. (2001). Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature*, **409**, 175–178.

Williams, A. F. (1932). *The Genesis of Diamonds*. London: Ernest Benn, 2 vols., 636 pp.

Williams, H. M., Turner, S. P., Pearce, J. A., Kelley, S. P., and Harris, N. B. W. (2004). Nature of the source regions of post-collisional, potassic magmatism in southern and northern Tibet from geochemical variations and inverse trace element modeling. *Journal of Petrology*, **45**, 555–607.

Williams, M. L., Jercinovic, M. J., and Hetherington, C. J. (2007). Microprobe monazite geochronology: understanding geologic processes by integrating composition and chronology. *Annual Review of Earth and Planetary Sciences* **35**, 137 175.

Wilson, C. J. N. (1984). The role of fluidization in the emplacement of pyroclastic flows, 2: experimental results and their interpretation. *Journal of Volcanology and Geothermal Research*, **20**, 55–84.

Wilson, J. R., and Larsen, S. B. (1985). Two-dimensional study of a layered intrusion – the Hyllingen Series, Norway. *Geological Magazine*, **122**, 97–124.

Wilson, J. R., Robins, B., Nielsen, F. M. Duchesne, J.-C., and Vander Auwera, J. (1996). The Bjerkreim–Sokndal layered intrusion. In *Layered Intrusions*, ed. R. G. Cawthorn. Amsterdam: Elsevier, 303–329.

Wilson, J. T. (1963). A possible origin of the Hawaiian Islands. *Canadian Journal of Physics*, **41**, 863–870.

Wilson, J. T. (1973). Mantle plumes and plate motions. *Tectonophysics*, **19**, 149–164.

Wilson, L., and Head, J. W., III (1981). Ascent and eruption of basaltic magma on the Earth and Moon. *Journal of Geophysical Research*, **86**, 2971–3001.

Wing, B. A., and Ferry, J. M. (2007). Magnitude and geometry of reactive fluid flow from direct inversion of spatial patterns of geochemical alteration. *American Journal of Science*, **307**, 793–832.

Winkler, H. G. F. (1949). Crystallization of basaltic magma as recorded by variations of crystal size in dikes. *Mineralogical Magazine*, **28**, 557–574.

Wolfe, C. J. and Solomon, S. C. (1998). Shear-wave splitting and implications for mantle flow beneath the MELT region of the East Pacific Rise. *Science*, **280**, 1230–1232.

Wood, B. J., and Fraser, D. G. (1976). *Elementary Thermodynamics for Geologists*. Oxford: Oxford University Press, 303 pp.

Woolsey, T. S., McCallum, M. E., and Schumm, S. A. (1975). Modelling of diatreme emplacement by fluidization. *Physics and Chemistry of the Earth*, **9**, 30–42.

Wopenka, B., and Pasteris, J. D. (1993). Structural characterization of kerogens to granulite-facies graphite: applicability of Raman microprobe spectroscopy. *American Mineralogist*, **78**, 533–557.

Wright, T. L., and Doherty, P. C. (1970). A linear programming and least squares computer method for solving petrologic mixing problems. *Geological Society of America Bulletin*, **81**, 1995–2008.

Wright, T. L., and Fiske, R. S. (1971). Origin of the differentiated and hybrid lavas of Kilauea volcano, Hawaii. *Journal of Petrology*, **12**, 1–65.

Wright, T. J., Ebinger, C., Biggs, J., *et al*. (2006). Magma-maintained rift segmentation at continental rupture in the 2005 Afar dyking episode. *Nature*, **442**, 291–294.

Wyllie, P. J. (1977). Crustal anatexis: an experimental review. *Tectonophysics*, **13**, 41–71.

Wyllie, P. J. (1980). The origin of kimberlites. *Journal of Geophysical Research*, **85**, 6702–6910.

Wyllie, P. J. (1988). Magma genesis, plate tectonics, and chemical differentiation of the Earth. *Reviews of Geophysics*, **26**, 370–404.

Wyllie, P. J., and Huang, W. L. (1976). Carbonation and melting reactions in the system CaO–MgO–SiO_2–CO_2 at mantle pressures with geophysical and petrological applications. *Contributions to Mineralogy and Petrology*, **54**, 79–107.

Wyllie, P. J., and Tuttle, O. F. (1960). The system CaO-CO_2-H_2O and the origin of carbonatites. *Journal of Petrology*, **1**, 1–46.

Yardley, B. W. D. (1975). On some quartz-plagioclase veins in the Connemara Schists, Ireland. *Geological Magazine*, **112**, 183–190.

Yardley, B. W. D. (1986a). Fluid migration and veining in the Connemara Schists, Ireland. In *Fluid-rock Interactions During Metamorphism*, ed. J. V. Walther, and B. J. Wood. New York: Springer, 109–131.

Yardley, B. W. D. (1986b). Is there water in the deep continental crust? *Nature*, **323**, 111.

Yirgu, G. (2007). Manda Hararo. *Smithsonian Bulletin of the Global Volcanism Network*, **32**(7), 2–3.

Yochelson, E. L. (1980). The scientific ideas of G. K. Gilbert. *Geological Society of America Special Paper*, **183**, 148 pp.

Yoder, H. S., Jr. (1965). Diopside–anorthite–water at five and ten kilobars and its bearing on explosive volcanism. *Carnegie Institution of Washington Yearbook*, **64**, 82–89.

Yoder, H. S., Jr., (1976). *Generation of Basaltic Magma*. Washington, DC: National Academy of Sciences, 265 pp.

Yoder, H. S., Jr., and Tilley, C. E. (1962). Origin of basalt magmas: an experimental study of natural and synthetic rock systems. *Journal of Petrology*, **3**, 342–532.

Yoder, H. S., Jr., Stewart, D. B., and Smith, J. R. (1957). Ternary feldspars. *Carnegie Institution of Washington Yearbook*, **55**, 206–214.

Young, E. D., and Rumble, D. (1993). The origin of correlated variations in in-situ $^{18}O/^{16}O$ and elemental concentrations in metamorphic garnet from southeastern Vermont, USA. *Geochimica et Cosmochimica Acta*, **57**, 2585–2597.

Zack, T., Moraes, R., and Kronz, A. (2005). Temperature dependence of Zr in rutile: empirical calibration of a rutile thermometer. *Contributions to Mineralogy and Petrology*, **148**, 471–488.

Zegers, T. E., and van Keken, P. E. (2001). Middle Archean continent formation by crustal delamination. *Geology*, **29**, 1083–1086.

Zen, E-an. (1966). Construction of pressure temperature diagrams for multicomponent systems after the method of Schreinemakers – a geometric approach. *U.S. Geological Survey Bulletin*, **1225**, 56 pp.

Zhang, Y. (2002). The age and accretion of the Earth. *Earth Science Reviews*, **59**, 235–263.

Zieg, M. J., and Marsh, B. D. (2002). Crystal size distributions and scaling laws in the quantification of igneous textures. *Journal of Petrology*, **43**, 85–101.

Zindler, A., and Hart, S. (1986). Chemical geodynamics. *Annual Review of Earth and Planetary Sciences*, **14**, 493–571.

Index