

# Fundamentals of Geophysics

Second Edition

WILLIAM LOWRIE

*Swiss Federal Institute of Technology, Zürich*



CAMBRIDGE  
UNIVERSITY PRESS

(5G)

BC 806

.L617

2007

C.2

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

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](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9780521675963](http://www.cambridge.org/9780521675963)

© W. Lowrie 2007

This publication is in copyright. Subject to statutory exception  
and to the provisions of relevant collective licensing agreements,  
no reproduction of any part may take place without  
the written permission of Cambridge University Press.

First published 2007

Printed in United Kingdom at the University Press, Cambridge

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

ISBN-13 978-0-521-85902-8 hardback

ISBN-13 978-0-521-67596-3 paperback

Cambridge University Press has no responsibility for the persistence or  
accuracy of URLs for external or third-party websites referred to  
in this publication, and does not guarantee that any content on such  
websites is, or will remain, accurate or appropriate

# 3 Seismology and the internal structure of the Earth

## 3.1 INTRODUCTION

Seismology is a venerable science with a long history. The Chinese scientist Chang Heng is credited with the invention in 132 AD, nearly two thousand years ago, of the first functional seismoscope, a primitive but ingenious device of elegant construction and beautiful design that registered the arrival of seismic waves and enabled the observer to infer the direction they came from. The origins of earthquakes were not at all understood. For centuries these fearsome events were attributed to supernatural powers. The accompanying destruction and loss of life were often understood in superstitious terms and interpreted as punishment inflicted by the gods on a sinful society. Biblical mentions of earthquakes – e.g., in the destruction of Sodom and Gomorrah – emphasize this vengeful theme. Although early astronomers and philosophers sought to explain earthquakes as natural phenomena unrelated to spiritual factors, the belief that earthquakes were an expression of divine anger prevailed until the advent of the Age of Reason in the eighteenth century. The path to a logical understanding of natural phenomena was laid in the seventeenth century by the systematic observations of scientists like Galileo, the discovery and statement of physical laws by Newton and the development of rational thought by contemporary philosophers.

In addition to the development of the techniques of scientific observation, an understanding of the laws of elasticity and the limited strength of materials was necessary before seismology could progress as a science. In a pioneering study, Galileo in 1638 described the response of a beam to loading, and in 1660 Hooke established the law of the spring. However, another 150 years passed before the generalized equations of elasticity were set down by Navier. During the early decades of the nineteenth century Cauchy and Poisson completed the foundations of modern elasticity theory.

Early descriptions of earthquake characteristics were necessarily restricted to observations and measurements in the “near-field” region of the earthquake, i.e. in comparatively close proximity to the place where it occurred. A conspicuous advance in the science of seismology was accomplished with the invention of a sensitive and reliable seismograph by John Milne in 1892. Although massive and primitive by comparison with modern instruments, the precision and sensitivity of this revolu-

tional new device permitted accurate, quantitative descriptions of earthquakes at large distances from their source, in their “far-field” region. The accumulation of reliable records of distant earthquakes (designated as “teleseismic” events) made possible the systematic study of the Earth’s seismicity and its internal structure.

The great San Francisco earthquake of 1906 was intensively studied and provided an impetus to efforts at understanding the origin of these natural phenomena, which were clarified in the same year by the elastic rebound model of H. F. Reid. Also, in 1906, R. D. Oldham proposed that the best explanation for the travel-times of teleseismic waves through the body of the Earth required a large, dense and probably fluid core; the depth to its outer boundary was calculated in 1913 by B. Gutenberg. From the analysis of the travel-times of seismic body waves from near earthquakes in Yugoslavia, A. Mohorovičić in 1909 inferred the existence of the crust–mantle boundary, and in 1936 the existence of the solid inner core was deduced by I. Lehmann. The definitions of these and other discontinuities associated with the deep internal structure of the Earth have since been greatly refined.

The needs of the world powers to detect incontrovertibly the testing of nuclear bombs by their adversaries provided considerable stimulus to the science of seismology in the 1950s and 1960s. The amount of energy released in a nuclear explosion is comparable to that of an earthquake, but the phenomena can be discriminated by analyzing the directions of first motion recorded by seismographs. The accurate location of the event required improved knowledge of seismic body-wave velocities throughout the Earth’s interior. These political necessities of the Cold War led to major improvements in seismological instrumentation, and to the establishment of a new world-wide network of seismic stations with the same physical characteristics. These developments had an important feedback to the earth sciences, because they resulted in more accurate location of earthquake epicenters and a better understanding of the Earth’s structure. The pattern of global seismicity, with its predominant concentration in narrow active zones, was an important factor in the development of the theory of plate tectonics, as it allowed the identification of plate margins and the sense of relative plate motions.

The techniques of refraction and reflection seismology, using artificial, controlled explosions as sources, were developed in the search for petroleum. Since the 1960s these

methods have been applied with notable success to the resolution of detailed crustal structure under continents and oceans. The development of powerful computer technology enabled refinements in earthquake location and in the determination of travel-times of seismic body waves. These advances led to the modern field of seismic tomography, a powerful and spectacular technique for revealing regions of the Earth's interior that have anomalous seismic velocities. In the field of earthquake seismology, the need to protect populations and man-made structures has resulted in the investment of considerable effort in the study of earthquake prediction and the development of construction codes to reduce earthquake damage.

To appreciate how seismologists have unravelled the structure of the Earth's interior it is necessary to understand what types of seismic waves can be generated by an earthquake or man-made source (such as a controlled explosion). The propagation of a seismic disturbance through the Earth is governed by physical properties such as density, and by the way in which the material of the Earth's interior reacts to the disturbance. Material within the seismic source suffers permanent deformation, but outside the source the passage of a seismic disturbance takes place predominantly by elastic displacement of the medium; that is, the medium suffers no permanent deformation. Before analyzing the different kinds of seismic waves, it is important to have a good grasp of elementary elasticity theory. This requires understanding the concepts of stress and strain, and the various elastic constants that relate them.

## 3.2 ELASTICITY THEORY

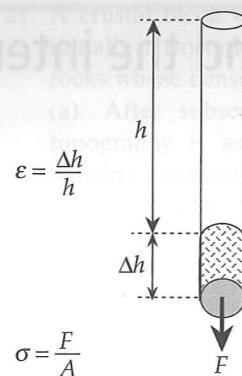
### 3.2.1 Elastic, anelastic and plastic behavior of materials

When a force is applied to a material, it deforms. This means that the particles of the material are displaced from their original positions. Provided the force does not exceed a critical value, the displacements are reversible; the particles of the material return to their original positions when the force is removed, and no permanent deformation results. This is called *elastic* behavior.

The laws of elastic deformation are illustrated by the following example. Consider a right cylindrical block of height  $h$  and cross-sectional area  $A$ , subjected to a force  $F$  which acts to extend the block by the amount  $\Delta h$  (Fig. 3.1). Experiments show that for elastic deformation  $\Delta h$  is directly proportional to the applied force and to the unstretched dimension of the block, but is inversely proportional to the cross-section of the block. That is,  $\Delta h \propto Fh/A$ , or

$$\frac{F}{A} \propto \frac{\Delta h}{h} \quad (3.1)$$

When the area  $A$  becomes infinitesimally small, the limiting value of the force per unit area ( $F/A$ ) is called the stress,  $\sigma$ . The units of stress are the same as the units of



**Fig. 3.1** A force  $F$  acting on a bar with cross-sectional area  $A$  extends the original length  $h$  by the amount  $\Delta h$ . Hooke's law of elastic deformation states that  $\Delta h/h$  is proportional to  $F/A$ .

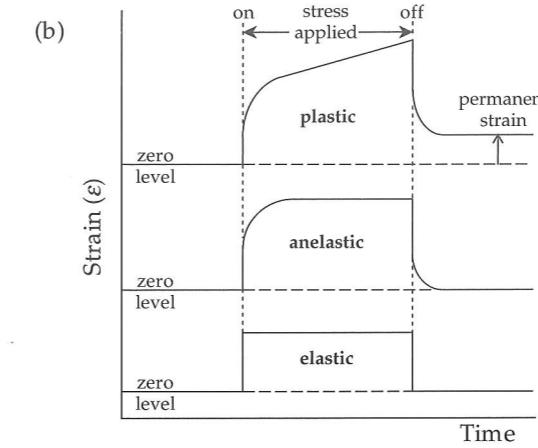
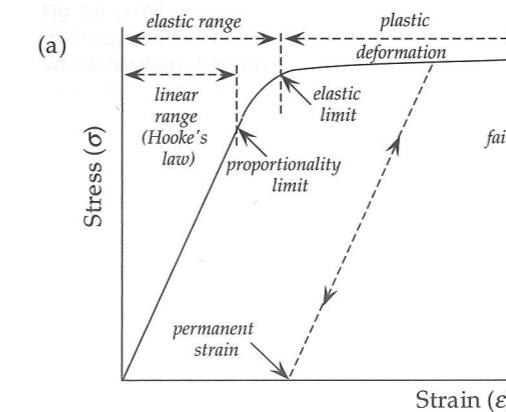
pressure. The SI unit is the pascal, equivalent to a force of 1 newton per square meter ( $1 \text{ Pa} = 1 \text{ N m}^{-2}$ ); the c.g.s. unit is the bar, equal to  $10^6 \text{ dyne cm}^{-2}$ .

When  $h$  is infinitesimally small, the fractional change in dimension ( $\Delta h/h$ ) is called the *strain*  $\epsilon$ , which is a dimensionless quantity. Equation (3.1) states that, for elastic behavior, the strain in a body is proportional to the stress applied to it. This linear relationship is called *Hooke's law*. It forms the basis of elasticity theory.

Beyond a certain value of the stress, called the *proportionality limit*, Hooke's law no longer holds (Fig. 3.2a). Although the material is still elastic (it returns to its original shape when stress is removed), the stress-strain relationship is non-linear. If the solid is deformed beyond a certain point, known as the *elastic limit*, it will not recover its original shape when stress is removed. In this range a small increase in applied stress causes a disproportionately large increase in strain. The deformation is said to be *plastic*. If the applied stress is removed in the plastic range, the strain does not return to zero; a permanent strain has been produced. Eventually the applied stress exceeds the strength of the material and *failure* occurs. In some rocks failure can occur abruptly within the elastic range; this is called *brittle* behavior.

The non-brittle, or *ductile*, behavior of materials under stress depends on the timescale of the deformation (Fig. 3.2b). An *elastic* material deforms immediately upon application of a stress and maintains a constant strain until the stress is removed, upon which the strain returns to its original state. A strain-time plot has a box-like shape. However, in some materials the strain does not reach a stable value immediately after application of a stress, but rises gradually to a stable value. This type of strain response is characteristic of *anelastic* materials. After removal of the stress, the time-dependent strain returns reversibly to the original level. In *plastic* deformation the strain keeps increasing as long as the stress is applied. When the stress is removed, the strain does not return to the original level; a permanent strain is left in the material.

## 3.2 ELASTICITY THEORY

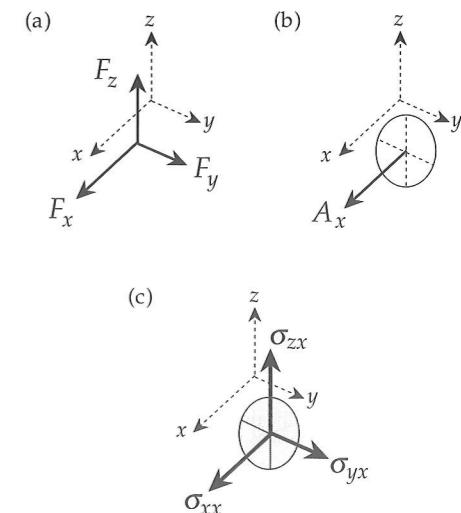


**Fig. 3.2** (a) The stress-strain relation for a hypothetical solid is linear (Hooke's law) until the proportionality limit, and the material deforms elastically until it reaches the elastic limit; plastic deformation produces further strain until failure occurs. (b) Variations of elastic, anelastic and plastic strains with time, during and after application of a stress.

Our knowledge of the structure and nature of the Earth's interior has been derived in large part from studies of seismic waves released by earthquakes. An earthquake occurs in the crust or upper mantle when the tectonic stress exceeds the local strength of the rocks and failure occurs. Away from the region of failure seismic waves spread out from an earthquake by elastic deformation of the rocks through which they travel. Their propagation depends on elastic properties that are described by the relationships between stress and strain.

### 3.2.2 The stress matrix

Consider a force  $\mathbf{F}$  acting on a rectangular prism  $P$  in a reference frame defined by orthogonal Cartesian coordinate axes  $x$ ,  $y$  and  $z$  (Fig. 3.3a). The component of  $\mathbf{F}$  which acts in the direction of the  $x$ -axis is designated  $F_x$ ; the force  $\mathbf{F}$  is fully defined by its components  $F_x$ ,  $F_y$  and  $F_z$ . The size of a small surface element is characterized by its area  $A$ , while its orientation is described by the direction normal to the surface (Fig. 3.3b). The small surface with area normal to the  $x$ -axis is designated  $A_x$ . The component of force  $F_x$  acting normal to the surface  $A_x$  produces a



**Fig. 3.3** (a) Components  $F_x$ ,  $F_y$  and  $F_z$  of the force  $\mathbf{F}$  acting in a reference frame defined by orthogonal Cartesian coordinate axes  $x$ ,  $y$  and  $z$ . (b) The orientation of a small surface element with area  $A_x$  is described by the direction normal to the surface. (c) The components of force parallel to the  $x$ -axis result in the normal stress  $\sigma_{xx}$ ; the components parallel to the  $y$ - and  $z$ -axes cause shear stresses  $\sigma_{yx}$  and  $\sigma_{zx}$ .

normal stress, denoted by  $\sigma_{xx}$ . The components of force along the  $y$ - and  $z$ -axes result in shear stresses  $\sigma_{yx}$  and  $\sigma_{zx}$  (Fig. 3.3c), given by

$$\sigma_{xx} = \lim_{A_x \rightarrow 0} \left( \frac{F_x}{A_x} \right) \quad \sigma_{yx} = \lim_{A_x \rightarrow 0} \left( \frac{F_y}{A_x} \right) \quad \sigma_{zx} = \lim_{A_x \rightarrow 0} \left( \frac{F_z}{A_x} \right) \quad (3.2)$$

Similarly, the components of the force  $\mathbf{F}$  acting on an element of surface  $A_y$  normal to the  $y$ -axis define a normal stress  $\sigma_{yy}$  and shear stresses  $\sigma_{xy}$  and  $\sigma_{zy}$ , while the components of  $\mathbf{F}$  acting on an element of surface  $A_z$  normal to the  $z$ -axis define a normal stress  $\sigma_{zz}$  and shear stresses  $\sigma_{xz}$  and  $\sigma_{yz}$ . The nine stress components completely define the state of stress of a body. They are described conveniently by the *stress matrix*

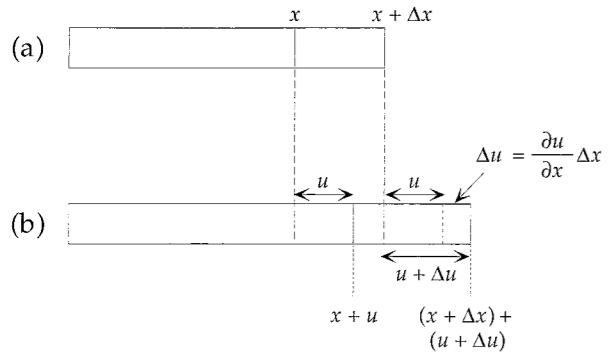
$$\begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \quad (3.3)$$

If the forces on a body are balanced to give no rotation, this  $3 \times 3$  matrix is symmetric (i.e.,  $\sigma_{xy} = \sigma_{yx}$ ,  $\sigma_{yz} = \sigma_{zy}$ ,  $\sigma_{zx} = \sigma_{xz}$ ) and contains only six independent elements.

### 3.2.3 The strain matrix

#### 3.2.3.1 Longitudinal strain

The strains produced in a body can also be expressed by a  $3 \times 3$  matrix. Consider first the one-dimensional case shown in Fig. 3.4 of two points in a body located close together at the positions  $x$  and  $(x + \Delta x)$ . If the point  $x$  is displaced by an infinitesimally small amount  $u$  in the



**Fig. 3.4** Infinitesimal displacements  $u$  and  $(u + \Delta u)$  of two points in a body that are located close together at the positions  $x$  and  $(x + \Delta x)$ , respectively.

direction of the  $x$ -axis, the point  $(x + \Delta x)$  will be displaced by  $(u + \Delta u)$ , where  $\Delta u$  is equal to  $(\partial u / \partial x) \Delta x$  to first order. The *longitudinal strain* or *extension* in the  $x$ -direction is the fractional change in length of an element along the  $x$ -axis. The original separation of the two points was  $\Delta x$ ; one point was displaced by  $u$ , the other by  $(u + \Delta u)$ , so the new separation of the points is  $(\Delta x + \Delta u)$ . The component of strain parallel to the  $x$ -axis resulting from a small displacement parallel to the  $x$ -axis is denoted  $\varepsilon_{xx}$ , and is given by

$$\varepsilon_{xx} = \frac{(\Delta x + \frac{\partial u}{\partial x} \Delta x) - \Delta x}{\Delta x} = \frac{\partial u}{\partial x} \quad (3.4)$$

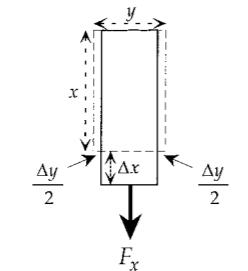
The description of longitudinal strain can be expanded to three dimensions. If a point  $(x, y, z)$  is displaced by an infinitesimal amount to  $(x + u, y + v, z + w)$ , two further longitudinal strains  $\varepsilon_{yy}$  and  $\varepsilon_{zz}$  are defined by

$$\varepsilon_{yy} = \frac{\partial v}{\partial y} \quad \text{and} \quad \varepsilon_{zz} = \frac{\partial w}{\partial z} \quad (3.5)$$

In an elastic body the transverse strains  $\varepsilon_{yy}$  and  $\varepsilon_{zz}$  are not independent of the strain  $\varepsilon_{xx}$ . Consider the change of shape of the bar in Fig. 3.5. When it is stretched parallel to the  $x$ -axis, it becomes thinner parallel to the  $y$ -axis and parallel to the  $z$ -axis. The transverse longitudinal strains  $\varepsilon_{yy}$  and  $\varepsilon_{zz}$  are of opposite sign but proportional to the extension  $\varepsilon_{xx}$  and can be expressed as

$$\varepsilon_{yy} = -\nu \varepsilon_{xx} \quad \text{and} \quad \varepsilon_{zz} = -\nu \varepsilon_{xx} \quad (3.6)$$

The constant of proportionality  $\nu$  is called *Poisson's ratio*. The values of the elastic constants of a material constrain  $\nu$  to lie between 0 (no lateral contraction) and a maximum value of 0.5 (no volume change) for an incompressible fluid. In very hard, rigid rocks like granite  $\nu$  is about 0.45, while in soft, poorly consolidated sediments it is about 0.05. In the interior of the Earth,  $\nu$  commonly has a value around 0.24–0.27. A body for which the value of  $\nu$  equals 0.25 is sometimes called an ideal Poisson body.



Poisson's ratio:

$$\nu = -\frac{\Delta y/y}{\Delta x/x} = -\frac{\varepsilon_{yy}}{\varepsilon_{xx}}$$

**Fig. 3.5** Change of shape of a rectangular bar under extension. When stretched parallel to the  $x$ -axis, it becomes thinner parallel to the  $y$ -axis and  $z$ -axis.

### 3.2.3.2 Dilatation

The *dilatation*  $\theta$  is defined as the fractional change in volume of an element in the limit when its surface area decreases to zero. Consider an undeformed volume element (as in the description of longitudinal strain) which has sides  $\Delta x$ ,  $\Delta y$  and  $\Delta z$  and undistorted volume  $V = \Delta x \Delta y \Delta z$ . As a result of the infinitesimal displacements  $\Delta u$ ,  $\Delta v$  and  $\Delta w$  the edges increase to  $\Delta x + \Delta u$ ,  $\Delta y + \Delta v$ , and  $\Delta z + \Delta w$ , respectively. The fractional change in volume is

$$\begin{aligned} \Delta V &= \frac{(\Delta x + \Delta u)(\Delta y + \Delta v)(\Delta z + \Delta w) - \Delta x \Delta y \Delta z}{\Delta x \Delta y \Delta z} \\ &= \frac{\Delta x \Delta y \Delta z + \Delta u \Delta y \Delta z + \Delta v \Delta z \Delta x + \Delta w \Delta x \Delta y - \Delta x \Delta y \Delta z}{\Delta x \Delta y \Delta z} \\ &= \frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta y} + \frac{\Delta w}{\Delta z} \end{aligned} \quad (3.7)$$

where very small quantities like  $\Delta u \Delta v$ ,  $\Delta v \Delta w$ ,  $\Delta w \Delta u$  and  $\Delta u \Delta v \Delta w$  have been ignored. In the limit, as  $\Delta x$ ,  $\Delta y$  and  $\Delta z$  all approach zero, we get the dilatation

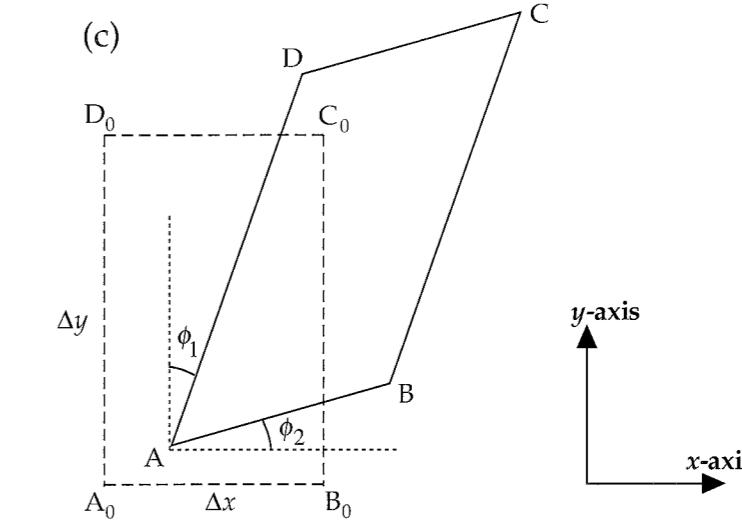
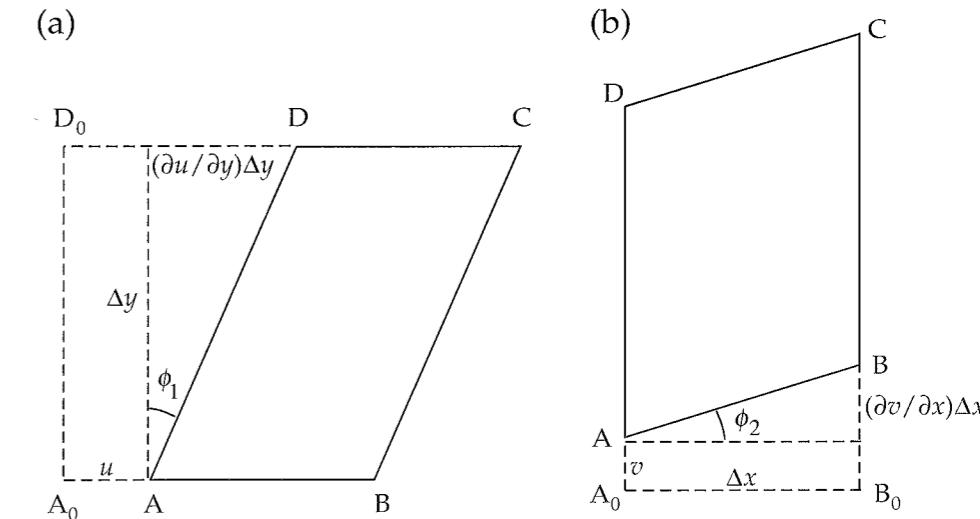
$$\begin{aligned} \theta &= \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \\ \theta &= \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} \end{aligned} \quad (3.8)$$

### 3.2.3.3 Shear strain

During deformation a body generally experiences not only longitudinal strains as described above. The shear components of stress ( $\sigma_{xy}$ ,  $\sigma_{yz}$ ,  $\sigma_{zx}$ ) produce *shear strains*, which are manifest as changes in the angular relationships between parts of a body. This is most easily illustrated in two dimensions. Consider a rectangle ABCD with sides  $\Delta x$  and  $\Delta y$  and its distortion due to shear stresses acting in the  $x$ - $y$  plane (Fig. 3.6). As in the earlier example of longitudinal strain, the point A is displaced parallel to the  $x$ -axis by an amount  $u$  (Fig. 3.6a). Because of the shear deformation, points between A and D experience larger  $x$ -displacements the further they are from A.

## 3.2 ELASTICITY THEORY

**Fig. 3.6** (a) When a square is sheared parallel to the  $x$ -axis, side AD parallel to the  $y$ -axis rotates through a small angle  $\phi_1$ ; (b) when it is sheared parallel to the  $y$ -axis, side AB parallel to the  $x$ -axis rotates through a small angle  $\phi_2$ . In general, shear causes both sides to rotate, giving a total angular deformation ( $\phi_1 + \phi_2$ ). In each case the diagonal AC is extended.



The point D which is at a vertical distance  $\Delta y$  above A is displaced by the amount  $(\partial u / \partial y) \Delta y$  in the direction of the  $x$ -axis. This causes a clockwise rotation of side AD through a small angle  $\phi_1$  given by

$$\tan \phi_1 = \frac{(\partial u / \partial y) \Delta y}{\Delta x} = \frac{\partial u}{\partial y} \quad (3.9)$$

Similarly, the point A is displaced parallel to the  $y$ -axis by an amount  $v$  (Fig. 3.6b), while the point B which is at a horizontal distance  $\Delta x$  from A is displaced by the amount  $(\partial v / \partial x) \Delta x$  in the direction of the  $y$ -axis. As a result side AB rotates counterclockwise through a small angle  $\phi_2$  given by

$$\tan \phi_2 = \frac{(\partial v / \partial x) \Delta x}{\Delta y} = \frac{\partial v}{\partial x} \quad (3.10)$$

Elastic deformation involves infinitesimally small displacements and distortions, and for small angles we can write  $\tan \phi_1 = \phi_1$  and  $\tan \phi_2 = \phi_2$ . The shear strain in the

$x$ - $y$  plane ( $\varepsilon_{xy}$ ) is defined as half the total angular distortion (Fig. 3.6c):

$$\varepsilon_{xy} = \frac{1}{2} \left( \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \quad (3.11)$$

By transposing  $x$  and  $y$ , and the corresponding displacements  $u$  and  $v$ , the shear component  $\varepsilon_{yx}$  is obtained:

$$\varepsilon_{yx} = \frac{1}{2} \left( \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \quad (3.12)$$

This is identical to  $\varepsilon_{xy}$ . The total angular distortion in the  $x$ - $y$  plane is  $(\varepsilon_{xy} + \varepsilon_{yx}) = 2\varepsilon_{xy} = 2\varepsilon_{yx}$ . Similarly, strain components  $\varepsilon_{yz}$  ( $=\varepsilon_{zy}$ ) and  $\varepsilon_{xz}$  ( $=\varepsilon_{zx}$ ) are defined for angular distortions in the  $y$ - $z$  and  $z$ - $x$  planes, respectively.

$$\varepsilon_{yz} = \varepsilon_{zy} = \frac{1}{2} \left( \frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right)$$

$$\varepsilon_{zx} = \varepsilon_{xz} = \frac{1}{2} \left( \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) \quad (3.13)$$

The longitudinal and shear strains define the symmetric  $3 \times 3$  strain matrix

$$\begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} \quad (3.14)$$

### 3.2.4 The elastic constants

According to Hooke's law, when a body deforms elastically, there is a linear relationship between stress and strain. The ratio of stress to strain defines an elastic constant (or elastic modulus) of the body. Strain is itself a ratio of lengths and therefore dimensionless. Thus the elastic moduli must have the units of stress ( $\text{N m}^{-2}$ ). The elastic moduli, defined for different types of deformation, are Young's modulus, the rigidity modulus and the bulk modulus.

*Young's modulus* is defined from the extensional deformations. Each longitudinal strain is proportional to the corresponding stress component, that is,

$$\sigma_{xx} = E\varepsilon_{xx} \quad \sigma_{yy} = E\varepsilon_{yy} \quad \sigma_{zz} = E\varepsilon_{zz} \quad (3.15)$$

where the constant of proportionality,  $E$ , is Young's modulus.

The *rigidity modulus* (or *shear modulus*) is defined from the shear deformation. Like the longitudinal strains, the total shear strain in each plane is proportional to the corresponding shear stress component:

$$\sigma_{xy} = 2\mu\varepsilon_{xy} \quad \sigma_{yz} = 2\mu\varepsilon_{yz} \quad \sigma_{zx} = 2\mu\varepsilon_{zx} \quad (3.16)$$

where the proportionality constant,  $\mu$ , is the rigidity modulus and the factor 2 arises as explained for Eqs. (3.11) and (3.12).

The *bulk modulus* (or *incompressibility*) is defined from the dilatation experienced by a body under hydrostatic pressure. Shear components of stress are zero for hydrostatic conditions ( $\sigma_{xy} = \sigma_{yz} = \sigma_{zx} = 0$ ), and the inwards pressure (negative normal stress) is equal in all directions ( $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = -p$ ). The bulk modulus,  $K$ , is the ratio of the hydrostatic pressure to the dilatation, that is,

$$p = -K\theta \quad (3.17)$$

The inverse of the bulk modulus ( $K^{-1}$ ) is called the *compressibility*.

#### 3.2.4.1 Bulk modulus in terms of Young's modulus and Poisson's ratio

Consider a rectangular volume element subjected to normal stresses  $\sigma_{xx}$ ,  $\sigma_{yy}$  and  $\sigma_{zz}$  on its end surfaces. Each longitudinal strain  $\varepsilon_{xx}$ ,  $\varepsilon_{yy}$  and  $\varepsilon_{zz}$  results from the combined effects of  $\sigma_{xx}$ ,  $\sigma_{yy}$  and  $\sigma_{zz}$ . For example,

applying Hooke's law, the stress  $\sigma_{xx}$  produces an extension equal to  $\sigma_{xx}/E$  in the  $x$ -direction. The stress  $\sigma_{yy}$  causes an extension  $\sigma_{yy}/E$  in the  $y$ -direction, which results in an accompanying transverse strain  $-\nu(\sigma_{yy}/E)$  in the  $x$ -direction, where  $\nu$  is Poisson's ratio. Similarly, the stress component  $\sigma_{zz}$  makes a contribution  $-\nu(\sigma_{zz}/E)$  to the total longitudinal strain  $\varepsilon_{xx}$  in the  $x$ -direction. Therefore,

$$\varepsilon_{xx} = \frac{\sigma_{xx}}{E} - \nu \frac{\sigma_{yy}}{E} - \nu \frac{\sigma_{zz}}{E} \quad (3.18)$$

Similar equations describe the total longitudinal strains  $\varepsilon_{yy}$  and  $\varepsilon_{zz}$ . They can be rearranged as

$$\begin{aligned} E\varepsilon_{xx} &= \sigma_{xx} - \nu\sigma_{yy} - \nu\sigma_{zz} \\ E\varepsilon_{yy} &= \sigma_{yy} - \nu\sigma_{zz} - \nu\sigma_{xx} \\ E\varepsilon_{zz} &= \sigma_{zz} - \nu\sigma_{xx} - \nu\sigma_{yy} \end{aligned} \quad (3.19)$$

Adding these three equations together we get

$$E(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}) = (1 - 2\nu)(\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) \quad (3.20)$$

Consider now the effect of a constraining hydrostatic pressure,  $p$ , where  $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = -p$ . Using the definition of dilatation ( $\theta$ ) in Eq. (3.8) we get

$$E\theta = (1 - 2\nu)(-3p)$$

$$E = (1 - 2\nu) \left( -3 \frac{p}{\theta} \right) \quad (3.21)$$

from which, using the definition of bulk modulus ( $K$ ) in Eq. (3.17),

$$K = \frac{E}{3(1 - 2\nu)} \quad (3.22)$$

#### 3.2.4.2 Shear modulus in terms of Young's modulus and Poisson's ratio

The relationship between  $\mu$  and  $E$  can be appreciated by considering the shear deformation of a rectangular prism that is infinitely long in one dimension and has a square cross-section in the plane of deformation. The shear causes shortening of one diagonal and extension of the other. Let the length of the side of the square be  $a$  (Fig. 3.7a) and that of its diagonal be  $d_0 (= a\sqrt{2})$ . The small shear through the angle  $\phi$  displaces one corner by the amount ( $a \tan \phi$ ) and stretches the diagonal to the new length  $d$  (Fig. 3.7b), which is given by Pythagoras' theorem:

$$\begin{aligned} d^2 &= a^2 + (a + a \tan \phi)^2 \\ &= a^2 + a^2 + a^2 \tan^2 \phi + 2a^2 \tan \phi \\ &= 2a^2 \left( 1 + \tan \phi + \frac{1}{2} \tan^2 \phi \right) \approx d_0^2 (1 + \phi) \end{aligned} \quad (3.23a)$$

## 3.2 ELASTICITY THEORY

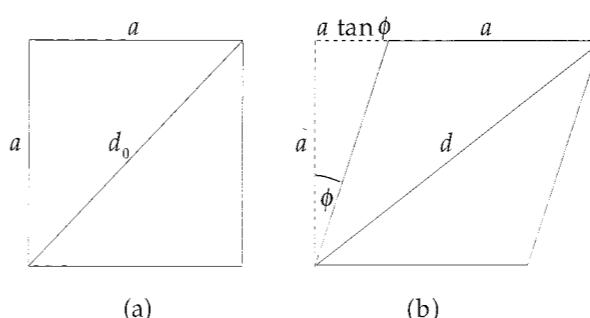


Fig. 3.7 (a) In the undeformed state  $d_0$  is the length of the diagonal of a square with side length  $a$ . (b) When the square is deformed by shear through an angle  $\phi$ , the diagonal is extended to the new length  $d$ .

$$d \approx d_0 \left( 1 + \frac{1}{2}\phi \right) \quad (3.23b)$$

where for an infinitesimally small strain  $\tan \phi = \phi$ , and powers of  $\phi$  higher than first order are negligibly small. The extension of the diagonal is

$$\frac{\Delta d}{d_0} = \frac{d - d_0}{d_0} = \frac{\phi}{2} \quad (3.24)$$

This extension is related to the *normal* stresses  $\sigma_{xx}$  and  $\sigma_{yy}$  in the  $x$ - $y$  plane of the cross-section (Fig. 3.8a), which are in general unequal. Let  $p$  represent their average value:  $p = (\sigma_{xx} + \sigma_{yy})/2$ . The change of shape of the square cross-section results from the differences  $\Delta p$  between  $p$  and  $\sigma_{xx}$  and  $\sigma_{yy}$ , respectively (Fig. 3.8b). The outwards stress difference  $\Delta p$  along the  $x$ -axis produces an  $x$ -extension equal to  $\Delta p/E$ , while the inwards stress difference along the  $y$ -axis causes contraction along the  $y$ -axis and a corresponding contribution to the  $x$ -extension equal to  $\nu(\Delta p/E)$ , where  $\nu$  is Poisson's ratio as before. The total  $x$ -extension  $\Delta x/x$  is therefore given by

$$\frac{\Delta x}{x} = \frac{\Delta p}{E}(1 + \nu) \quad (3.25)$$

Let each edge of the square represent an arbitrary area  $A$  normal to the plane of the figure. The stress differences  $\Delta p$  produce forces  $f = \Delta p A$  on the edges of the square, which resolve to shear forces  $f/\sqrt{2}$  parallel to the sides of the inner square defined by joining the mid-points of the sides of the original square (Fig. 3.8c). Normal to the plane of the figure the surface area represented by each inner side is  $A/\sqrt{2}$ , and therefore the tangential (shear) stress acting on these sides simply equals  $\Delta p$  (Fig. 3.8d). The inner square shears through an angle  $\phi$ , and so we can write

$$\Delta p = \mu\phi \quad (3.26)$$

One diagonal becomes stretched in the  $x$ -direction while the other diagonal is shortened in the  $y$ -direction. The extension of the diagonal of a sheared square was shown above to be  $\phi/2$ . Thus,

$$\frac{\Delta p}{E}(1 + \nu) = \frac{\phi}{2} = \frac{\Delta p}{2\mu} \quad (3.27)$$

Rearranging terms we get the relationship between  $\mu$ ,  $E$  and  $\nu$ :

$$\mu = \frac{E}{2(1 + \nu)} \quad (3.28)$$

#### 3.2.4.3 The Lamé constants

The first line of Eq. (3.19) can be rewritten as

$$E\varepsilon_{xx} = (1 + \nu)\sigma_{xx} - \nu(\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) \quad (3.29)$$

and from Eq. (3.20) we have

$$\begin{aligned} (\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) &= \frac{E}{(1 - 2\nu)}(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}) \\ &= \frac{E}{(1 - 2\nu)}\theta \end{aligned} \quad (3.30)$$

where  $\theta$  is the dilatation, as defined in Eq. (3.8). After substituting Eq. (3.30) in Eq. (3.29) and rearranging we get

$$\sigma_{xx} = \frac{\nu E}{(1 + \nu)(1 - 2\nu)}\theta + \frac{E}{(1 + \nu)}\varepsilon_{xx} \quad (3.31)$$

Writing

$$\lambda = \frac{\nu E}{(1 + \nu)(1 - 2\nu)}$$

and substituting from Eq. (3.28) we can write Eq. (3.31) in the simpler form

$$\sigma_{xx} = \lambda\theta + 2\mu\varepsilon_{xx} \quad (3.32)$$

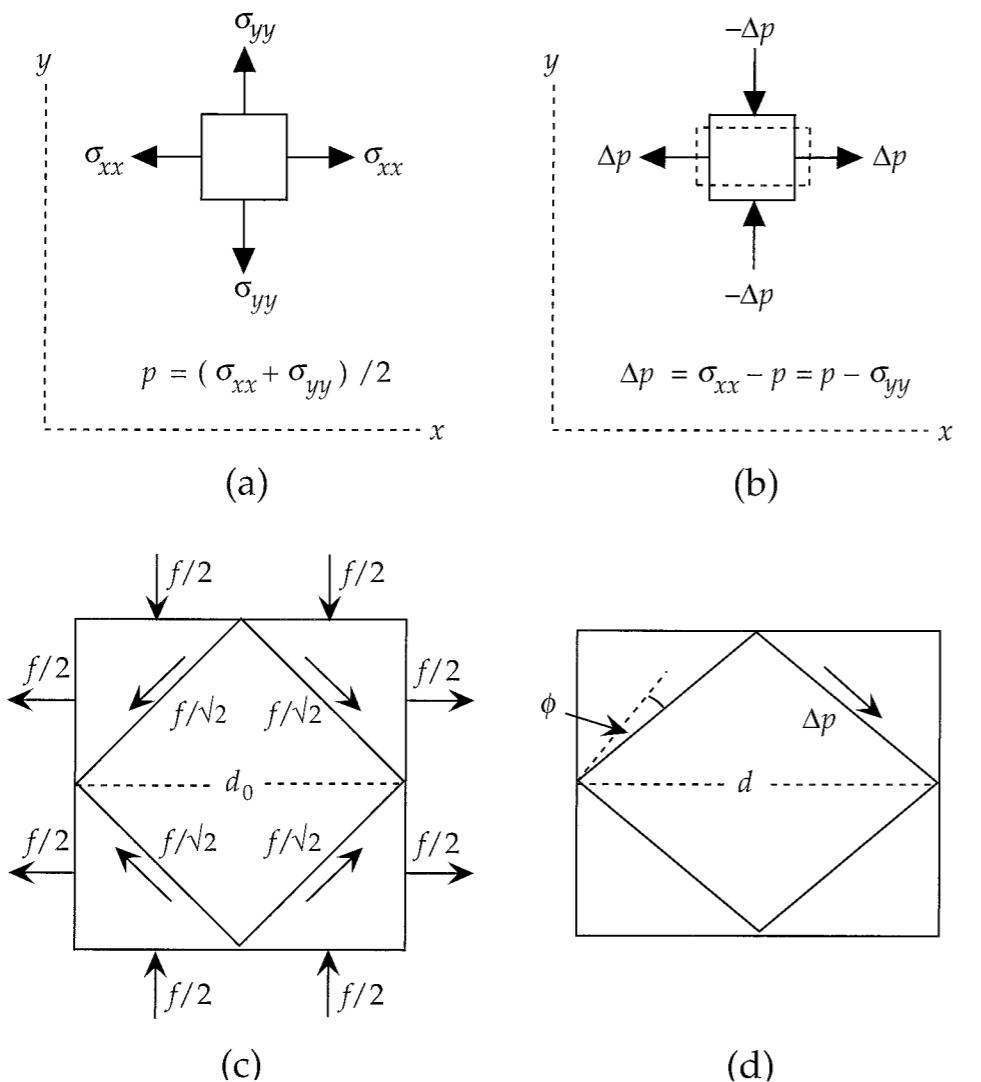
with similar expressions for  $\sigma_{yy}$  and  $\sigma_{zz}$ .

The constants  $\lambda$  and  $\mu$  are known as the Lamé constants. They are related to the elastic constants defined physically above.  $\mu$  is equivalent to the rigidity modulus, while the bulk modulus  $K$ , Young's modulus  $E$  and Poisson's ratio  $\nu$  can each be expressed in terms of both  $\lambda$  and  $\mu$  (Box 3.1).

#### 3.2.4.4 Anisotropy

The foregoing discussion treats the elastic parameters as constants. In fact they are dependent on pressure and temperature and so can only be considered constant for specified conditions. The variations of temperature and pressure in the Earth ensure that the elastic parameters vary with depth. Moreover, it has been assumed that the relationships between stress and strain hold equally for all directions, a property called *isotropy*. This condition is not fulfilled in many minerals. For example, if a mineral has uniaxial symmetry in the arrangement of the atoms in its unit cell, the physical properties of the mineral parallel and perpendicular to the axis of symmetry are different. The mineral is *anisotropic*. The relations between components of stress and strain in an anisotropic substance are

**Fig. 3.8** (a) Unequal normal stresses  $\sigma_{xx}$  and  $\sigma_{yy}$  in the  $x-y$  plane, and their average value  $p$ . (b) Stress differences  $\Delta p$  between  $p$  and  $\sigma_{xx}$  and  $\sigma_{yy}$ , respectively, cause elongation parallel to  $x$  and shortening parallel to  $y$ . (c) Forces  $f = \Delta p A$  along the sides of the original square give shear forces  $f/\sqrt{2}$  along the edges of the inner square, each of which has area  $A\sqrt{2}$ . (d) The shear stress on each side of the inner square has value  $\Delta p$  and causes extension of the diagonal of the inner square and shear deformation through an angle  $\phi$ .



more complex than in the perfectly elastic, isotropic case examined in this chapter. The elastic parameters of an isotropic body are fully specified by the two parameters  $\lambda$  and  $\mu$ , but as many as 21 parameters may be needed to describe anisotropic elastic behavior. Seismic velocities, which depend on the elastic parameters, vary with direction in an anisotropic medium.

Normally, a rock contains so many minerals that it can be assumed that they are oriented at random and the rock can be treated as isotropic. This assumption can also be made, at least to first order, for large regions of the Earth's interior. However, if anisotropic minerals are subjected to stress they develop a preferred alignment with the stress field. For example, platy minerals tend to align with their tabular shapes normal to the compression axis, or parallel to the direction of flow of a fluid. Preferential grain alignment results in seismic anisotropy. This has been observed in seismic studies of the upper mantle, especially at oceanic ridges, where anisotropic velocities have been attributed to the alignment of crystals by convection currents.

Anelastic behavior in the Earth is related to the petrophysical properties of rocks and minerals. If a material is not perfectly elastic, a seismic wave passing through it loses energy to the material (e.g., as frictional heating) and the amplitude of the wave gradually diminishes. The decrease in amplitude is called *attenuation*, and it is due to anelastic *damping* of the vibration of particles of the material (see Section 3.3.2.7). For example, the passage of seismic waves through the asthenosphere is damped owing

### 3.2.5 Imperfect elasticity in the Earth

A seismic wave passes through the Earth as an elastic disturbance of very short duration lasting only some seconds or minutes. Elasticity theory is used to explain seismic wave propagation. However, materials may react differently to brief, sudden stress than they do to long-lasting steady stress. The stress response of rocks and minerals in the Earth is affected by various factors, including temperature, hydrostatic confining pressure, and time. As a result, elastic, anelastic and plastic behavior occur with various degrees of importance at different depths.

The values of  $\lambda$  and  $\mu$  are almost equal in some materials, and it is possible to assume  $\lambda = \mu$ , from which it follows that  $\nu = 0.25$ . Such a material is called a *Poisson solid*.

## 3.2 ELASTICITY THEORY

### Box 3.1: Elastic parameters in terms of the Lamé constants

$$\sigma_{xx} = \lambda \frac{\sigma_{xx}}{(3\lambda + 2\mu)} + 2\mu \epsilon_{xx} \quad (10)$$

Gathering and rearranging terms gives the following succession:

$$\sigma_{xx} \left( 1 - \frac{\lambda}{3\lambda + 2\mu} \right) = 2\mu \epsilon_{xx} \quad (11)$$

$$\sigma_{xx} \left( \frac{\lambda + \mu}{3\lambda + 2\mu} \right) = \mu \epsilon_{xx} \quad (12)$$

$$\sigma_{xx} = \mu \left( \frac{3\lambda + 2\mu}{\lambda + \mu} \right) \epsilon_{xx} \quad (13)$$

The definition of Young's modulus is  $E = \sigma_{xx}/\epsilon_{xx}$  and so in terms of the Lamé constants

$$E = \mu \left( \frac{3\lambda + 2\mu}{\lambda + \mu} \right) \quad (14)$$

### 3. Poisson's ratio ( $\nu$ )

Poisson's ratio is defined as  $\nu = -\epsilon_{yy}/\epsilon_{xx} = -\epsilon_{zz}/\epsilon_{xx}$ . It relates the bulk modulus  $K$  and Young's modulus  $E$  as developed in Eq. (3.22):

$$K = \frac{E}{3(1 - 2\nu)} \quad (15)$$

Substituting the expressions derived above for  $K$  and  $E$  we get

$$\frac{3\lambda + 2\mu}{3} = \frac{1}{3(1 - 2\nu)} \mu \left( \frac{3\lambda + 2\mu}{\lambda + \mu} \right) \quad (16)$$

Rearranging terms leads to the expression for Poisson's ratio  $\nu$  in terms of the Lamé constants:

$$\frac{\lambda + \mu}{\mu} = \frac{1}{(1 - 2\nu)} \quad (17)$$

$$(1 - 2\nu) = \frac{\mu}{\lambda + \mu} \quad (18)$$

$$\nu = \frac{\lambda}{2(\lambda + \mu)} \quad (19)$$

A material that reacts elastically to a sudden stress may deform and flow plastically under a stress that acts over a long time interval. Plastic behavior in the asthenosphere and in the deeper mantle may allow material to

to anelastic behavior at the grain level of the minerals. This may consist of time-dependent slippage between grains; alternatively, fluid phases may be present at the grain boundaries.

This expression is now substituted in Eq. (6a), which becomes

$$\sigma_{xx} = \lambda \theta + 2\mu \epsilon_{xx} \quad (6a)$$

$$0 = \lambda \theta + 2\mu \epsilon_{yy} \quad (6b)$$

$$0 = \lambda \theta + 2\mu \epsilon_{zz} \quad (6c)$$

Adding equations (6a), (6b), and (6c) together gives

$$\sigma_{xx} = 3\lambda \theta + 2\mu (\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz}) = 3\lambda \theta + 2\mu \theta \quad (7)$$

$$\sigma_{xx} = (3\lambda + 2\mu) \theta \quad (8)$$

$$\theta = \frac{\sigma_{xx}}{(3\lambda + 2\mu)} \quad (9)$$

flow, perhaps due to the motion of dislocations within crystal grains. The flow takes place over times on the order of hundreds of millions of years, but it provides an efficient means of transporting heat out of the deep interior.

### 3.3 SEISMIC WAVES

#### 3.3.1 Introduction

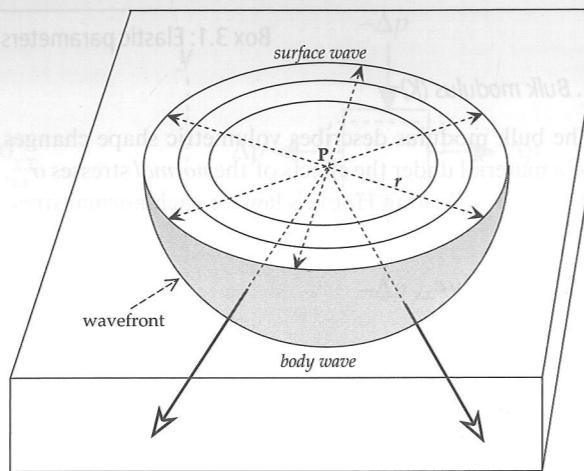
The propagation of a seismic disturbance through a heterogeneous medium is extremely complex. In order to derive equations that describe the propagation adequately, it is necessary to make simplifying assumptions. The heterogeneity of the medium is often modelled by dividing it into parallel layers, in each of which homogeneous conditions are assumed. By suitable choice of the thickness, density and elastic properties of each layer, the real conditions can be approximated. The most important assumption about the propagation of a seismic disturbance is that it travels by elastic displacements in the medium. This condition certainly does not apply close to the seismic source. In or near an earthquake focus or the shot point of a controlled explosion the medium is destroyed. Particles of the medium are displaced permanently from their neighbors; the deformation is anelastic. However, when a seismic disturbance has travelled some distance away from its source, its amplitude decreases and the medium deforms elastically to permit its passage. The particles of the medium carry out simple harmonic motions, and the seismic energy is transmitted as a complex set of wave motions.

When seismic energy is released suddenly at a point P near the surface of a homogeneous medium (Fig. 3.9), part of the energy propagates through the body of the medium as seismic *body waves*. The remaining part of the seismic energy spreads out over the surface as a seismic *surface wave*, analogous to the ripples on the surface of a pool of water into which a stone has been thrown.

#### 3.3.2 Seismic body waves

When a body wave reaches a distance  $r$  from its source in a homogeneous medium, the *wavefront* (defined as the surface in which all particles vibrate with the same phase) has a spherical shape, and the wave is called a *spherical wave*. As the distance from the source increases, the curvature of the spherical wavefront decreases. At great distances from the source the wavefront is so flat that it can be considered to be a plane and the seismic wave is called a *plane wave*.

The direction perpendicular to the wavefront is called the *seismic ray path*. The description of the harmonic motion in plane waves is simpler than for spherical waves, because for plane waves we can use orthogonal Cartesian coordinates. Even for plane waves the mathematical description of the three-dimensional



**Fig. 3.9** Propagation of a seismic disturbance from a point source P near the surface of a homogeneous medium; the disturbance travels as a body wave through the medium and as a surface wave along the free surface.

displacements of the medium is fairly complex. However, we can learn quite a lot about body-wave propagation from a simpler, less rigorous description.

##### 3.3.2.1 Compressional waves

Let Cartesian reference axes be defined such that the  $x$ -axis is parallel to the direction of propagation of the plane wave; the  $y$ - and  $z$ -axes then lie in the plane of the wavefront (Fig. 3.10). A generalized vibration of the medium can be reduced to components parallel to each of the reference axes. In the  $x$ -direction the particle motion is back and forward parallel to the direction of propagation. This results in the medium being alternately stretched and condensed in this direction (Fig. 3.11a). This harmonic motion produces a body wave that is transmitted as a sequence of rarefactions and condensations parallel to the  $x$ -axis.

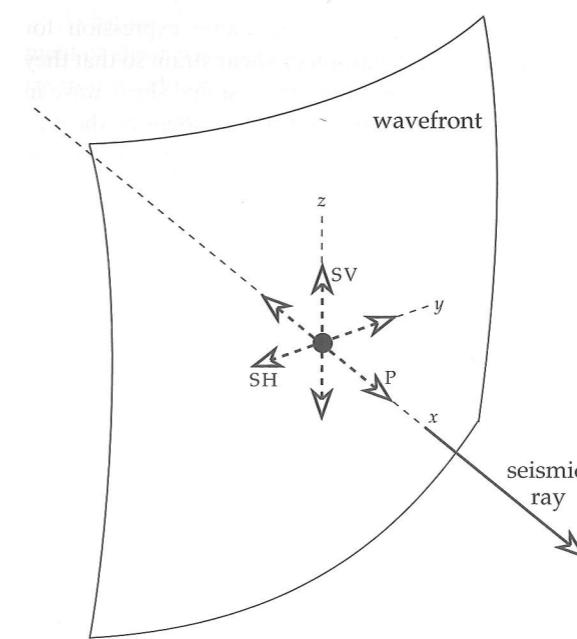
Consider the disturbance of the medium shown in Fig. 3.11b. The area of the wavefront normal to the  $x$ -direction is  $A_x$ , and the wave propagation is treated as one-dimensional. At an arbitrary position  $x$  (Fig. 3.11c), the passage of the wave produces a displacement  $u$  and a force  $F_x$  in the  $x$ -direction. At the position  $x + dx$  the displacement is  $u + du$  and the force is  $F_x + dF_x$ . Here  $dx$  is the infinitesimal length of a small volume element which has mass  $\rho dx A_x$ . The net force acting on this element in the  $x$ -direction is given by

$$(F_x + dF_x) - F_x = dF_x = \frac{\partial F_x}{\partial x} dx \quad (3.33)$$

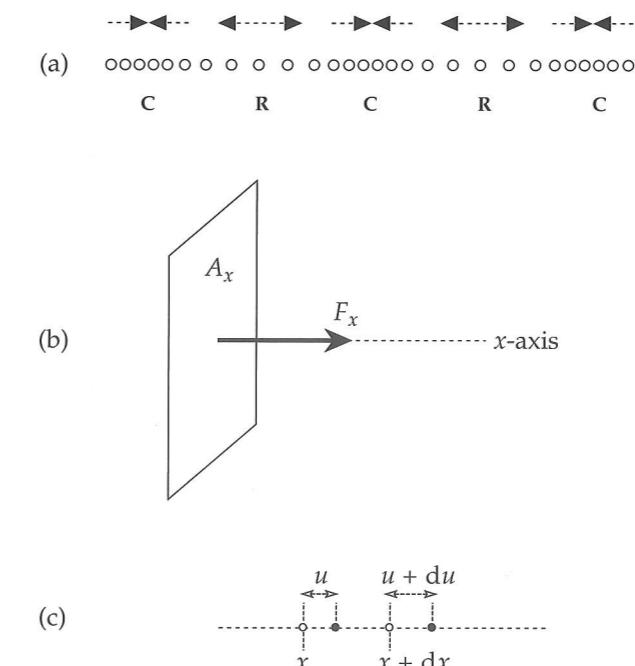
The force  $F_x$  is caused by the stress element  $\sigma_{xx}$  acting on the area  $A_x$ , and is equal to  $\sigma_{xx} A_x$ . This allows us to write the one-dimensional equation of motion

$$(\rho dx A_x) \frac{\partial^2 u}{\partial t^2} = dx A_x \frac{\partial \sigma_{xx}}{\partial x} \quad (3.34)$$

### 3.3 SEISMIC WAVES



**Fig. 3.10** Representation of a generalized vibration as components parallel to three orthogonal reference axes. Particle motion in the  $x$ -direction is back and forth parallel to the direction of propagation, corresponding to the P-wave. Vibrations along the  $y$ - and  $z$ -axes are in the plane of the wavefront and normal to the direction of propagation. The  $z$ -vibration in a vertical plane corresponds to the SV-wave; the  $y$ -vibration is horizontal and corresponds to the SH-wave.



**Fig. 3.11** (a) The particle motion in a one-dimensional P-wave transmits energy as a sequence of rarefactions (R) and condensations (C) parallel to the  $x$ -axis. (b) Within the wavefront the component of force  $F_x$  in the  $x$ -direction of propagation is distributed over an element of area  $A_x$  normal to the  $x$ -axis. (c) A particle at position  $x$  experiences a longitudinal displacement  $u$  in the  $x$ -direction, while at the nearby position  $x + dx$  the corresponding displacement is  $u + du$ .

The definitions of Young's modulus,  $E$ , (Eq. (3.15)) and the normal strain  $\varepsilon_{xx}$  (Eq. (3.4)) give, for a one-dimensional deformation

$$\sigma_{xx} = E \varepsilon_{xx} = E \frac{\partial u}{\partial x} \quad (3.35)$$

Substitution of Eq. (3.35) into Eq. (3.34) gives the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = V^2 \frac{\partial^2 u}{\partial x^2} \quad (3.36)$$

where  $V$  is the velocity of the wave, given by

$$V = \sqrt{\frac{E}{\rho}} \quad (3.37)$$

A one-dimensional wave is rather restrictive. It represents the stretching and compressing in the  $x$ -direction as effects that are independent of what happens in the  $y$ - and  $z$ -directions. In an elastic solid the elastic strains in any direction are coupled to the strains in transverse directions by Poisson's ratio for the medium. A three-dimensional analysis is given in Appendix A that takes into account the simultaneous changes perpendicular to the direction of propagation. In this case the area  $A_x$  can no longer be considered constant. Instead of looking at the displacements in one direction only, all three axes must be taken into account. This is achieved by analyzing the changes in volume. The longitudinal (or compressional) body wave passes through a medium as a series of *dilatations* and *compressions*. The equation of the compressional wave in the  $x$ -direction is

$$\frac{\partial^2 \theta}{\partial t^2} = \alpha^2 \frac{\partial^2 \theta}{\partial x^2} \quad (3.38)$$

where  $\alpha$  is the wave velocity and is given by

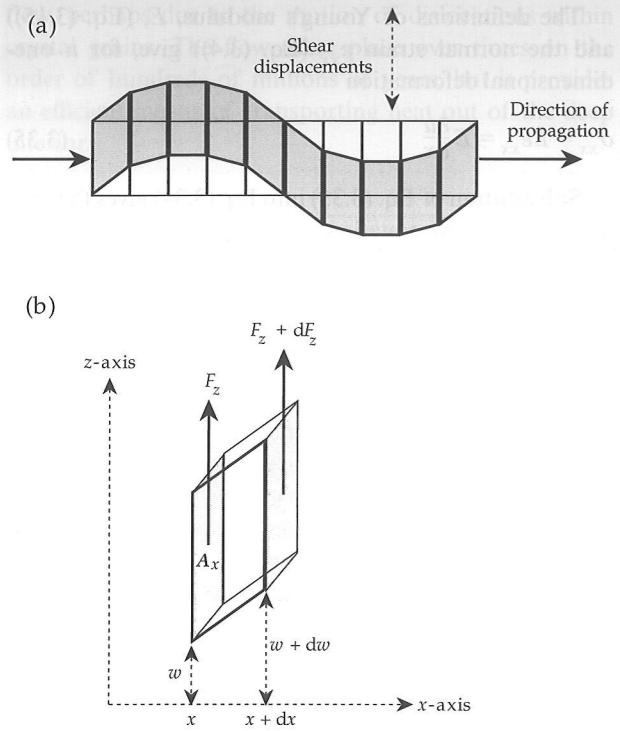
$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}} = \sqrt{\frac{K + \frac{4}{3}\mu}{\rho}} \quad (3.39)$$

The longitudinal wave is the fastest of all seismic waves. When an earthquake occurs, this wave is the first to arrive at a recording station. As a result it is called the *primary wave*, or *P-wave*. Eq. (3.39) shows that P-waves can travel through solids, liquids and gases, all of which are compressible ( $K \neq 0$ ). Liquids and gases do not allow shear. Consequently,  $\mu = 0$ , and the compressional wave velocity in a liquid or gas is given by

$$\alpha = \sqrt{\frac{K}{\rho}} \quad (3.40)$$

##### 3.3.2.2 Transverse waves

The vibrations along the  $y$ - and  $z$ -axes (Fig. 3.10) are parallel to the wavefront and transverse to the direction of propagation. If we wish, we can combine the  $y$ - and  $z$ -components into a single transverse motion. It is more



**Fig. 3.12** (a) Shear distortion caused by the passage of a one-dimensional S-wave. (b) Displacements and forces in the z-direction at the positions  $x$  and  $x+dx$  bounding a small sheared element.

convenient, however, to analyze the motions in the vertical and horizontal planes separately. Here we discuss the disturbance in the vertical plane defined by the  $x$ - and  $z$ -axes; an analogous description applies to the horizontal plane.

The transverse wave motion is akin to that seen when a rope is shaken. Vertical planes move up and down and adjacent elements of the medium experience shape distortions (Fig. 3.12a), changing repeatedly from a rectangle to a parallelogram and back. Adjacent elements of the medium suffer vertical shear.

Consider the distortion of an element bounded by vertical planes separated by a small horizontal distance  $dx$  (Fig. 3.12b) at an arbitrary horizontal position  $x$ . The passage of a wave in the  $x$ -direction produces a displacement  $w$  and a force  $F_z$  in the  $z$ -direction. At the position  $x+dx$  the displacement is  $w+dw$  and the force is  $F_z+dF_z$ . The mass of the small volume element bounded by the vertical planes is  $\rho dx A_x$ , where  $A_x$  is the area of the bounding plane. The net force acting on this element in the  $z$ -direction is given by

$$(F_z + dF_z) - F_z = dF_z = \frac{\partial F_z}{\partial x} dx \quad (3.41)$$

The force  $F_z$  arises from the shear stress  $\sigma_{xz}$  on the area  $A_x$ , and is equal to  $\sigma_{xz} A_x$ . The equation of motion of the vertically sheared element is

$$\left( \rho dx A_x \right) \frac{\partial^2 w}{\partial t^2} = dx A_x \frac{\partial \sigma_{xz}}{\partial x} \quad (3.42)$$

We now have to modify the Lamé expression for Hooke's law and the definition of shear strain so that they apply to the passage of a one-dimensional shear wave in the  $x$ -direction. In this case, because the areas of the parallelograms between adjacent vertical planes are equal, there is no volume change. The dilatation  $\theta$  is zero, and Hooke's law is as given in Eq. (3.16):

$$\sigma_{xz} = 2\mu \epsilon_{xz} \quad (3.43)$$

Following the definition of shear-strain components in Eq. (3.12) we have

$$\epsilon_{xz} = \frac{1}{2} \left( \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) \quad (3.44)$$

For a one-dimensional shear wave there is no change in the distance  $dx$  between the vertical planes;  $du$  and  $\partial u / \partial z$  are zero and  $\epsilon_{xz}$  is equal to  $(\partial w / \partial x) / 2$ . On substitution into Eq. (3.43) this gives

$$\sigma_{xz} = \mu \frac{\partial w}{\partial x} \quad (3.45)$$

and on further substitution into Eq. (3.42) and rearrangement of terms we get

$$\frac{\partial^2 w}{\partial t^2} = \beta^2 \frac{\partial^2 w}{\partial x^2} \quad (3.46)$$

where  $\beta$  is the velocity of the shear wave, given by

$$\beta = \sqrt{\frac{\mu}{\rho}} \quad (3.47)$$

The only elastic property that determines the velocity of the shear wave is the rigidity or shear modulus,  $\mu$ . In liquids and gases  $\mu$  is zero and shear waves cannot propagate. In solids, a quick comparison of Eqs. (3.39) and (3.47) gives

$$\alpha^2 - \frac{4}{3} \beta^2 = \frac{K}{\rho} \quad (3.48)$$

By definition, the bulk modulus  $K$  is positive (if it were negative, an increase in confining pressure would cause an increase in volume), and therefore  $\alpha$  is always greater than  $\beta$ . Shear waves from an earthquake travel more slowly than P-waves and are recorded at an observation station as later arrivals. Shear waves are often referred to as *secondary waves* or *S-waves*.

The general shear-wave motion within the plane of the wavefront can be resolved into two orthogonal components, one being horizontal and the other lying in the vertical plane containing the ray path (Fig. 3.10). Equation (3.46) describes a one-dimensional shear wave which travels in the  $x$ -direction, but which has particle displacements ( $w$ ) in the  $z$ -direction. This wave can be considered to be polarized in the vertical plane. It is called the *SV-wave*. A similar equation describes the shear wave in the  $x$ -direction with particle displacements ( $v$ ) in the  $y$ -direction. A shear wave that is polarized in the horizontal plane is called an *SH-wave*.

The force  $F_z$  arises from the shear stress  $\sigma_{xz}$  on the area  $A_x$ , and is equal to  $\sigma_{xz} A_x$ . The equation of motion of the vertically sheared element is

$$\left( \rho dx A_x \right) \frac{\partial^2 w}{\partial t^2} = dx A_x \frac{\partial \sigma_{xz}}{\partial x} \quad (3.42)$$

### 3.3 SEISMIC WAVES

As for the description of longitudinal waves, this treatment of shear-wave transmission is over-simplified; a more rigorous treatment is given in Appendix A. The passage of a shear wave involves rotations of volume elements within the plane normal to the ray path, without changing their volume. For this reason, shear waves are also sometimes called rotational (or equivoluminal) waves. The rotation is a vector,  $\psi$ , with  $x$ -,  $y$ - and  $z$ -components given by

$$\psi_x = \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \quad \psi_y = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \quad \psi_z = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad (3.49)$$

A more appropriate equation for the shear wave in the  $x$ -direction is then

$$\frac{\partial^2 \psi}{\partial t^2} = \beta^2 \frac{\partial^2 \psi}{\partial x^2} \quad (3.50)$$

where  $\beta$  is again the shear-wave velocity as given by Eq. (3.47).

Until now we have chosen the direction of propagation along one of the reference axes so as to simplify the mathematics. If we remove this restriction, additional second-order differentiations with respect to the  $y$ - and  $z$ -coordinates must be introduced. The P-wave and S-wave equations become, respectively,

$$\frac{\partial^2 \theta}{\partial t^2} = \alpha^2 \left( \frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} + \frac{\partial^2 \theta}{\partial z^2} \right) \quad (3.51)$$

$$\frac{\partial^2 \psi}{\partial t^2} = \beta^2 \left( \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) \quad (3.52)$$

#### 3.3.2.3 The solution of the seismic wave equation

Two important characteristics of a wave motion are: (1) it transmits energy by means of elastic displacements of the particles of the medium, i.e., there is no net transfer of mass, and (2) the wave pattern repeats itself in both time and space. The harmonic repetition allows us to express the amplitude variation by a *sine* or *cosine* function. As the wave passes any point, the amplitude of the disturbance is repeated at regular time intervals,  $T$ , the *period* of the wave. The number of times the amplitude is repeated per second is the *frequency*,  $f$ , which is equal to the inverse of the period ( $f = 1/T$ ). At any instant in time, the disturbance in the medium is repeated along the direction of travel at regular distances,  $\lambda$ , the *wavelength* of the wave. During the passage of a P-wave in the  $x$ -direction, the harmonic displacement ( $u$ ) of a particle from its mean position can be written

$$u = A \sin 2\pi \left( \frac{x}{\lambda} - \frac{t}{T} \right) \quad (3.53)$$

where  $A$  is the amplitude.

The quantity in brackets is called the *phase* of the wave. Any value of the phase corresponds to a particular amplitude and direction of motion of the particles of the

medium. The *wave number* ( $k$ ), *angular frequency* ( $\omega$ ) and *velocity* ( $c$ ) are defined and related by

$$k = \frac{2\pi}{\lambda} \quad \omega = 2\pi f = \frac{2\pi}{T} \quad c = \lambda f = \frac{\omega}{k} \quad (3.54)$$

Equation (3.53) for the displacement ( $u$ ) can then be written

$$u = A \sin(kx - \omega t) = A \sin(k(x - ct)) \quad (3.55)$$

The velocity  $c$  introduced here is called the *phase velocity*. It is the velocity with which a constant phase (e.g., the “peak” or “trough,” or one of the zero displacements) is transmitted. This can be seen by equating the phase to a constant and then differentiating the expression with respect to time, as follows:

$$kx - \omega t = \text{constant} \quad (3.56)$$

$$\begin{aligned} k \frac{dx}{dt} - \omega &= 0 \\ \frac{dx}{dt} &= \frac{\omega}{k} \end{aligned} \quad (3.56)$$

To demonstrate that the displacement given by Eq. (3.55) is a solution of the one-dimensional wave equation (Eq. (3.38)) we must partially differentiate  $u$  in Eq. (3.55) twice with respect to time ( $t$ ) and twice with respect to position ( $x$ ):

$$\begin{aligned} \frac{\partial u}{\partial x} &= Ak \cos(kx - \omega t) \\ \frac{\partial^2 u}{\partial x^2} &= -Ak^2 \sin(kx - \omega t) = -k^2 u \end{aligned} \quad (3.57)$$

$$\begin{aligned} \frac{\partial u}{\partial t} &= -A\omega \cos(kx - \omega t) \\ \frac{\partial^2 u}{\partial t^2} &= -A\omega^2 \sin(kx - \omega t) = -\omega^2 u \end{aligned} \quad (3.57)$$

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2} &= \frac{\omega^2 \partial^2 u}{k^2 \partial x^2} = c^2 \frac{\partial^2 u}{\partial x^2} \end{aligned} \quad (3.57)$$

For a P-wave travelling along the  $x$ -axis the dilatation  $\theta$  is given by an equation similar to Eq. (3.57), with substitution of the P-wave velocity ( $\alpha$ ) for the velocity  $c$ . Similarly, for an S-wave along the  $x$ -axis the rotation  $\psi$  is given by an equation like Eq. (3.57) with appropriate substitutions of  $\psi$  for  $u$  and the S-wave velocity ( $\beta$ ) for the velocity  $c$ . However, in general, the solutions of the three-dimensional compressional and shear wave equations (Eqs. (3.51) and (3.52), respectively) are considerably more complicated than those given by Eq. (3.55).

#### 3.3.2.4 D'Alembert's principle

Equation (3.55) describes the particle displacement during the passage of a wave that is travelling in the direction of the positive  $x$ -axis with velocity  $c$ . Because the velocity enters the wave equation as  $c^2$ , the one-dimensional wave equation is also satisfied by the displacement

$$u = B \sin k(x + ct) \quad (3.58)$$

which corresponds to a wave travelling with velocity  $c$  in the direction of the negative  $x$ -axis.

In fact, any function of  $(x \pm ct)$  that is itself continuous and that has continuous first and second derivatives is a solution of the one-dimensional wave equation. This is known as *D'Alembert's principle*. It can be simply demonstrated for the function  $F = f(x - ct) = f(\phi)$  as follows:

$$\begin{aligned}\frac{\partial F}{\partial x} &= \frac{\partial F}{\partial \phi} \frac{\partial \phi}{\partial x} = \frac{\partial F}{\partial \phi} & \frac{\partial F}{\partial t} &= \frac{\partial F}{\partial \phi} \frac{\partial \phi}{\partial t} = -c \frac{\partial F}{\partial \phi} \\ \frac{\partial^2 F}{\partial x^2} &= \frac{\partial}{\partial x} \frac{\partial F}{\partial x} = \frac{\partial}{\partial x} \frac{\partial \phi}{\partial \phi} \frac{\partial F}{\partial x} = \frac{\partial^2 F}{\partial \phi^2} \\ \frac{\partial^2 F}{\partial t^2} &= \frac{\partial}{\partial t} \frac{\partial F}{\partial t} = \frac{\partial}{\partial t} \frac{\partial \phi}{\partial \phi} \frac{\partial F}{\partial t} = -c \frac{\partial}{\partial \phi} \left( -c \frac{\partial F}{\partial \phi} \right) = c^2 \frac{\partial^2 F}{\partial \phi^2} \\ \frac{\partial^2 F}{\partial t^2} &= c^2 \frac{\partial^2 F}{\partial x^2} \end{aligned}\quad (3.59)$$

Because Eq. (3.59) is valid for positive and negative values of  $c$ , its general solution  $F$  represents the superposition of waves travelling in opposite directions along the  $x$ -axis, and is given by

$$F = f(x - ct) + g(x + ct) \quad (3.60)$$

### 3.3.2.5 The eikonal equation

Consider a wave travelling with constant velocity  $c$  along the axis  $x'$  which has direction cosines  $(l, m, n)$ . If  $x'$  is measured from the center of the coordinate axes  $(x, y, z)$  we can substitute  $x' = lx + my + nz$  for  $x$  in Eq. (3.60). If we consider for convenience only the wave travelling in the direction of  $+x'$ , we get as the general solution to the wave equation

$$F = f(lx + my + nz - ct) \quad (3.61)$$

The wave equation is a second-order differential equation. However, the function  $F$  is also a solution of a first-order differential equation. This is seen by differentiating  $F$  with respect to  $x, y, z$ , and  $t$ , respectively, which gives

$$\begin{aligned}\frac{\partial F}{\partial x} &= \frac{\partial F}{\partial \phi} \frac{\partial \phi}{\partial x} = l \frac{\partial F}{\partial \phi} & \frac{\partial F}{\partial y} &= \frac{\partial F}{\partial \phi} \frac{\partial \phi}{\partial y} = m \frac{\partial F}{\partial \phi} \\ \frac{\partial F}{\partial z} &= \frac{\partial F}{\partial \phi} \frac{\partial \phi}{\partial z} = n \frac{\partial F}{\partial \phi} & \frac{\partial F}{\partial t} &= \frac{\partial F}{\partial \phi} \frac{\partial \phi}{\partial t} = -c \frac{\partial F}{\partial \phi}\end{aligned}\quad (3.62)$$

The direction cosines  $(l, m, n)$  are related by  $l^2 + m^2 + n^2 = 1$ , and so, as can be verified by substitution, the expressions in Eq. (3.62) satisfy the equation

$$\left( \frac{\partial F}{\partial x} \right)^2 + \left( \frac{\partial F}{\partial y} \right)^2 + \left( \frac{\partial F}{\partial z} \right)^2 = \left( \frac{1}{c} \right)^2 \left( \frac{\partial F}{\partial t} \right)^2 \quad (3.63)$$

In seismic wave theory, the progress of a wave is described by successive positions of its wavefront, defined as the surface in which all particles at a given instant in time are moving with the same phase. For a particular

value of  $t$  a constant phase of the wave equation solution given by Eq. (3.61) requires that

$$lx + my + nz = \text{constant} \quad (3.64)$$

From analytical geometry we know that Eq. (3.64) represents a family of planes perpendicular to a line with direction cosines  $(l, m, n)$ . We began this discussion by describing a wave moving with velocity  $c$  along the direction  $x'$ , and now we see that this direction is normal to the plane wavefronts. This is the direction that we defined earlier as the ray path of the wave.

In a medium like the Earth the elastic properties and density – and therefore also the velocity – vary with position. The ray path is no longer a straight line and the wavefronts are not planar. Instead of Eq. (3.61) we write

$$F = f[S(x, y, z) - c_0 t] \quad (3.65)$$

where  $S(x, y, z)$  is a function of position only and  $c_0$  is a constant reference velocity. Substitution of Eq. (3.65) into Eq. (3.63) gives

$$\left( \frac{\partial S}{\partial x} \right)^2 + \left( \frac{\partial S}{\partial y} \right)^2 + \left( \frac{\partial S}{\partial z} \right)^2 = \left( \frac{c_0}{c} \right)^2 = \zeta^2 \quad (3.66)$$

where  $\zeta$  is known as the refractive index of the medium. Equation (3.66) is called the *eikonal equation*. It establishes the equivalence of treating seismic wave propagation by describing the wavefronts or the ray paths. The surfaces  $S(x, y, z) = \text{constant}$  represent the wavefronts (no longer planar). The direction cosines of the ray path (normal to the wavefront) are in this case given by

$$\lambda = \zeta \frac{\partial S}{\partial x} \quad \mu = \zeta \frac{\partial S}{\partial y} \quad \nu = \zeta \frac{\partial S}{\partial z} \quad (3.67)$$

### 3.3.2.6 The energy in a seismic disturbance

It is important to distinguish between the velocity with which a seismic disturbance travels through a material and the speed with which the particles of the material vibrate during the passage of the wave. The vibrational speed ( $v_p$ ) is obtained by differentiating Eq. (3.55) with respect to time, which yields

$$v_p = \frac{\partial u}{\partial t} = -\omega A \cos(kx - \omega t) \quad (3.68)$$

The *intensity* or *energy density* of a wave is the energy per unit volume in the wavefront and consists of kinetic and potential energy. The kinetic part is given by

$$I = \frac{1}{2} \rho v_p^2 = \frac{1}{2} \rho \omega^2 A^2 \cos^2(kx - \omega t) \quad (3.69)$$

The energy density averaged over a complete harmonic cycle consists of equal parts of kinetic and potential energy; it is given by

### 3.3 SEISMIC WAVES

$$I_{av} = \frac{1}{2} \rho \omega^2 A^2 \quad (3.70)$$

i.e., the mean intensity of the wave is proportional to the square of its amplitude.

#### 3.3.2.7 Attenuation of seismic waves

The further a seismic signal travels from its source the weaker it becomes. The decrease of amplitude with increasing distance from the source is referred to as *attenuation*. It is partly due to the geometry of propagation of seismic waves, and partly due to anelastic properties of the material through which they travel.

The most important reduction is due to *geometric attenuation*. Consider the seismic body waves generated by a seismic source at a point P on the surface of a uniform half-space (see Fig. 3.9). If there is no energy loss due to friction, the energy ( $E_b$ ) in the wavefront at distance  $r$  from its source is distributed over the surface of a hemisphere with area  $2\pi r^2$ . The *intensity* (or energy density,  $I_b$ ) of the body waves is the energy per unit area of the wavefront, and at distance  $r$  is:

$$I_b(r) = \frac{E_b}{2\pi r^2} \quad (3.71)$$

The surface wave is constricted to spread out laterally. The disturbance affects not only the free surface but extends downwards into the medium to a depth  $d$ , which we can consider to be constant for a given wave (Fig. 3.9). When the wavefront of a surface wave reaches a distance  $r$  from the source, the initial energy ( $E_s$ ) is distributed over a circular cylindrical surface with area  $2\pi r d$ . At a distance  $r$  from its source the intensity of the surface wave is given by:

$$I_s(r) = \frac{E_s}{2\pi r d} \quad (3.72)$$

These equations show that the decrease in intensity of body waves is proportional to  $1/r^2$  while the decrease in surface wave intensity is proportional to  $1/r$ . As shown in Eq. (3.70), the intensity of a wave-form, or harmonic vibration, is proportional to the square of its amplitude. The corresponding amplitude attenuations of body waves and surface waves are proportional to  $1/r$  and  $1/\sqrt{r}$ , respectively. Thus, seismic body waves are attenuated more rapidly than surface waves with increasing distance from the source. This explains why, except for the records of very deep earthquakes that do not generate strong surface waves, the surface-wave train on a seismogram is more prominent than that of the body waves.

Another reason for attenuation is the *absorption* of energy due to imperfect elastic properties. If the particles of a medium do not react perfectly elastically with their neighbors, part of the energy in the wave is lost (reappearing, for example, as frictional heat) instead of being transferred through the medium. This type of attenuation of the seismic wave is referred to as *anelastic damping*.

The damping of seismic waves is described by a parameter called the *quality factor* ( $Q$ ), a concept borrowed from electric circuit theory where it describes the performance of an oscillatory circuit. It is defined as the fractional loss of energy per cycle

$$\frac{2\pi}{Q} = -\frac{\Delta E}{E} \quad (3.73)$$

In this expression  $\Delta E$  is the energy lost in one cycle and  $E$  is the total elastic energy stored in the wave. If we consider the damping of a seismic wave as a function of the distance that it travels, a cycle is represented by the wavelength ( $\lambda$ ) of the wave. Equation (3.73) can be rewritten for this case as

$$\begin{aligned}\frac{2\pi}{Q} &= -\frac{1}{E} \lambda \frac{dE}{dr} \\ \frac{dE}{E} &= -\frac{2\pi dr}{Q \lambda}\end{aligned}\quad (3.74)$$

It is conventional to measure damping by its effect on the amplitude of a seismic signal, because that is what is observed on a seismic record. We have seen that the energy in a wave is proportional to the square of its amplitude  $A$  (Eq. (3.70)). Thus we can write  $dE/E = 2dA/A$  in Eq. (3.74), and on solving we get the damped amplitude of a seismic wave at distance  $r$  from its source:

$$A = A_0 \exp\left(-\frac{\pi r}{Q \lambda}\right) = A_0 \exp\left(-\frac{r}{D}\right) \quad (3.75)$$

In this equation  $D$  is the distance within which the amplitude falls to  $1/e$  (36.8%, or roughly a third) of its original value. The inverse of this distance ( $D^{-1}$ ) is called the *absorption coefficient*. For a given wavelength,  $D$  is proportional to the  $Q$ -factor of the region through which the wave travels. A rock with a high  $Q$ -factor transmits a seismic wave with relatively little energy loss by absorption, and the distance  $D$  is large. For body waves  $D$  is generally of the order of 10,000 km and damping of the waves by absorption is not a very strong effect. It is slightly stronger for seismic surface waves, for which  $D$  is around 5000 km.

Equation (3.75) shows that the damping of a seismic wave is dependent on the  $Q$ -factor of the region of the Earth that the wave has travelled through. In general the  $Q$ -factor for P-waves is higher than the  $Q$ -factor for S-waves. This may indicate that anelastic damping is determined primarily by the shear component of strain. In solids with low rigidity, the shear strain can reach high levels and the damping is greater than in materials with high rigidity. In fluids the  $Q$ -factor is high and damping is low, because shear strains are zero and the seismic wave is purely compressional. The values of  $Q$  are quite variable in the Earth: values of around  $10^2$  are found for the mantle, and around  $10^3$  for P-waves in the liquid core. Because  $Q$  is a measure of the deviation from perfect elasticity, it is also encountered in the theory of natural oscillations of the

Earth, and has an effect on fluctuations of the Earth's free rotation, as in the damping of the Chandler wobble.

It follows from Eq. (3.75) that the absorption coefficient ( $D^{-1}$ ) is inversely proportional to the wavelength  $\lambda$ . Thus the attenuation of a seismic wave by absorption is dependent upon the frequency of the signal. High frequencies are attenuated more rapidly than are lower frequencies. As a result, the frequency spectrum of a seismic signal changes as it travels through the ground. Although the original signal may be a sharp pulse (resulting from a shock or explosion), the preferential loss of high frequencies as it travels away from the source causes the signal to assume a smoother shape. This selective loss of high frequencies by absorption is analogous to removing high frequencies from a sound source using a filter. Because the low frequencies are not affected so markedly, they pass through the ground with less attenuation. The ground acts as a low-pass filter to seismic signals.

### 3.3.3 Seismic surface waves

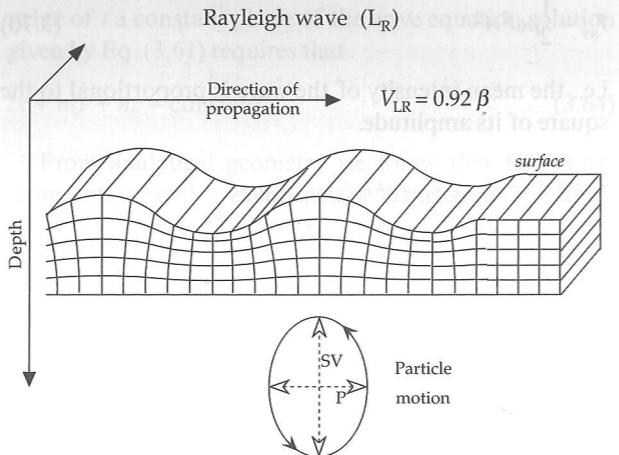
A disturbance at the free surface of a medium propagates away from its source partly as seismic surface waves. Just as seismic body waves can be classified as P- or S-waves, there are two categories of seismic surface waves, sometimes known collectively as L-waves (Section 3.4.4.3), and subdivided into *Rayleigh waves* ( $L_R$ ) and *Love waves* ( $L_Q$ ), which are distinguished from each other by the types of particle motion in their wavefronts. In the description of body waves, the motion of particles in the wavefront was resolved into three orthogonal components – a longitudinal vibration parallel to the ray path (the P-wave motion), a transverse vibration in the vertical plane containing the ray path (the vertical shear or SV-wave) and a horizontal transverse vibration (the horizontal shear or SH-wave). These components of motion, restricted to surface layers, also determine the particle motion and character of the two types of surface waves.

#### 3.3.3.1 Rayleigh waves ( $L_R$ )

In 1885 Lord Rayleigh described the propagation of a surface wave along the free surface of a semi-infinite elastic half-space. The particles in the wavefront of the Rayleigh wave are polarized to vibrate in the vertical plane. The resulting particle motion can be regarded as a combination of the P- and SV-vibrations. If the direction of propagation of the Rayleigh wave is to the right of the viewer (as in Fig. 3.13), the particle motion describes a *retrograde ellipse* in the vertical plane with its major axis vertical and minor axis in the direction of wave propagation.

If Poisson's relation holds for a solid (i.e., Poisson's ratio  $\nu = 0.25$ ) the theory of Rayleigh waves gives a speed ( $V_{LR}$ ) equal to  $\sqrt{2 - 2/\sqrt{3}} = 0.9194$  of the speed ( $\beta$ ) of S-waves (i.e.,  $V_{LR} = 0.9194\beta$ ). This is approximately the case in the Earth.

The particle displacement is not confined entirely to the surface of the medium. Particles below the free



**Fig. 3.13** The particle motion in the wavefront of a Rayleigh wave consists of a combination of P- and SV-vibrations in the vertical plane. The particles move in retrograde sense around an ellipse that has its major axis vertical and minor axis in the direction of wave propagation.

surface are also affected by the passage of the Rayleigh wave; in a uniform half-space the amplitude of the particle displacement decreases exponentially with increasing depth. The penetration depth of the surface wave is typically taken to be the depth at which the amplitude is attenuated to  $(e^{-1})$  of its value at the surface. For Rayleigh waves with wavelength  $\lambda$  the characteristic penetration depth is about  $0.4\lambda$ .

#### 3.3.3.2 Love waves ( $L_Q$ )

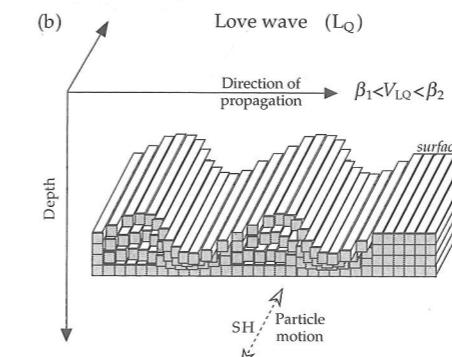
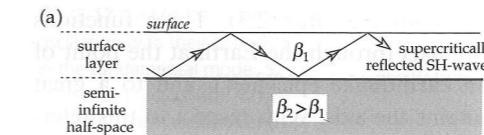
The boundary conditions which govern the components of stress at the free surface of a semi-infinite elastic half-space prohibit the propagation of SH-waves along the surface. However, A. E. H. Love showed in 1911 that if a horizontal layer lies between the free surface and the semi-infinite half-space (Fig. 3.14a), SH-waves within the layer that are reflected at supercritical angles (see Section 3.6) from the top and bottom of the layer can interfere constructively to give a surface wave with horizontal particle motions (Fig. 3.14b). The velocity ( $\beta_1$ ) of S-waves in the near-surface layer must be lower than in the underlying half-space ( $\beta_2$ ). The velocity of the Love waves ( $V_{LQ}$ ) lies between the two extreme values:  $\beta_1 < V_{LQ} < \beta_2$ .

Theory shows that the speed of Love waves with very short wavelengths is close to the slower velocity  $\beta_1$  of the upper layer, while long wavelengths travel at a speed close to the faster velocity  $\beta_2$  of the lower medium. This dependence of velocity on wavelength is termed *dispersion*. Love waves are always dispersive, because they can only propagate in a velocity-layered medium.

#### 3.3.3.3 The dispersion of surface waves

The dispersion of surface waves provides an important tool for determining the vertical velocity structure of the lower crust and upper mantle. Love waves are intrinsically

### 3.3 SEISMIC WAVES



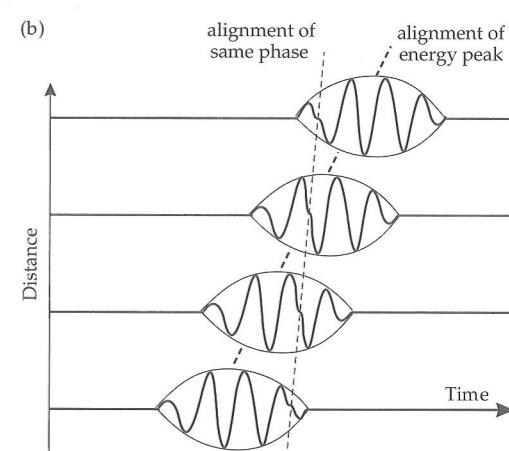
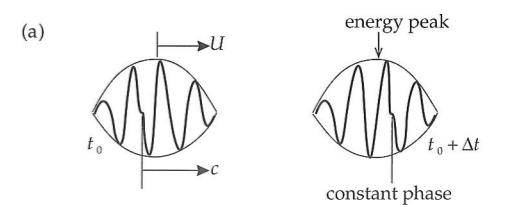
**Fig. 3.14** In a Love wave the particle motion is horizontal and perpendicular to the direction of propagation. The amplitude of the wave decreases with depth below the free surface.

dispersive even when the surface layer and underlying half-space are uniform. Rayleigh waves over a uniform half-space are non-dispersive. However, horizontal layers with different velocities are usually present or there is a vertical velocity gradient. Rayleigh waves with long wavelengths penetrate more deeply into the Earth than those with short wavelengths. The speed of Rayleigh waves is proportional to the shear-wave velocity ( $V_{LR} \approx 0.92\beta$ ), and in the crust and uppermost mantle  $\beta$  generally increases with depth. Thus, the deeper penetrating long wavelengths travel with faster seismic velocities than the short wavelengths. As a result, the Rayleigh waves are dispersive.

The packet of energy that propagates as a surface wave contains a spectrum of wavelengths. The energy in the wave propagates as the envelope of the wave packet (Fig. 3.15a), at a speed that is called the *group velocity* ( $U$ ). The individual waves that make up the wave packet travel with *phase velocity* ( $c$ ), as defined in Eq. (3.56). If the phase velocity is dependent on the wavelength, the group velocity is related to it by

$$U = \frac{\partial \omega}{\partial k} = \frac{\partial (ck)}{\partial k} = c + k \frac{\partial c}{\partial k} = c - \lambda \frac{\partial c}{\partial \lambda} \quad (3.76)$$

The situation in which phase velocity increases with increasing wavelength (i.e., the longer wavelengths propagate faster than the short wavelengths) is called *normal dispersion*. In this case, because  $\partial c / \partial \lambda$  is positive, the group velocity  $U$  is slower than the phase velocity  $c$ . The shape of the wave packet changes systematically as the faster moving long wavelengths pass through the packet (Fig. 3.15b). As time elapses, an initially concentrated pulse becomes progressively stretched out into a long train of waves. Consequently, over a medium in which velocity increases with depth, the long wavelengths arrive as the first part of the surface-wave record at large distances from the seismic source.

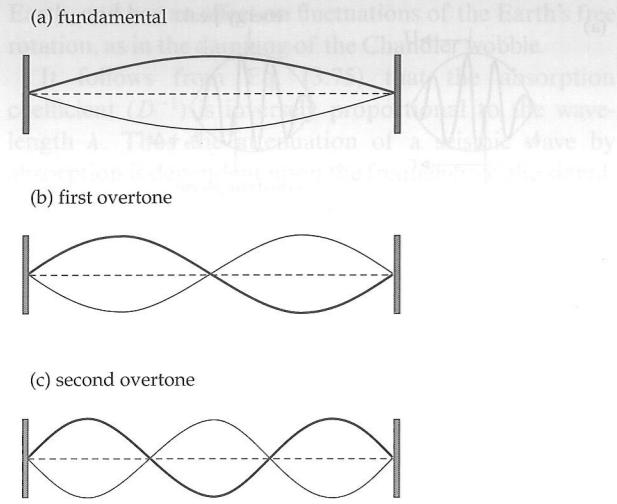


**Fig. 3.15** (a) The surface-wave energy propagates as the envelope of the wave packet with the group velocity  $U$ , while the individual wavelengths travel with the phase velocity  $c$ . (b) Change of shape of a wave packet due to normal dispersion as the faster-moving long wavelengths pass through the packet. At large distances from the source, the long wavelengths arrive as the first part of the surface-wave train (modified from Telford et al., 1990).

### 3.3.4 Free oscillations of the Earth

When a bell is struck with a hammer, it vibrates freely at a number of natural frequencies. The combination of natural oscillations that are excited gives each bell its particular sonority. In an analogous way, the sudden release of energy in a very large earthquake can set the entire Earth into vibration, with natural frequencies of oscillation that are determined by the elastic properties and structure of the Earth's interior. The free oscillations involve three-dimensional deformation of the Earth's spherical shape and can be quite complex. Before discussing the Earth's free oscillations it is worth reviewing some concepts of vibrating systems that can be learned from the one-dimensional excitation of a vibrating string that is fixed at both ends.

Any complicated vibration of the string can be represented by the superposition of a number of simpler vibrations, called the *normal modes* of vibration. These arise when travelling waves reflected from the boundaries at the ends of the string interfere with each other to give a *standing wave*. Each normal mode corresponds to a standing wave with frequency and wavelength determined by the condition that the length of the string must always equal an integral number of half-wavelengths (Fig. 3.16). As well as the fixed ends, there are other points on the string that have zero displacement; these are called the *nodes* of the vibration. The first normal (or *fundamental*)



**Fig. 3.16** Normal modes of vibration for a standing wave on a string fixed at both ends.

mode of vibration has no nodes. The second normal mode (sometimes called the *first overtone*) has one node; its wavelength and period are half those of the fundamental mode. The third normal mode (second overtone) has three times the frequency of the first mode, and so on. Modes with one or more node are called *higher-order* modes.

The concepts of modes and nodes are also applicable to a vibrating sphere. The complex general vibration of a sphere can be resolved into the superposition of a number of normal modes. The nodes of zero displacement become *nodal surfaces* on which the amplitude of the vibration is zero. The free oscillations of the Earth can be divided into three categories. In *radial oscillations* the displacements are purely radial, in *spheroidal oscillations* they are partly radial and partly tangential, and in *toroidal oscillations* they are purely tangential.

#### 3.3.4.1 Radial oscillations

The simplest kind of free oscillations are the *radial oscillations*, in which the shape of the Earth remains “spherical” and all particles vibrate purely radially (Fig. 3.17a). In the fundamental mode of this type of oscillation the entire Earth expands and contracts in unison with a period of about 20.5 minutes. The second normal mode (first overtone) of radial oscillations has a single internal spherical nodal surface. While the inner sphere is contracting, the part outside the nodal surface is expanding, and vice versa. The nodal surfaces of higher modes are also spheres internal to the Earth and concentric with the outer surface.

#### 3.3.4.2 Spheroidal oscillations

A general spheroidal oscillation involves both radial and tangential displacements that can be described by *spheri-*

*cal harmonic functions* (see Box 2.3). These functions are referred to an axis through the Earth at the point of interest (e.g., an earthquake epicenter), and to a great circle which contains the axis. With respect to this reference frame they describe the latitudinal and longitudinal variations of the displacement of a surface from a sphere. They allow complete mathematical description and concise identification of each mode of oscillation with the aid of three indices. The *longitudinal order m* is the number of nodal lines on the sphere that are great circles, the *order l* is determined from the  $(l - m)$  latitudinal nodal lines, and the *overtone number n* describes the number of internal nodal surfaces. The notation  ${}_nS_l^m$  denotes a spheroidal oscillation of order  $l$ , longitudinal order  $m$ , and overtone number  $n$ . In practice, only oscillations with longitudinal order  $m=0$  (rotationally symmetric about the reference axis) are observed, and this index is usually dropped. Also, the oscillation of order  $l=1$  does not exist; it would have only a single equatorial nodal plane and the vibration would involve displacement of the center of gravity. The spheroidal oscillations  ${}_0S_2$  and  ${}_0S_3$  are shown in Fig. 3.17b. Spheroidal oscillations displace the Earth’s surface and alter the internal density distribution. After large earthquakes they produce records on highly sensitive gravity meters used for bodily Earth-tide observations, and also on strain gauges and tilt meters.

The *radial oscillations* can be regarded as a special type of spheroidal oscillation with  $l=0$ . The fundamental radial oscillation is also the fundamental spheroidal oscillation, denoted  ${}_0S_0$ ; the next higher mode is called  ${}_1S_0$  (Fig. 3.17a).

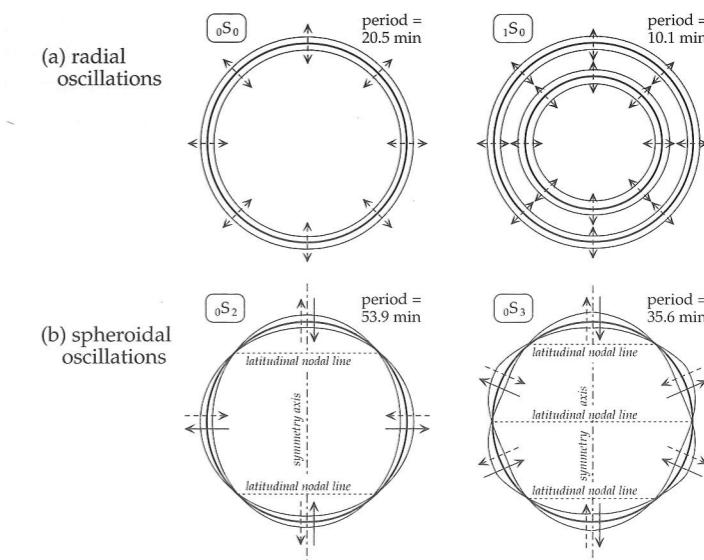
#### 3.3.4.3 Toroidal oscillations

The third category of oscillation is characterized by displacements that are purely tangential. The spherical shape and volume of the Earth are unaffected by a toroidal oscillation, which involves only longitudinal displacements about an axis (Fig. 3.18a). The amplitude of the longitudinal displacement varies with latitude. (Note that, as for spheroidal oscillations, “latitude” and “longitude” refer to the symmetry axis and are different from their geographic definitions.) The toroidal modes have nodal planes that intersect the surface on circles of “latitude,” on which the toroidal displacement is zero. Analogously to the spheroidal oscillations, the notation  ${}_nT_l$  is used to describe the spatial geometry of toroidal modes. The mode  ${}_0T_0$  has zero displacement and  ${}_0T_1$  does not exist, because it describes a constant azimuthal twist of the entire Earth, which would change its angular momentum. The simplest toroidal mode is  ${}_0T_2$ , in which two hemispheres oscillate in opposite senses across a single nodal plane (Fig. 3.18a). Higher toroidal modes of order  $l$  oscillate across  $(l - 1)$  nodal planes perpendicular to the symmetry axis.

The amplitudes of toroidal oscillations inside the Earth change with depth. The displacements of internal

### 3.3 SEISMIC WAVES

**Fig. 3.17** (a) Modes of radial oscillation with their periods: in the fundamental mode  ${}_0S_0$  (also called the balloon mode) the entire Earth expands and contracts in unison; higher modes, such as  ${}_1S_0$ , have internal spherical nodal surfaces concentric with the outer surface. (b) Modes of spheroidal oscillation  ${}_0S_2$  (“football mode”) and  ${}_0S_3$  with their periods; the nodal lines are small circles perpendicular to the symmetry axis.



these oscillations, and they are therefore restricted to the Earth’s rigid mantle and crust.

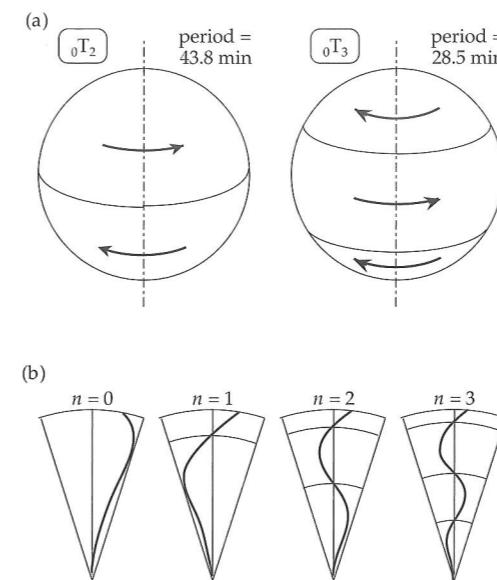
#### 3.3.4.4 Comparison with surface waves

The higher-order free oscillations of the Earth are related directly to the two types of surface wave.

- (i) In a Rayleigh wave the particle vibration is polarized in the vertical plane and has radial and tangential components (see Fig. 3.13). The higher-order *spheroidal oscillations* are equivalent to the standing wave patterns that arise from the interference of trains of long-period Rayleigh waves travelling in opposite directions around the Earth.
- (ii) In a Love wave the particle vibration is polarized horizontally. The *toroidal oscillations* may be regarded as the standing wave patterns due to the interference of oppositely travelling Love waves.

The similarity between surface waves and higher-order natural oscillations of the Earth is evident in the variations of displacement with depth. Like any vibration, a train of surface waves is made up of different modes. Theoretical analysis of surface waves shows that the amplitudes of different modes decay with depth in the Earth (Fig. 3.19) in an equivalent manner to the depth attenuation of natural oscillations (Fig. 3.18b).

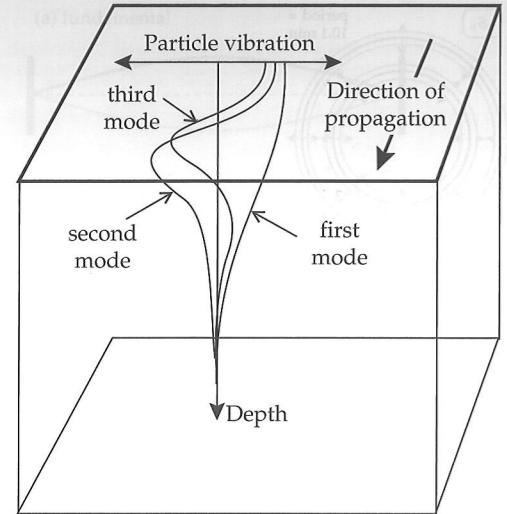
The periods of the normal modes of free oscillations were calculated before they were observed. They have long periods – the period of  ${}_0S_0$  is 20.5 minutes, that of  ${}_0T_2$  is 43.9 minutes and that of  ${}_0S_2$  is 53.9 minutes – and pendulum seismographs are not suitable for recording them. Their recognition had to await the development of long-period seismographs. The spheroidal oscillations have a radial component of displacement and can be recorded with long-period, vertical-motion seismographs. Continually recording gravimeters used for the observation of bodily Earth-tides also record the spheroidal oscillations but not the toroidal oscillations, which have



**Fig. 3.18** (a) The modes of toroidal oscillation  ${}_0T_2$  and  ${}_0T_3$  with their periods; these modes involve oscillation in opposite senses across nodal planes normal to the symmetry axis. (b) In toroidal oscillations, displacements of internal spherical surfaces from their equilibrium positions vary with depth and are zero at internal nodal surfaces.

spherical surfaces from their equilibrium positions are zero at internal nodal surfaces (Fig. 3.18b). The nomenclature for toroidal oscillations also represents these internal nodal surfaces. Thus,  ${}_1T_1$  denotes a general toroidal mode with one internal nodal surface; the internal sphere twists in an opposite sense to the outer spherical shell;  ${}_2T_1$  has two internal nodal surfaces, etc. (Fig. 3.18b).

The twisting motion in toroidal oscillations does not alter the radial density distribution in the Earth, and so they do not show in gravity meter records. They cause changes in strain and displacement parallel to the Earth’s surface and can be recorded by strain meters. Toroidal oscillations are dependent on the shear strength of the Earth’s interior. The Earth’s fluid core cannot take part in



**Fig. 3.19** The attenuation with depth of some low-order modes of Love waves.

no vertical component. These usually must be recorded with an instrument that is sensitive to horizontal displacements, such as the strain meter designed by H. Benioff in 1935. Long-period, horizontal-motion seismographs are capable of recording the toroidal oscillations induced by great earthquakes.

Strain meter records of the November 4, 1952, magnitude 8 earthquake in Kamchatka exhibited a long-period surface wave with a period of 57 minutes (Benioff, 1958). This is much longer than the known periods of travelling surface waves, and was interpreted as a free oscillation of the Earth. Several independent investigators, using bodily Earth-tide gravity meters and different kinds of seismograph, recorded long-period waves excited by the massive 1960 earthquake in Chile (surface-wave magnitude  $M_s = 8.5$ , moment magnitude  $M_w = 9.5$ ). These were conclusively identified with spheroidal and toroidal oscillations.

Figure 3.20 illustrates some of the normal modes of free oscillation of the Earth set up by the huge 2004 Sumatra–Andaman Islands earthquake ( $M_w = 9.0$ ). Vertical motions were recorded during 240 hours by the Canberra, Australia, station of the Geoscope global network of digital broadband seismic stations (see Section 3.4.3.3). The power spectrum in Fig. 3.20 is a plot of the energy associated with different frequencies of vibration. Some spheroidal and toroidal free oscillations are identified and illustrated schematically. The splitting of some normal modes is caused by the non-spherical shape and rotation of the Earth.

The study of the natural oscillations of the Earth set up by large earthquakes is an important branch of seismology, because the normal modes are strongly dependent on the Earth's internal structure. The low-order modes are affected by the entire interior of the Earth, while the higher-order modes react primarily to movements of the upper mantle. The periods of free oscillation are determined by the radial distributions of elastic prop-

erties and densities in the Earth. Comparison of the observed periods of different modes with values computed for different models of the Earth's velocity and density structure provides an important check on the validity of the Earth model. The oscillations are damped by the anelasticity of the Earth. The low-order mode oscillations with periods up to 40–50 minutes persist much longer than the higher-order modes with short periods of only a few minutes. By studying the decay times of different modes, a profile of the anelastic quality factor ( $Q$ ) within the Earth can be calculated that is compatible with results obtained for body waves.

### 3.4 THE SEISMOGRAPH

#### 3.4.1 Introduction

The earliest known instrument for indicating the arrival of a seismic tremor from a distant source is reputed to have been invented by a Chinese astronomer called Chang Heng in 132 AD. The device consisted of eight inverted dragons placed at equal intervals around the rim of a vase. Under each dragon sat an open-mouthed metal toad. Each dragon held a bronze ball in its mouth. When a slight tremor shook the device, an internal mechanism opened the mouth of one dragon, releasing its bronze ball, which fell into the open mouth of the metal toad beneath, thereby marking the direction of arrival of the tremor. The principle of this instrument was used in eighteenth-century European devices that consisted of brimful bowls of water or mercury with grooved rims under which tiny collector bowls were placed to collect the overflow occasioned by a seismic tremor. These instruments gave visible evidence of a seismic event but were unable to trace a permanent record of the seismic wave itself. They are classified as *seismoscopes*.

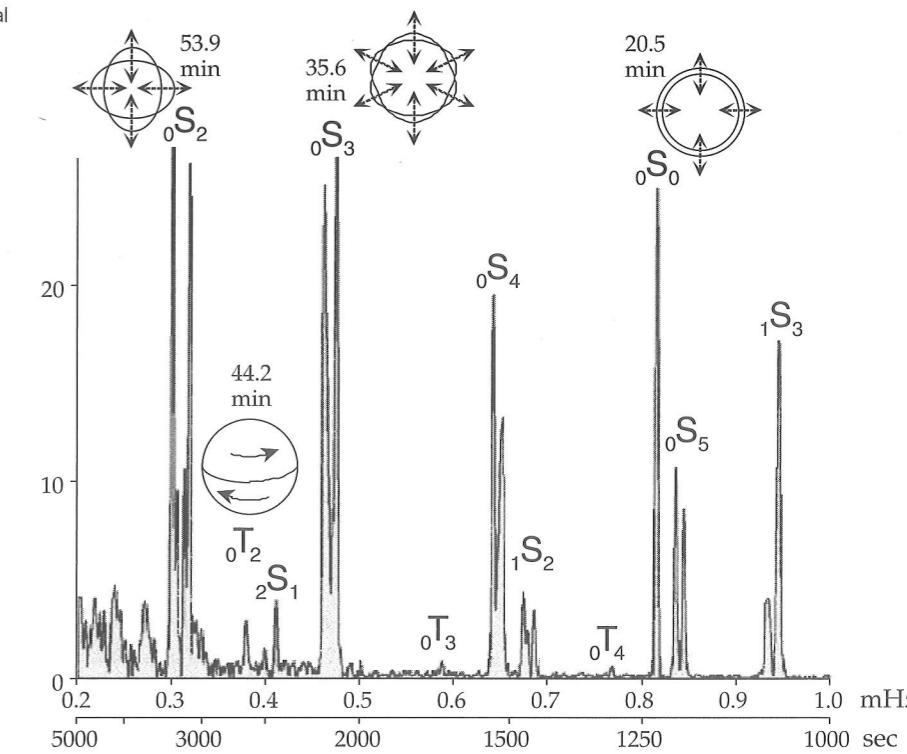
The science of seismology dates from the invention of the *seismograph* by the English scientist John Milne in 1892. Its name derives from its ability to convert an unfelt ground vibration into a visible record. The seismograph consists of a receiver and a recorder. The ground vibration is detected and amplified by a sensor, called the *seismometer* or, in exploration seismology, the *geophone*. In modern instruments the vibration is amplified and filtered electronically. The amplified ground motion is converted to a visible record, called the *seismogram*.

The seismometer makes use of the principle of inertia. If a heavy mass is only loosely coupled to the ground (for example, by suspending it from a wire like a pendulum as in Fig. 3.21), the motion of the Earth caused by a seismic wave is only partly transferred to the mass. While the ground vibrates, the inertia of the heavy mass assures that it does not move as much, if at all. The seismometer amplifies and records the relative motion between the mass and the ground.

Early seismographs were undamped and reacted only to a limited band of seismic frequencies. Seismic waves

### 3.4 THE SEISMOGRAPH

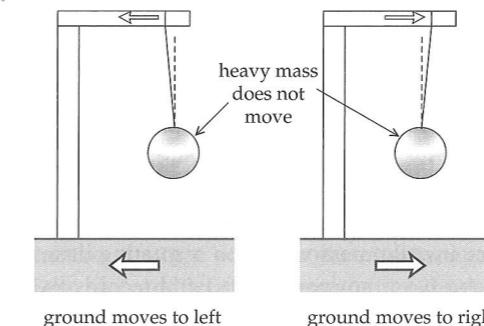
**Fig. 3.20** Spectrum of natural oscillations of the Earth following the magnitude 9.0 Sumatra–Andaman earthquake of December 26, 2004 (after Park et al., 2005).



broadband instruments with digital recording on magnetic tape, hard disk, or solid state memory device.

#### 3.4.2 Principle of the seismometer

Seismometers are designed to react to motion of the Earth in a given direction. Mechanical instruments record the amplified displacement of the ground; electromagnetic instruments respond to the velocity of ground motion. Depending on the design, either type may respond to vertical or horizontal motion. Some modern electromagnetic instruments are constructed so as to record simultaneously three orthogonal components of motion. Most designs employ variations on the pendulum principle.



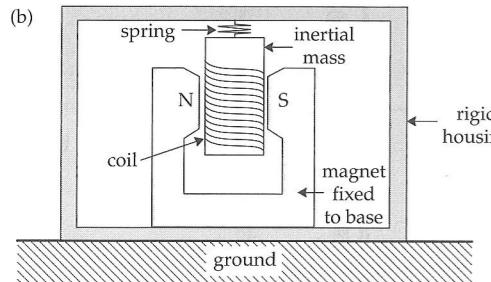
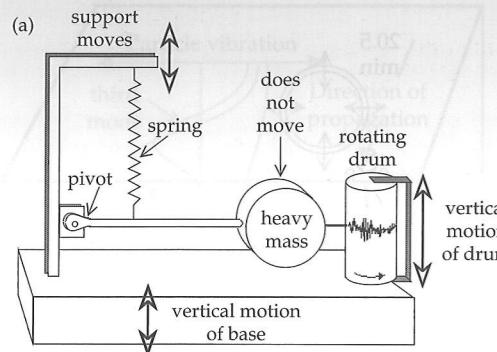
**Fig. 3.21** The principle of the seismometer. Because of its inertia, a suspended heavy mass remains almost stationary when the ground and suspension move to the left or to the right.

with inappropriate frequencies were barely recorded at all, but strong waves could set the instrument into resonant vibration. In 1903, the German seismologist Emil Wiechert substantially increased the accuracy of the seismograph by improving the amplification method and by damping the instrument. These early instruments relied on mechanical levers for amplification and recording signals on smoked paper. This made them both bulky and heavy, which severely restricted their application.

A major technological improvement was achieved in 1906, when Prince Boris Galitzin of Russia introduced the electromagnetic seismometer, which allowed galvanometric recording on photographic paper. This electrical method had the great advantage that the recorder could now be separated from the seismometer. The seismometer has evolved constantly, with improvements in seismometer design and recording method, culminating in modern

##### 3.4.2.1 Vertical-motion seismometer

In the mechanical type of vertical-motion seismometer (Fig. 3.22a), a large mass is mounted on a horizontal bar hinged at a pivot so that it can move only in the vertical plane. A pen attached to the bar writes on a horizontal rotating drum that is fixed to the housing of the instrument. The bar is held in a horizontal position by a weak spring. This assures a loose coupling between the mass and the housing, which is connected rigidly to the ground. Vertical ground motion, as sensed during the passage of a seismic wave, is transmitted to the housing but not to the inertial mass and the pen, which remain stationary. The pen inscribes a trace of the vertical vibration of the housing on a paper fixed to the rotating drum. This trace is the vertical-motion seismogram of the seismic wave.

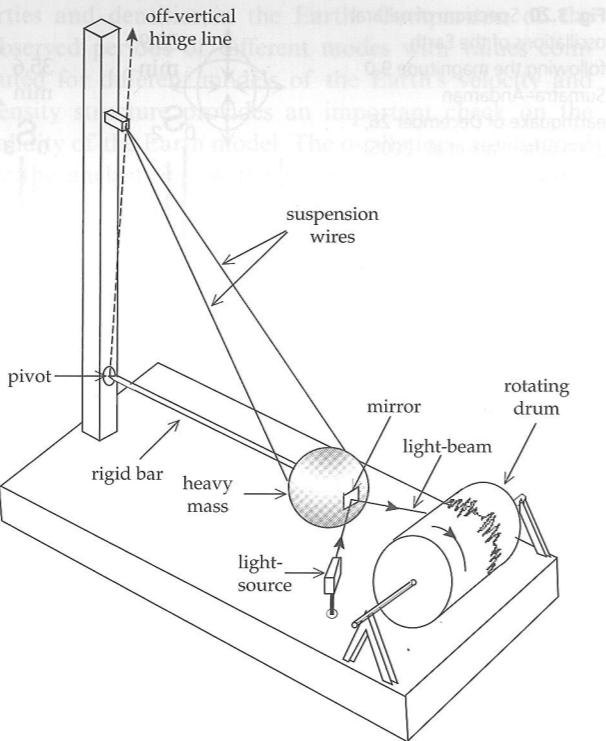


**Fig. 3.22** Schematic diagrams illustrating the principle of operation of the vertical-motion seismometer: (a) mechanical pendulum type (after Strahler, 1963), (b) electromagnetic, moving-coil type.

The electromagnetic seismometer responds to the relative motion between a magnet and a coil of wire. One of these members is fixed to the housing of the instrument and thereby to the Earth. The other is suspended by a spring and forms the inertial member. Two basic designs are possible. In the moving-magnet type, the coil is fixed to the housing and the magnet is inertial. In the moving-coil type the roles are reversed (Fig. 3.22b). A coil of wire fixed to the inertial mass is suspended between the poles of a strong magnet, which in turn is fixed to the ground by the rigid housing. Any motion of the coil within the magnetic field induces a voltage in the coil proportional to the rate of change of magnetic flux. During a seismic arrival the vibration of the ground relative to the mass is converted to an electrical voltage by induction in the coil. The voltage is amplified and transmitted through an electrical circuit to the recorder.

#### 3.4.2.2 Horizontal-motion seismometer

The principle of the mechanical type of horizontal-motion seismometer is similar to that of the vertical-motion instrument. As before the inertial mass is mounted on a horizontal bar, but the fulcrum is now hinged almost vertically so that the mass is confined to swing sideways in a nearly horizontal plane (Fig. 3.23). The behavior of the system is similar to that of a gate when its hinges are slightly out of vertical alignment. If the hinge axis is tilted slightly forward, the stable position of the gate is where its center of mass is at its lowest point. In any displacement of the hinge, the restoring gravitational forces try to return it to



**Fig. 3.23** Schematic design of the pendulum type of horizontal-motion seismometer (after Strahler, 1963).

this stable position. Similarly, the horizontal-motion seismometer swings about its equilibrium position like a horizontal pendulum (in fact it is the housing of the instrument and thereby to the Earth. The other is suspended by a spring and forms the inertial member. Two basic designs are possible. In the moving-magnet type, the coil is fixed to the housing and the magnet is inertial. In the moving-coil type the roles are reversed (Fig. 3.22b). A coil of wire fixed to the inertial mass is suspended between the poles of a strong magnet, which in turn is fixed to the ground by the rigid housing. Any motion of the coil within the magnetic field induces a voltage in the coil proportional to the rate of change of magnetic flux. During a seismic arrival the vibration of the ground relative to the mass is converted to an electrical voltage by induction in the coil. The voltage is amplified and transmitted through an electrical circuit to the recorder.

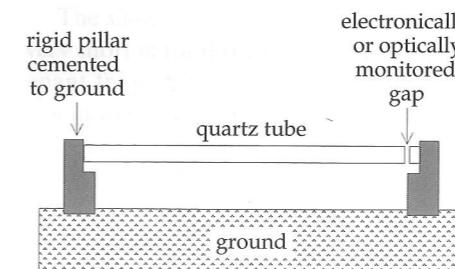
The design of an electromagnetic horizontal-motion seismometer is similar to that of the vertical-motion type, with the exception that the axis of the moving member (coil or magnet) is horizontal.

#### 3.4.2.3 Strain seismometer

The pendulum seismometers described above are inertial devices, which depend on the resistance of a loosely coupled mass to a change in its momentum. At about the same time that he developed the inertial seismograph, Milne also conducted experiments with a primitive strain seismograph that measured the change in distance between two posts during the passage of a seismic wave. The gain of early strain seismographs was low. However, in 1935 H. Benioff invented a sensitive strain seismograph from which modern versions are descended.

The principle of the instrument is shown in Fig. 3.24. It can record only horizontal displacements. Two collinear

## 3.4 THE SEISMOGRAPH



**Fig. 3.24** Schematic design of a strain seismometer (after Press and Siever, 1985).

horizontal rods made of fused quartz so as to be insensitive to temperature change are attached to posts about 20m apart, fixed to the ground; their near ends are separated by a small gap. The changes in separation of the two fixed posts result in changes in the gap width, which are detected with a capacitance or variable-reluctance transducer. In modern instruments the variation in gap width may be observed optically, using the interference between laser light-beams reflected from mirrors attached to the opposite sides of the gap. The strain instrument is capable of resolving strains of the order of  $10^{-8}$  to  $10^{-10}$ .

#### 3.4.3 The equation of the seismometer

Inertial seismometers for recording horizontal and vertical ground motion function on the pendulum principle. When the instrument frame is displaced from its equilibrium position relative to the inertial mass, a restoring force arises that is, to first order, proportional to the displacement. Let the vertical or horizontal displacement, dependent on the type of seismometer, be  $u$  and the restoring force  $-ku$ , and let the corresponding displacement of the ground be  $q$ . The total displacement of the inertial mass  $M$  is then  $(u+q)$ , and the equation of motion is

$$M \frac{d^2}{dt^2}(u+q) = -ku \quad (3.77)$$

We now divide throughout by  $M$ , write  $k/M = \omega_0^2$ , and after rearranging the equation, we get the familiar equation of forced simple harmonic motion

$$\frac{\partial^2 u}{\partial t^2} + \omega_0^2 u = -\frac{\partial^2 q}{\partial t^2} \quad (3.78)$$

In this equation  $\omega_0$  is the *natural frequency* (or *resonant frequency*) of the instrument. For a ground motion with this frequency, the seismometer would execute large uncontrolled vibrations and the seismic signal could not be recorded accurately. To get around this problem, the seismometer motion is damped by providing a velocity-dependent force that opposes the motion. A damping term enters into the equation of motion, which becomes

$$\frac{\partial^2 u}{\partial t^2} + 2\lambda\omega_0 \frac{\partial u}{\partial t} + \omega_0^2 u = -\frac{\partial^2 q}{\partial t^2} \quad (3.79)$$

The constant  $\lambda$  in this equation is called the *damping factor* of the instrument. It plays an important role in determining how the seismometer responds to a seismic wave.

A seismic signal is generally composed of numerous superposed harmonic vibrations with different frequencies. We can determine how a seismometer with natural frequency  $\omega_0$  responds to a seismic signal with any frequency  $\omega$  by solving Eq. (3.78) with  $q = A \cos \omega t$  (Box 3.2). Here,  $A$  is the magnified amplitude of the ground motion, equal to the true ground motion multiplied by a magnification factor that depends on the sensitivity of the instrument. Let the displacement  $u$  recorded by the seismometer be

$$u = U \cos(\omega t - \Delta) \quad (3.80)$$

where  $U$  is the amplitude of the recorded signal and  $\Delta$  is the phase difference between the record and the ground motion. As derived in Box 3.2, the phase lag  $\Delta$  is given by

$$\Delta = \tan^{-1}\left(\frac{2\lambda\omega\omega_0}{\omega_0^2 - \omega^2}\right) \quad (3.81)$$

The solution of the equation of motion (Eq. (3.79)) gives the displacement  $u$  on the seismic record as

$$u = \frac{A\omega^2}{[(\omega_0^2 - \omega^2)^2 + 4\lambda^2\omega^2\omega_0^2]^{1/2}} \cos(\omega t - \Delta) \quad (3.82)$$

#### 3.4.3.1 Effect of instrumental damping

The ground motion caused by a seismic wave contains a broad spectrum of frequencies. Equation (3.82) shows that the response of the seismometer to different signal frequencies is strongly dependent on the value of the damping factor  $\lambda$  (Fig. 3.25). A completely *undamped* seismometer has  $\lambda = 0$ , and for small values of  $\lambda$  the response of the seismometer is said to be *underdamped*. An undamped or greatly underdamped seismometer preferentially amplifies signals near the natural frequency, and therefore cannot make an accurate record of the ground motion; the undamped instrument will resonate at its natural frequency  $\omega_0$ . For all damping factors  $\lambda < 1/\sqrt{2}$  the instrument response function has a peak, indicating preferential amplification of a particular frequency.

The value  $\lambda = 1$  corresponds to *critical damping*, so-called because it delineates two different types of seismometer response in the absence of a forcing vibration. If  $\lambda < 1$ , the damped, free seismometer responds to a disturbance by swinging *periodically* with decreasing amplitude about its rest position. If  $\lambda \geq 1$ , the disturbed seismometer behaves *aperiodically*, moving smoothly back to its rest position. However, if the damping is too severe ( $\lambda \gg 1$ ), the instrument is *overdamped* and all frequencies in the ground motion are suppressed.

The optimum behavior of a seismometer requires that the instrument should respond to a wide range of frequencies in the ground motion, without preferential amplification or excessive suppression of frequencies.

**Box 3.2: The seismometer equation**

Equation (3.79) is the damped equation of motion for the signal  $u$  recorded by a seismometer with natural frequency  $\omega_0$  and damping factor  $\lambda$ , when the ground displacement is  $q$ . Let the ground displacement be  $q = A \cos \omega t$  and the recorded seismic signal be  $u = U \cos(\omega t - \Delta)$ . Substituting in Eq. (3.79) we get the following equation:

$$-\omega^2 U \cos(\omega t - \Delta) - 2\lambda \omega \omega_0 U \sin(\omega t - \Delta) + \omega_0^2 U \cos(\omega t - \Delta) = A \omega^2 \cos \omega t \quad (1)$$

i.e.,

$$U[(\omega_0^2 - \omega^2) \cos(\omega t - \Delta) - 2\lambda \omega \omega_0 \sin(\omega t - \Delta)] = A \omega^2 \cos \omega t \quad (2)$$

If we now write

$$(\omega_0^2 - \omega^2) = R \cos \varphi \quad 2\lambda \omega \omega_0 = R \sin \varphi \quad (3)$$

where

$$R = [(\omega_0^2 - \omega^2)^2 + 4\lambda^2 \omega^2 \omega_0^2]^{1/2} \quad \tan \varphi = \left( \frac{2\lambda \omega \omega_0}{\omega_0^2 - \omega^2} \right) \quad (4)$$

the equations reduce to

$$U[R \cos \varphi \cos(\omega t - \Delta) - R \sin \varphi \sin(\omega t - \Delta)] = A \omega^2 \cos \omega t \quad (5)$$

$$UR \cos(\omega t - \Delta + \varphi) = A \omega^2 \cos \omega t \quad (6)$$

Complex numbers (see Box 2.6) allow a simple solution to this equation. The function  $\cos \theta$  is the real part of the complex number (i.e.,  $\cos \theta = \operatorname{Re}\{e^{i\theta}\}$ ). Therefore,

$$UR \operatorname{Re}\{e^{i(\omega t - \Delta + \varphi)}\} = A \omega^2 \operatorname{Re}\{e^{i\omega t}\} \quad (7)$$

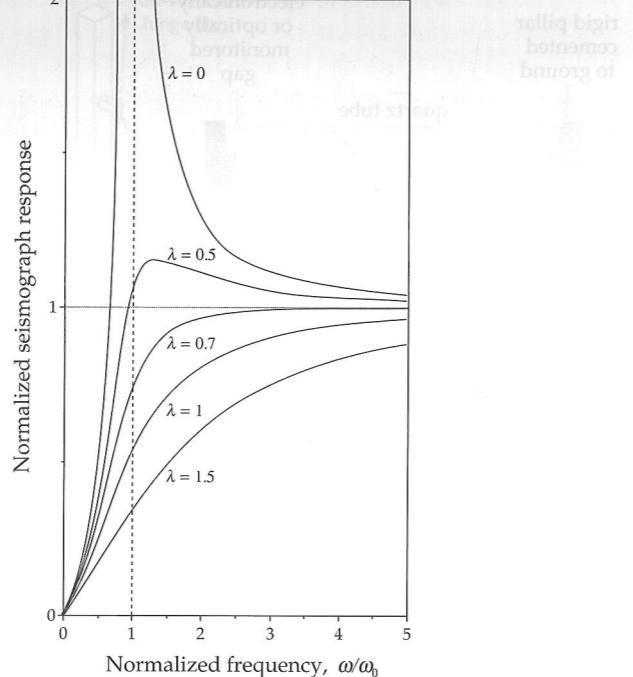
$$U = \frac{A \omega^2}{R} \operatorname{Re}\{e^{i\omega t} e^{-(i(\omega t - \Delta + \varphi))}\} = \frac{A \omega^2}{R} \operatorname{Re}\{e^{i(\Delta - \varphi)}\} = \frac{A \omega^2}{R} \cos(\Delta - \varphi) \quad (8)$$

The maximum amplitude of the record,  $U$ , is when  $\cos(\Delta - \varphi) = 1$ . The corresponding phase lag  $\Delta$  between the recorded signal and the ground motion is

$$\Delta = \tan^{-1} \left( \frac{2\lambda \omega \omega_0}{\omega_0^2 - \omega^2} \right) \quad (9)$$

and the equation for the amplitude of the seismic record is given by

$$u = A \cos \omega t = q \quad (3.84)$$



**Fig. 3.25** Effect of the damping factor  $\lambda$  on the response of a seismometer to different signal frequencies. Critical damping corresponds to  $\lambda = 1$ . Satisfactory operation corresponds to a damping factor between 0.7 and 1 (i.e., 70–100% of critical damping).

This requires that the damping factor should be close to the critical value. It is usually chosen to be in the range 70% to 100% of critical damping (i.e.,  $1/\sqrt{2} \leq \lambda < 1$ ). At critical damping the response of the seismometer to a periodic disturbing signal with frequency  $\omega$  is given by

$$u = \frac{A \omega^2}{(\omega_0^2 + \omega^2)} \cos(\omega t - \Delta) \quad (3.83)$$

### 3.4.3.2 Long-period and short-period seismometers

The natural period ( $2\pi/\omega_0$ ) of a seismometer is an important factor in determining what it actually records. Two examples of special interest correspond to instruments with very long and very short natural periods, respectively.

The *long-period seismometer* is an instrument in which the resonant frequency  $\omega_0$  is very low. For all but the lowest frequencies we can write  $\omega_0 \ll \omega$ . The phase lag  $\Delta$  between the seismometer and the ground motion becomes zero (see Box 3.2), and the amplitude of the seismometer displacement becomes equal to the amplified ground displacement  $q$ :

$$u = A \cos \omega t = q \quad (3.84)$$

The long-period seismometer is sometimes called a *displacement meter*. It is usually designed to record seismic signals with frequencies of 0.01–0.1 Hz (i.e., periods in the range 10–100 s).

## 3.4 THE SEISMOGRAPH

The *short-period seismometer* is constructed to have a very short natural period and a correspondingly high resonant frequency  $\omega_0$ , which is higher than most frequencies in the seismic wave. Under these conditions we have  $\omega_0 \gg \omega$ , the phase difference  $\Delta$  is again small and Eq. (3.83) becomes

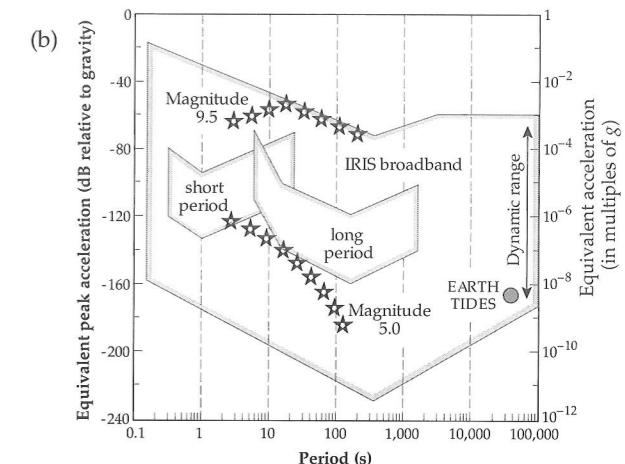
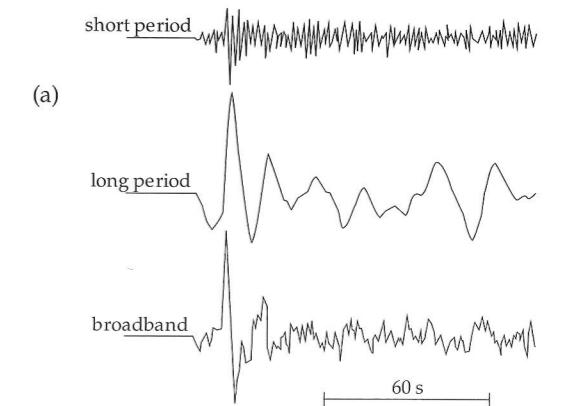
$$u = \frac{\omega^2}{\omega_0^2} A \cos \omega t = -\frac{1}{\omega_0^2} \ddot{q} \quad (3.85)$$

This equation shows that the displacement of the short-period seismometer is proportional to the acceleration of the ground, and the instrument is accordingly called an *accelerometer*. It is usually designed to respond to seismic frequencies of 1–10 Hz (periods in the range 0.1–1 s). An accelerometer is particularly suitable for recording strong motion earthquakes, when the amplitude of the ground motion would send a normal type of displacement seismometer off-scale.

### 3.4.3.3 Broadband seismometers

Short-period seismometers operate with periods of 0.1–1 s and long-period instruments are designed for periods greater than 10 s. The resolution of seismic signals with intermediate frequencies of 0.1–1 Hz (periods of 1–10 s) is hindered by the presence in this range of a natural form of seismic background noise. The noise derives from a nearly continuous succession of small ground movements that are referred to as *microseisms*. Some microseismic noise is of local origin, related to such effects as vehicular traffic, rainfall, wind action on trees, etc. However, an important source is the action of storm waves at sea, which is detectable on seismic records far inland. The drumming of rough surf on a shoreline and the interference of sea waves over deep water are thought to be the principal causes of microseismic noise. The microseismic noise has a low amplitude on a seismic record, but it may be as strong as a weak signal from a distant earthquake, which cannot be selectively amplified without also magnifying the noise. The problem is exacerbated by the limited dynamic range of short- or long-period seismometers. Short-period instruments yield records dominated by high frequencies while long-period devices smooth these out, giving a record with only a low-frequency content (Fig. 3.26a).

The range between the strongest and weakest signals that can be recorded without distortion by a given instrument is called its dynamic range. Dynamic range is measured by the power (or energy density) of a signal, and is expressed in units of decibels (dB). A decibel is defined as  $10 \log_{10}(\text{signal power})$ . Because power is proportional to the square of amplitude (Section 3.3.2.6), a decibel is equivalent to  $20 \log_{10}(\text{amplitude})$ . So, for example, a range of 20 dB in power corresponds to a factor 10 variation in acceleration, and a dynamic range of 100 dB corresponds to a  $10^5$  variation in amplitude. Short- and long-period seismometers have narrow dynamic ranges



**Fig. 3.26** (a) Comparison of short-period and long-period records of a teleseismic P-wave with a broadband seismometer recording of the same event, which contains more information than the other two records separately or combined. (b) Ranges of the ground acceleration (in dB) and periods of ground motion encompassed by the very broadband seismic system of the IRIS Global Seismic Network, compared with the responses of short-period and long-period seismometers and expected ground accelerations from magnitude 5.0 and 9.5 earthquakes (stars) and from bodily Earth-tides (redrawn from Lay and Wallace, 1995).

because they are designed to give optimum performance in limited frequency ranges, below or above the band of ground noise. This handicap was overcome by the design of broadband seismometers that have high sensitivity over a very wide dynamic range.

The broadband seismometer has basically an inertial pendulum-type design, with enhanced capability due to a force-feedback system. This works by applying a force proportional to the displacement of the inertial mass to prevent it from moving significantly. The amount of feedback force applied is determined using an electrical transducer to convert motion of the mass into an electrical signal. The force needed to hold the mass stationary corresponds to the ground acceleration. The signal is digitized with 16-to-24 bit resolution, synchronized with accurate time signals, and recorded on magnetic tape, hard disk, or solid state memory. The

feedback electronics are critical to the success of this instrument.

Broadband design results in a seismometer with great bandwidth and linear response. It is no longer necessary to avoid recording in the 1–10 s bandwidth of ground noise avoided by short-period and long-period seismometers. The recording of an earthquake by a broadband seismometer contains more useable information than can be obtained from the short-period or long-period recordings individually or in combination (Fig. 3.26a).

Broadband seismometers can be used to register a wide range of signals (Fig. 3.26b). The dynamic range extends from ground noise up to the strong acceleration that would result from an earthquake with magnitude 9.5, and the periods that can be recorded range from high-frequency body waves to the very long period oscillations of the ground associated with bodily Earth-tides (Section 2.3.3.5). The instrument is employed world-wide in modern standardized seismic networks, replacing short-period and long-period seismometers.

#### 3.4.4 The seismogram

A seismogram represents the conversion of the signal from a seismometer into a time record of a seismic event. The commonest method of obtaining a directly visible record, in use since the earliest days of modern seismology, uses a drum that rotates at a constant speed to provide the time axis of the record, as shown schematically in Fig. 3.22a and Fig. 3.23. In early instruments a mechanical linkage provided the coupling between sensor and record. The invention of the electromagnetic seismometer by Galitzin allowed transmission of the seismic signal to the recorder as an electrical signal. For many years, a galvanometer was used to convert the electrical signal back to a mechanical form for visual display.

In a galvanometer a small coil is suspended on a fine wire between the poles of a magnet. The current in the coil creates a magnetic field that interacts with the field of the permanent magnet and causes a deflection of the coil. The electrical circuitry of the galvanometer is designed with appropriate damping so that the deflection of the galvanometer is a faithful record of the seismic signal. The deflection can be transferred to a visible record in a variety of ways.

Mechanical and electromagnetic seismometers deliver continuous analog recordings of seismic events. These types of seismometer are now of mainly historic interest, having been largely replaced by broadband seismometers. Galvanometer-based analog recording has been superseded by digital recording.

##### 3.4.4.1 Analog recording

In an early method of recording, a smoked paper sheet was attached to the rotating drum. A fine stylus was

connected to the pendulum by a system of levers. The point of the stylus scratched a fine trace on the smoked paper. Later instruments employed a pen instead of the stylus and plain paper instead of the smoked paper. These methods make a “wiggly line” trace of the vibration.

In a further development of galvanometric recording a light-beam was reflected from a small mirror attached to the coil or its suspension to trace the record on photographic paper attached to the rotating drum. Photographic methods of recording are free of the slight friction of mechanical contact and allow inventive modifications of the form of the record. In the *variable density* method, the galvanometer current modulated the intensity of a light-bulb, so that the fluctuating signal showed on the photographic record as successive light and dark bands. A *variable area* trace was obtained by using the galvanometer current to vary the aperture of a narrow slit, through which a light-beam passed on to the photographic paper.

Every seismogram carries an accurate record of the elapsed time. This may be provided by a tuned electrical circuit whose frequency of oscillation is controlled by the natural frequency of vibration of a quartz crystal. At regular intervals the timing system delivers a short impulse to the instrument, causing a time signal to be imprinted on the seismogram. Modern usage is to employ a time signal transmitted by radio. The timing lines appear as regularly spaced blips on a wiggly line trace, or as bright lines on photographic records.

An important development after the early 1950s, especially in commercial seismology, was the replacement of photographic recording by *magnetic recording*. The electrical current from the seismometer was sent directly to the recording head of a tape recorder. The varying current in the recording head imprinted a corresponding magnetization on a magnetic tape. Magnetic recorders might have 24 to 50 parallel channels, each able to record from a different source.

Many years of development resulted in sophisticated methods of analog recording, but they have now been superseded universally by digital methods.

##### 3.4.4.2 Digital recording

In digital recording, the analog signal from a seismometer is passed through an electronic device called an analog-to-digital converter, which samples the continuous input signal at discrete, closely spaced time intervals and represents it as a sequence of *binary numbers*. Conventional representation on the familiar decimal scale expresses a number as a sum of powers of 10 multiplied by the digits 0–9. In contrast, a number is represented on the binary scale as a sum of powers of 2 multiplied by one of only two digits, 0 or 1. For example, in decimal notation the number 153 represents  $1 \times 10^2 + 5 \times 10^1 + 3 \times 10^0$ . In binary notation, the same number is represented as

#### 3.4 THE SEISMOGRAPH

$$\begin{aligned} 153 &= 128 + 16 + 8 + 1 \\ &= 1 \times 2^7 + 0 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 \\ &\quad + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 10011001. \end{aligned}$$

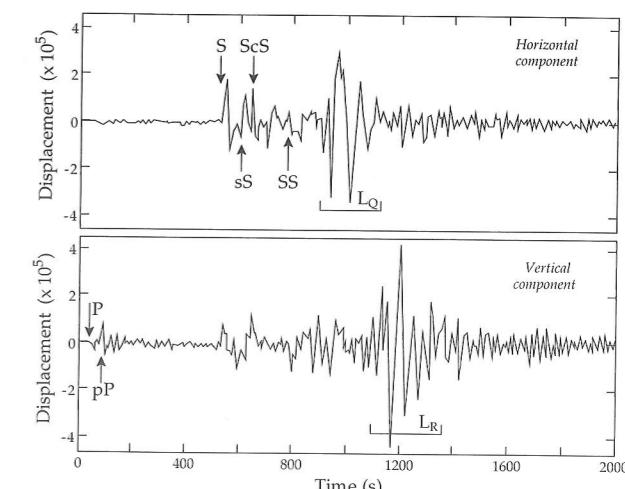
Each of the digits in a binary number is called a *bit* and the combination that expresses the digitized quantity is called a *word*; the binary number 10011001 is an *eight-bit word*. A binary number is much longer and thus more cumbersome for everyday use than the decimal form, but it is suitable for computer applications. Because it involves only two digits, a binary number can be represented by a simple condition, such as whether a switch is off or on, or the presence or absence of an electrical voltage or current.

Digital recording of seismic signals was developed in the 1960s and since the early 1970s it has virtually replaced analog recording. The analog method records the signal, usually employing the galvanometer principle, on photographic film or on magnetic tape as a continuous time-varying voltage whose amplitude is proportional to a characteristic of the ground disturbance (displacement, velocity or acceleration). The dynamic range of the analog method is limited and the system has to be adapted specifically to the characteristics of the signal to be recorded. For example, an analog device for recording strong signals lacks the sensitivity needed to record weak signals, whereas an analog recorder of weak signals will be overloaded by a strong signal. The digital recording technique samples the amplified output of the seismometer at time increments of a millisecond or so, and writes the digitized voltage directly to magnetic tape or to a computer hard-disk. This avoids possible distortion of the signal that can result from mechanical or optical recording. Digital recording has greater fidelity than the analog method, its dynamic range is wide, and the data are recorded in a form suitable for numerical processing by high-speed computers.

After processing, the digital record is usually converted back to an analog form for display and interpretation. The processed digital signal is passed through a digital-to-analog converter and displayed as a wiggle trace or variable density record. The familiar continuous paper record is still a common form of displaying earthquake records. However, instead of employing galvanometers to displace the pen in response to a signal, modern devices utilize a motor-driven pen; in this case the electrical signal from the earthquake record powers a small servo-motor which controls the pen displacement on the paper record.

##### 3.4.4.3 Phases on a seismogram

The seismogram of a distant earthquake contains the arrivals of numerous seismic waves that have travelled along different paths through the Earth from the source to the receiver. The appearance of the seismogram can therefore be very complicated, and its interpretation demands



**Fig. 3.27** Broadband seismograms of an earthquake in Peru recorded at Harvard, Massachusetts. Top: the SH body wave and Love ( $L_Q$ ) surface wave are prominent on the horizontal component record; bottom: the P and SV body waves and the Rayleigh ( $L_R$ ) surface waves are clear on the vertical component record. Both seismograms also show several other phases (redrawn from Lay and Wallace, 1995).

considerable experience. The analysis of seismic waves that have been multiply reflected and refracted in the Earth's interior will be treated in Section 3.7. Each event that is recorded in the seismogram is referred to as a *phase*.

As described in Section 3.3.2.1, the fastest seismic waves are the longitudinal waves. The first phase on the seismogram corresponds to the arrival of a longitudinal body wave, identified as the primary wave, or *P-wave* (Fig. 3.27). The next phase is the bodily shear wave, referred to as the secondary wave or *S-wave*, which usually has a larger amplitude than the P-wave. It is followed by the large-amplitude disturbance associated with surface waves, which are sometimes designated *L-waves* because of their much longer wavelengths. Dispersion (Section 3.3.3.3) causes the wavelengths at the head of the surface-wave train to be longer than those at the tail. Conventionally, Rayleigh waves are referred to as  $L_R$  waves, while Love waves are called  $L_Q$  waves.

The arrivals recorded on any seismogram depend on the type of sensor used. For example, a vertical-component seismometer responds to P-, SV- and Rayleigh waves but does not register SH- or Love waves; a horizontal-component seismometer can register P-, SH-, Rayleigh and Love waves. The amplitudes of the different phases on a seismogram are influenced by several factors: the orientation of the instrument axis to the wave path, the epicentral distance (see Section 3.5.2), the focal mechanism and the structure traversed by the waves.

Representative seismograms for a distant earthquake are shown in Fig. 3.27. They were recorded at Harvard, Massachusetts, for an earthquake that occurred deep beneath Peru on May 24, 1991. The seismic body waves from this earthquake have travelled through deep regions of the mantle and several seismic phases are recorded.

The upper seismogram is the trace of a horizontal-component seismometer oriented almost transverse to the seismic path, so the P-wave arrival is barely discernible. The first strong signal is the S body wave (in this case, an SH-wave), closely followed by several other phases (defined in Section 3.7.2) and the Love ( $L_Q$ ) surface-wave train. The lower seismogram, recorded by a vertical-component seismometer, shows the arrivals of P and SV body waves and the Rayleigh ( $L_R$ ) surface-wave train. Love waves travel along the surface at close to the near-surface S-wave velocity ( $V_{LQ} \approx \beta$ ); Rayleigh waves are slower, having  $V_{LR} \approx 0.92\beta$ , so they reach the recording station later than the Love waves.

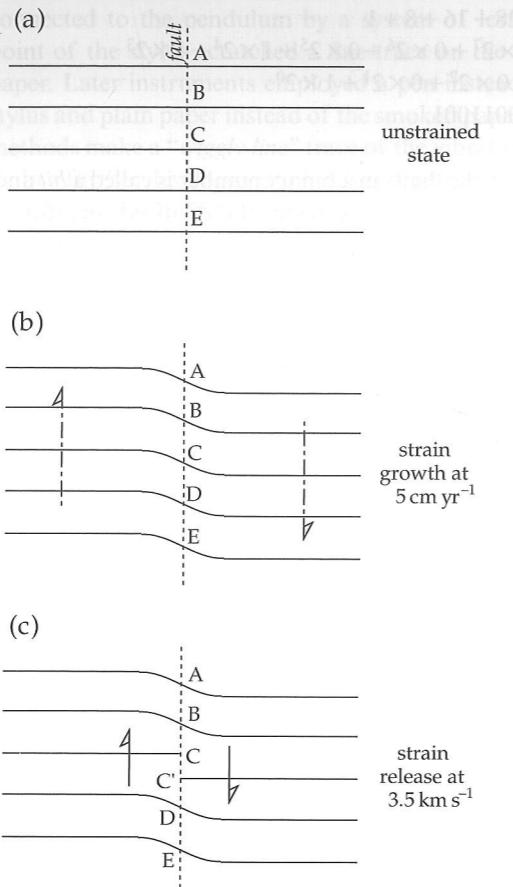
### 3.5 EARTHQUAKE SEISMOLOGY

#### 3.5.1 Introduction

Most of the earthquakes which shake the Earth each year are so weak that they are only registered by sensitive seismographs, but some are strong enough to have serious, even catastrophic, consequences for mankind and the environment. About 90% of all earthquakes result from tectonic events, primarily movements on faults. The remaining 10% are related to volcanism, collapse of subterranean cavities, or man-made effects.

Our understanding of the processes that lead to earthquakes derives to a large extent from observations of seismic events on the San Andreas fault in California. The average relative motion of the plates adjacent to the San Andreas fault is about  $5 \text{ cm yr}^{-1}$ , with the block to the west of the fault moving northward. On the fault-plane itself, this motion is not continuous but takes place spasmodically. According to modern plate tectonic theory this extensively studied fault system is a transform fault. This is a rather special type, so it cannot be assumed that the observations related to the San Andreas fault are applicable without reservation to all other faults. However, the *elastic rebound model*, proposed by H. F. Reid after the 1906 San Francisco quake, is a useful guide to how an earthquake may occur.

The model is illustrated in Fig. 3.28 by the changes to five parallel lines, drawn normal to the trace of the fault in the unstrained state and intersecting it at the points A–E. Strain due to relative motion of the blocks adjacent to the fault accumulates over several years. Far from the trace of the fault the five lines remain straight and parallel, but close to it they are bent. When the breaking point of the crustal rocks at C is exceeded, rupture occurs and there is a violent displacement on the fault-plane. The relative displacement that has been taking place progressively between the adjacent plates during years or decades is achieved on the fault-plane in a few seconds. The strained rocks adjacent to the fault “rebound” suddenly. The accumulated strain energy is released with the seismic speed of the ruptured rocks, which is several kilometers per second. The segments BC and C'D undergo compression, while



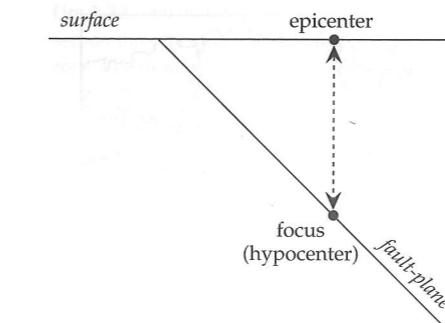
**Fig. 3.28** Elastic rebound model of the origin of earthquakes: (a) unstrained state of a fault segment, (b) accumulation of strain close to the fault due to relative motion of adjacent crustal blocks, and (c) “rebound” of strained segment as an earthquake with accompanying release of seismic energy.

CD and BC' experience dilatation. The points A and E do not move; the stored strain energy at these points is not released. The entire length of the fault-plane is not displaced, only the region in which the breaking point has been exceeded. The greater the length of the fault-plane that is activated, the larger is the ensuing earthquake.

The occurrence of a large earthquake is not necessarily as abrupt as described in the preceding paragraph, although it can be very sudden. In 1976 a major earthquake with magnitude 7.8 struck a heavily populated area of northern China near the city of Tangshan. Although there were known faults in the area, they had long been seismically inactive, and the large earthquake struck without warning. It completely devastated the industrial region and caused an estimated 243,000 fatalities. However, in many instances the accumulating strain is partially released locally as small earthquakes, or *foreshocks*. This is an indicator that strain energy is building up to the rupture level and is sometimes a premonition that a larger earthquake is imminent.

When an earthquake occurs, most of the stored energy is released in the main shock. However, for weeks or

### 3.5 EARTHQUAKE SEISMOLOGY



**Fig. 3.29** Vertical section perpendicular to the plane of a normal fault, defining the epicenter and hypocenter (focus) of an earthquake.

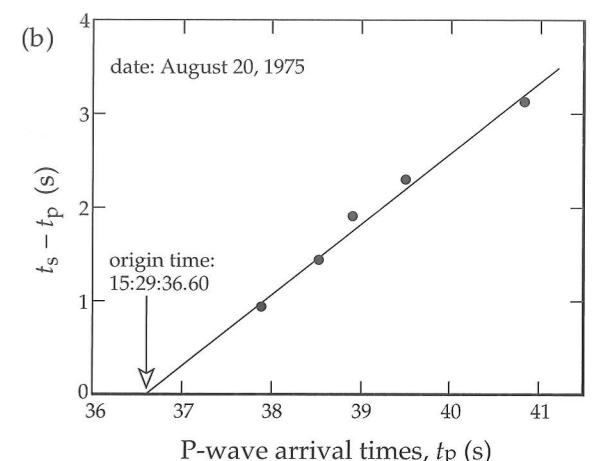
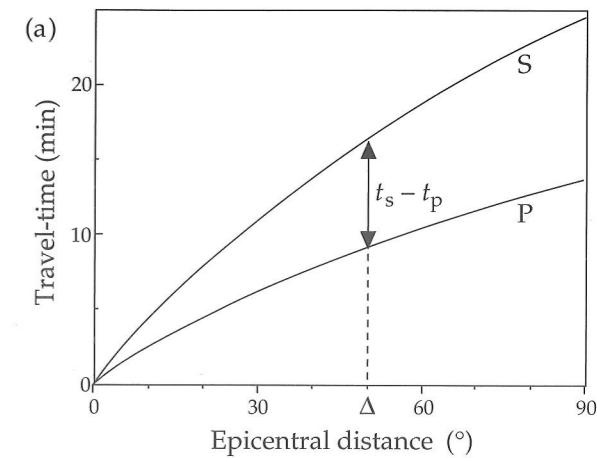
months after a large earthquake there may be numerous lesser shocks, known as *aftershocks*, some of which can be comparable in size to the main earthquake. Structures weakened by the main event often collapse in large aftershocks, which can cause physical damage as severe as the main shock. The death toll from aftershocks is likely to be less, because people have evacuated damaged structures.

Although in fact the earthquake involves a part of the fault-plane measuring many square kilometers in area, from the point of view of an observer at a distance of hundreds or even thousands of kilometers the earthquake appears to happen at a point. This point is called the *focus* or *hypocenter* of the earthquake (Fig. 3.29). It generally occurs at a *focal depth* many kilometers below the Earth's surface. The point on the Earth's surface vertically above the focus is called the *epicenter* of the earthquake.

#### 3.5.2 Location of the epicenter of an earthquake

The distance of a seismic station from the epicenter of an earthquake (the *epicentral distance*) may be expressed in kilometers  $\Delta_{\text{km}}$  along the surface, or by the angle  $\Delta^\circ = (180/\pi)(\Delta_{\text{km}}/R)$  subtended at the Earth's center. The travel-times of P- and S-waves from an earthquake through the body of the Earth to an observer are dependent on the epicentral distance (Fig. 3.30a). The travel-time versus distance plots are not linear, because the ray paths of waves travelling to distant seismographs are curved. However, the standard seismic velocity profile of the Earth's interior is well enough known that the travel-times for each kind of wave can be tabulated or graphed as a function of epicentral distance. In computing epicentral distance from earthquake data the total travel-time is not at first known, because an observer is rarely at the epicenter to record the exact time of occurrence  $t_0$  of the earthquake. However, the difference in travel-times for P- and S-waves ( $t_s - t_p$ ) can be obtained directly from the seismogram; it increases with increasing epicentral distance (Fig. 3.30a).

For local earthquakes we can assume that the seismic velocities  $\alpha$  and  $\beta$  are fairly constant in the near-surface layers. The time when the earthquake occurred,  $t_0$ , can then be obtained by plotting the differences ( $t_s - t_p$ )



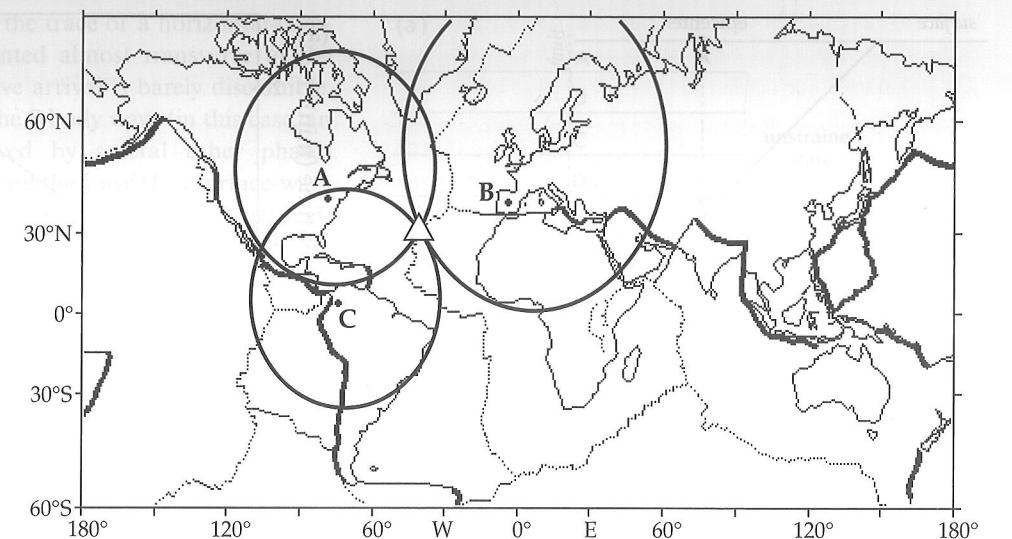
**Fig. 3.30** (a) Travel-times of P- and S-waves from an earthquake through the body of the Earth to an observer at epicentral distances up to  $90^\circ$ . The epicentral distance ( $\Delta$ ) of the earthquake is found from the difference in travel times ( $t_s - t_p$ ). (b) Wadati diagram for determining the time of occurrence of an earthquake.

against the arrival times  $t_p$  of the P-wave at different stations. The plot, called a *Wadati diagram*, is a straight line (Fig. 3.30b). If  $D$  is the distance travelled by the seismic wave, the travel-times of P- and S-waves are respectively  $t_p = D/\alpha$  and  $t_s = D/\beta$ , so

$$t_s - t_p = D\left(\frac{1}{\beta} - \frac{1}{\alpha}\right) = t_p\left(\frac{\alpha}{\beta} - 1\right) \quad (3.86)$$

The intercept  $t_0$  of the straight line with the arrival-time axis is the time of occurrence of the earthquake and the slope of the line is  $[(\alpha/\beta) - 1]$ . Knowing the P-wave velocity  $\alpha$ , the distance to the earthquake is obtained from  $D = \alpha(t_p - t_0)$ .

In order to determine the location of an earthquake, epicenter travel-times of P- and S-waves to at least three seismic stations are necessary (Fig. 3.31). The data from one station give only the distance of the epicenter from that station. It could lie anywhere on a circle centered at the station. The data from an additional station define a second circle which intersects the first circle at two points,



each of which could be the true epicenter. Data from a third station remove the ambiguity: the common point of intersection of the three circles is the epicenter.

Generally the circles do not intersect at a point, but form a small spherical triangle. The optimum location of the epicenter is at the center of the triangle. If data are available from more than three seismic stations, the epicentral location is improved; the triangle is replaced by a small polygon. This situation arises in part from observational errors, and because the theoretical travel-times are imperfectly known. The interior of the Earth is neither homogeneous nor isotropic, as must be assumed. The exact location of earthquake epicenters requires detailed knowledge of the seismic velocities along the entire path, but especially under the source area and the receiving station. The main reason for the intersection triangle or polygon is, however, that the seismic rays travel to the seismograph from the focus, and not from the epicenter. The focal depth of the earthquake,  $d$ , which may be up to several hundred kilometers, must be taken into account. It can be estimated from simple geometry. If  $\Delta_{\text{km}}$  is the epicentral distance and  $D$  the distance travelled by the wave, to a first approximation  $d = (D^2 - \Delta_{\text{km}}^2)^{1/2}$ . Combining several values of  $d$  from different recording stations gives a reasonable estimate of the focal depth.

### 3.5.3 Global seismicity

The epicenters of around 30,000 earthquakes are now reported annually by the International Seismological Center. The geographical distribution of world seismicity (see Fig. 1.10) dramatically illustrates the tectonically active regions of the Earth. The seismicity map is important evidence in support of plate tectonic theory, and delineates the presently active plate margins.

Earthquake epicenters are not uniformly distributed over the Earth's surface, but occur predominantly along rather narrow zones of *interplate* seismic activity. The

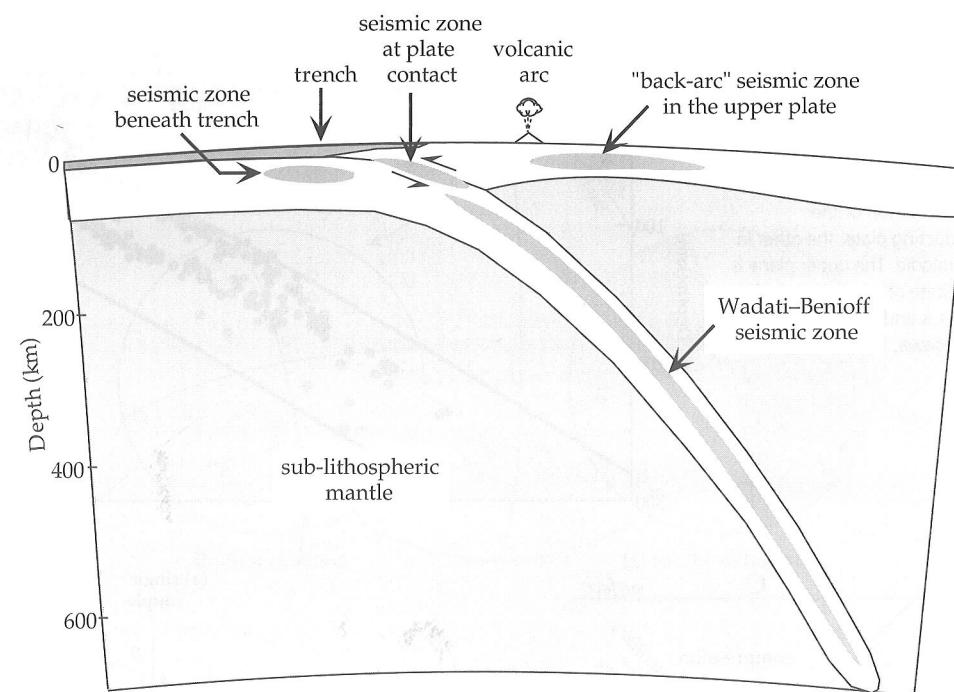
*circum-Pacific zone*, in which about 75–80% of the annual release of seismic energy takes place, forms a girdle that encompasses the mountain ranges on the west coast of the Americas and the island arcs along the east coast of Asia and Australasia. The *Mediterranean-transasiatic zone*, responsible for about 15–20% of the annual seismic energy release, begins at the Azores triple junction in the Atlantic Ocean and extends along the Azores–Gibraltar ridge; after passing through North Africa it makes a loop through the Italian peninsula, the Alps and the Dinarides; it then runs through Turkey, Iran, the Himalayan mountain chain and the island arcs of southeast Asia, where it terminates at the circum-Pacific zone. The system of *oceanic ridges and rises* forms the third most active zone of seismicity, with about 3–7% of the annually released seismic energy. In addition to their seismicity, each of these zones is also characterized by active volcanism.

The remainder of the Earth is considered to be *aseismic*. However, no region of the Earth can be regarded as completely earthquake-free. About 1% of the global seismicity is due to *intralate* earthquakes, which occur remote from the major seismic zones. These are not necessarily insignificant: some very large and damaging earthquakes (e.g. the New Madrid, Missouri, earthquakes of 1811 and 1812 in the Mississippi river valley) have been of the intralate variety.

Earthquakes can also be classified according to their focal depths. Earthquakes with *shallow* focal depths, less than 70 km, occur in all the seismically active zones; only shallow earthquakes occur on the oceanic ridge systems. The largest proportion (about 85%) of the annual release of seismic energy is liberated in shallow-focus earthquakes. The remainder is set free by earthquakes with *intermediate* focal depths of 70–300 km (about 12%) and by earthquakes with *deep* focal depths greater than 300 km (about 3%). These occur only in the circum-Pacific and Mediterranean-transasiatic seismic zones, and accompany the process of plate subduction.

### 3.5 EARTHQUAKE SEISMOLOGY

Fig. 3.32 Schematic cross-section through a subduction zone. The most active region is the zone of contact between the converging plates at depths of 10–60 km. There may be a "back-arc" seismic zone in the overriding plate. Below about 70 km depth a Wadati–Benioff seismic zone is described within the subducting plate (after Isacks, 1989).



The distributions of epicentral locations and focal depths of intermediate and deep earthquakes give important evidence for the processes at a subduction zone. When the earthquake foci along a subduction zone are projected onto a cross-section normal to the strike of the plate margin, they are seen to define a zone of seismicity about 30–40 km thick in the upper part of the 80–100 km thick subducting oceanic plate, which plunges at roughly 30–60° beneath the overriding plate (Fig. 3.32). For many years the inclined seismic zone was referred to in Western literature as a *Benioff zone* in recognition of the Californian scientist, Hugo Benioff. In the years following World War II Benioff carried out important pioneering studies that described the distribution of deep earthquakes on steeply dipping surfaces of seismicity. Many characteristics of the occurrence of deep earthquakes had been described in the late 1920s by a Japanese seismologist, Kiyoo Wadati. He discovered that the closer the epicenters of earthquakes lay to the Asian continent, the greater were their focal depths; the deep seismicity appeared to lie on an inclined plane. It was Benioff, however, who in 1954 proposed as an explanation of the phenomenon that the ocean floor was being "subducted" underneath the adjacent land. This was a bold proposal well in advance of the advent of plate tectonic theory. Today the zone of active seismicity is called a *Wadati–Benioff zone* in recognition of both discoverers.

In three dimensions the Wadati–Benioff zone constitutes an inclined slab dipping underneath the overriding plate. It marks the location and orientation of the upper surface of the subducting plate. The dip-angle of the zone varies between about 30° and 60°, becoming steeper with increasing depth, and it can extend to depths of several hundred kilometers into the Earth. The deepest reliably

located focal depths extend down to about 680 km. Important changes in the crystalline structure of mantle minerals occur below this depth.

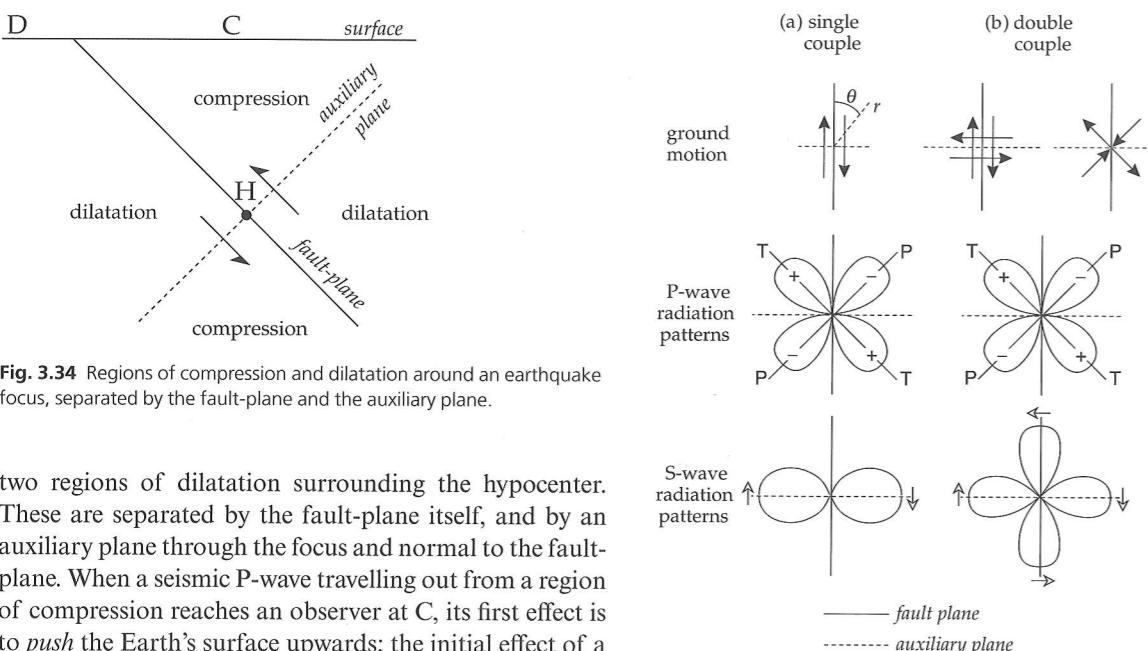
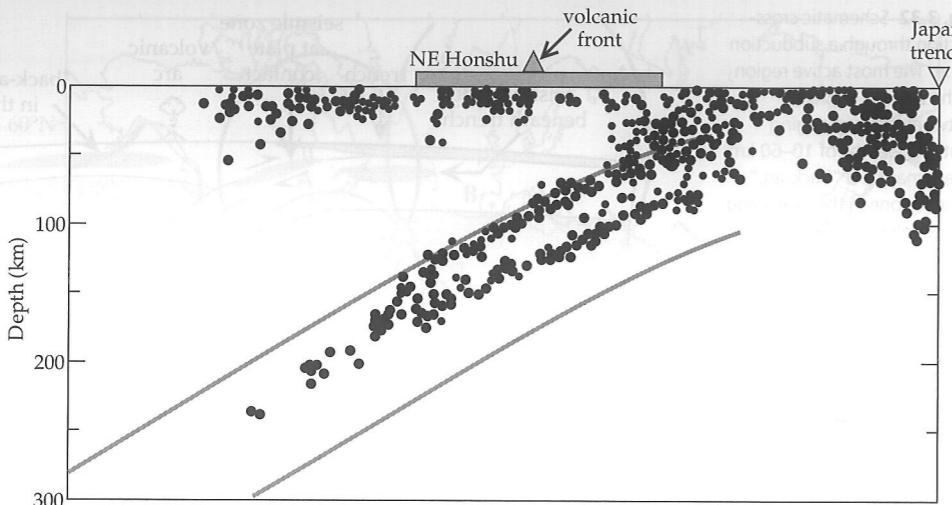
The structure of a subducting plate is not always as simple as described. A detailed study of the subducting Pacific plate revealed a double Wadati–Benioff zone under northeast Honshu, Japan (Fig. 3.33). The seismicity at depths below 100 km defines two parallel planes about 30–40 km apart. The upper plane, identified with the top of the subducting plate, is in a state of compression; the lower plane, in the middle of the slab, is in a state of extension. These stress states are the result of unbending of the subducting plate, which had previously undergone sharp bending at shallow depth below the trench axis. This information is inferred from analysis of the mechanisms by which the earthquakes occur.

### 3.5.4 Analysis of earthquake focal mechanisms

During an earthquake the accumulated elastic energy is released suddenly by physical displacement of the ground, as heat and as seismic waves that travel outwards from the focus. By studying the first motions recorded by seismographs at distant seismic stations, the focal mechanism of the earthquake can be inferred and the motion on the fault-plane interpreted.

Consider a vertical section perpendicular to the plane of a normal fault on which the hypocenter of an earthquake is located at the point H (Fig. 3.34). When the region above the fault moves up-slope, it produces a region of compression ahead of it and a region of dilatation (or expansion) behind it. In conjunction with the compensatory down-slope motion of the lower block, the earthquake produces two regions of compression and

**Fig. 3.33** Distribution of microearthquakes in a vertical section across a double subduction zone under the island of Honshu, Japan. Below 100 km the seismicity defines two parallel planes, one at the top of the subducting plate, the other in the middle. The upper plane is in a state of compression, the lower is under extension (after Hasegawa, 1989).



**Fig. 3.34** Regions of compression and dilatation around an earthquake focus, separated by the fault-plane and the auxiliary plane.

two regions of dilatation surrounding the hypocenter. These are separated by the fault-plane itself, and by an auxiliary plane through the focus and normal to the fault-plane. When a seismic P-wave travelling out from a region of compression reaches an observer at C, its first effect is to *push* the Earth's surface upwards; the initial effect of a P-wave that travels out from a region of dilatation to an observer at D is to *tug* the surface downwards. The P-wave is the earliest seismic wave to reach a seismograph at C or D and therefore the initial motion recorded by the instrument allows us to distinguish whether the first arrival was compressional or dilatational.

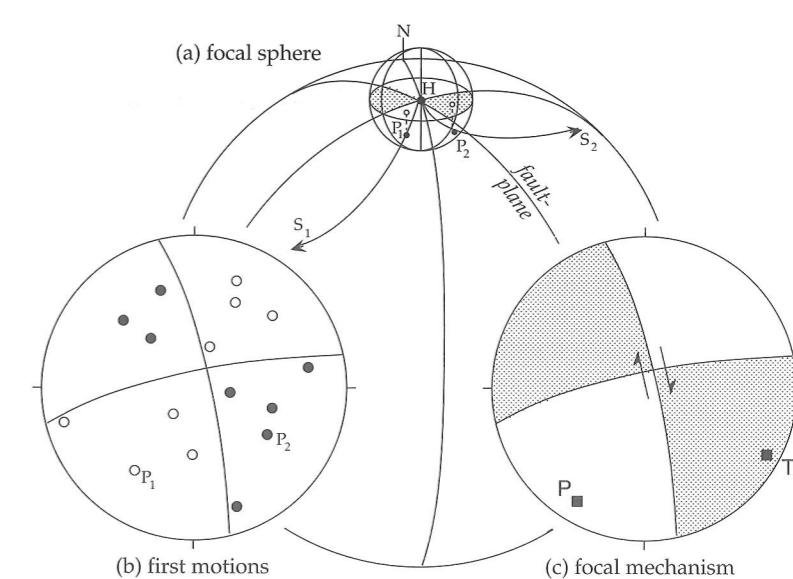
#### 3.5.4.1 Single-couple and double-couple radiation patterns

The amplitudes of P-waves and S-waves vary with distance from their source because of the effects of physical damping and geometric dispersion. The amplitudes also depend geometrically on the angle at which the seismic ray leaves the source. This geometric factor can be calculated mathematically, assuming a model for the source mechanism. The simplest is to represent the source by a single pair of antiparallel motions. Analysis of the amplitude of the P-wave as a function of the angle  $\theta$  between a ray and the plane of the fault (Fig. 3.35a) gives an equation of the form

$$A(r,t,\alpha,\theta) = A_0(r,t,\alpha)\sin^2 2\theta \quad (3.87)$$

in which  $A_0(r, t, \alpha)$  describes the decrease in amplitude with distance  $r$ , time  $t$ , and seismic P-wave velocity  $\alpha$ . A plot of the amplitude variation with  $\theta$  is called the radiation pattern of the P-wave amplitude, which for the single-couple model has a quadrupolar character (Fig. 3.35a). It consists of four lobes, two corresponding to the angular variation of amplitude where the first motion is compressional, and two where the first motion is

**Fig. 3.36** Method of determining the focal mechanism of an earthquake. (a) The focal sphere surrounding the earthquake focus, with two rays  $S_1$  and  $S_2$  that cut the sphere at  $P_1$  and  $P_2$ , respectively. (b) The points  $P_1$  and  $P_2$  are plotted on a lower-hemisphere stereogram as first-motion pushes (solid points) or tugs (open points). (c) The best-fitting great circles define regions of compression (shaded) and tension (unshaded). The P- and T-axes are located on the bisectors of the angles between the fault-plane and auxiliary plane.



dilatational. The lobes are separated by the fault-plane and the auxiliary plane.

The radiation pattern for S-waves from a single-couple source is described by an equation of the form

$$B(r,t,\beta,\theta) = B_0(r,t,\beta)\sin^2\theta \quad (3.88)$$

where the amplitude  $B$  is now dependent on the S-wave velocity  $\beta$ . The radiation pattern has a dipolar character consisting of two lobes in which the first motions are of opposite sense.

An alternative model of the earthquake source is to represent it by a pair of orthogonal couples (Fig. 3.35b). The double-couple source gives the same form of radiation pattern for P-waves as the single-couple source, but the radiation pattern for S-waves is quadrupolar instead of dipolar. This difference in the S-wave characteristics enables the seismologist to determine which of the two earthquake source models is applicable. S-waves arrive later than P-waves, so their first motions must be resolved from the background noise of earlier arrivals. They can be observed and are consistent with the double-couple model.

Note that the maximum P-wave amplitudes occur at  $45^\circ$  to the fault plane. The directions of maximum amplitude in the compressional and dilatational fields define the T-axis and P-axis, respectively. Here T and P imply "tension" and "compression," respectively, the stress conditions before faulting. Geometrically the P- and T-axes are the bisectors of the angles between the fault-plane and auxiliary plane. The orientations of these axes and of the fault-plane and auxiliary plane can be obtained even for distant earthquakes by analyzing the directions of first motions recorded in seismograms of the events. The analysis is called a fault-plane solution, or focal mechanism solution.

#### 3.5.4.2 Fault-plane solutions

The ray path along which a P-wave travels from an earthquake to the seismogram is curved because of the variation of seismic velocity with depth. The first step in the fault-plane solution is to trace the ray back to its source. A fictitious small sphere is imagined to surround the focus (Fig. 3.36a) and the point at which the ray intersects its surface is computed with the aid of standardized tables of seismic P-wave velocity within the Earth. The azimuth and dip of the angle of departure of the ray from the earthquake focus are calculated and plotted as a point on the lower hemisphere of the small sphere. This direction is then projected onto the horizontal plane through the epicenter. The projection of the entire lower hemisphere is called a stereogram. The direction of the ray is marked with a solid point if the first motion was a push away from the focus (i.e., the station lies in the field of compression). An open point indicates that the first motion was a tug towards the focus (i.e., the station lies in the field of dilatation).

First-motion data of any event are usually available from several seismic stations that lie in different directions from the focus. The solid and open points on the stereogram usually fall in distinct fields of compression and dilatation (Fig. 3.36b). Two mutually orthogonal planes are now drawn so as to delineate these fields as well as possible. The fit is best made mathematically by a least-squares technique, but often a visual fit is obvious and sufficient. The two mutually orthogonal planes correspond to the fault-plane and the auxiliary plane, although it is not possible to decide which is the active fault-plane from the seismic data alone. The regions of the stereogram corresponding to compressional first motions are usually shaded to distinguish them from the regions of dilatational first motions (Fig. 3.36c). The P- and T-axes are the lines that bisect the angles between the fault-plane

and auxiliary plane in the fields of dilatation and compression, respectively. To attach a physical meaning to the P- and T-axes we will have to take a closer look at the mechanics of faulting.

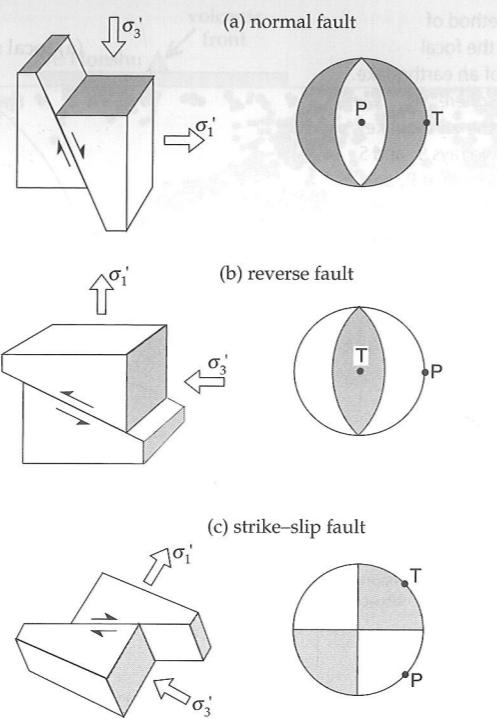
### 3.5.4.3 Mechanics of faulting

As discussed in elasticity theory, the state of stress can be represented by the magnitudes and directions of the three principal stresses  $\sigma_1 > \sigma_2 > \sigma_3$ . The directions of these principal stresses are by definition parallel to the coordinate axes and are therefore positive for *tensional stress*. The theory of faulting of homogeneous materials has been developed by studying the failure of materials under *compressional stress*, which is directed inwards toward the origin of the coordinate axes. The minimum tensional stress corresponds to the maximum compressional stress, and vice versa. The reason for taking this view is that geologists are interested in the behavior of materials within the Earth, where pressure builds up with increasing depth and faulting occurs under high confining pressures.

We can combine both points of view if we consider stress to consist of a part that causes change of volume and a part that causes distortion. The first of these is called the *hydrostatic stress*, and is defined as the mean ( $\sigma_m$ ) of the three principal stresses:  $\sigma_m = (\sigma_1 + \sigma_2 + \sigma_3)/3$ . If we now subtract this value from each of the principal stresses we get the *deviatoric stresses*:  $\sigma'_1 > \sigma'_2 > \sigma'_3$ . Note that  $\sigma'_1$  is positive; it is a tensional stress, directed outward. However,  $\sigma'_3$  is negative; it is a compressional stress, directed inward.

Failure of a material occurs on the plane of maximum shear stress. For a perfectly homogeneous material this is the plane that contains the intermediate principal stress axis  $\sigma_2$  (or  $\sigma'_2$ ). The fault-plane is oriented at 45° to the axes of maximum and minimum compressional stress, i.e., it bisects the angle between the  $\sigma'_1$  and  $\sigma'_3$  axes. In real materials inhomogeneity and the effects of internal friction result in failure on a fault-plane inclined at 20°–30° to the axis of maximum compression. Seismologists often ignore this complication when they make fault-plane analyses. The axis of maximum tensional stress  $\sigma'_1$  is often equated with the T-axis, in the field of compressional first motions on the bisector of the angle between the two principal planes. The axis of maximum compressional stress  $\sigma'_3$  is equated with the P-axis, in the field of dilatational first motions. In reality the direction of  $\sigma'_3$  will lie between the P-axis and the fault plane.

The locations of P and T may at first seem strange, for the axes appear to lie in the wrong quadrants. However, one must keep in mind that the orientations of the principal stress axes correspond to the stress pattern *before* the earthquake, while the fault-plane solution shows the ground motions that occurred *after* the earthquake. The focal mechanism analysis makes it possible to interpret the directions of the principal axes of stress in the Earth's crust that led to the earthquake.



**Fig. 3.37** The three main types of fault and their focal mechanisms. Left: the orientations of each fault-plane and the principal deviatoric stresses,  $\sigma'_1$  and  $\sigma'_3$ . Right: focal mechanisms and orientations of P- and T-axes.

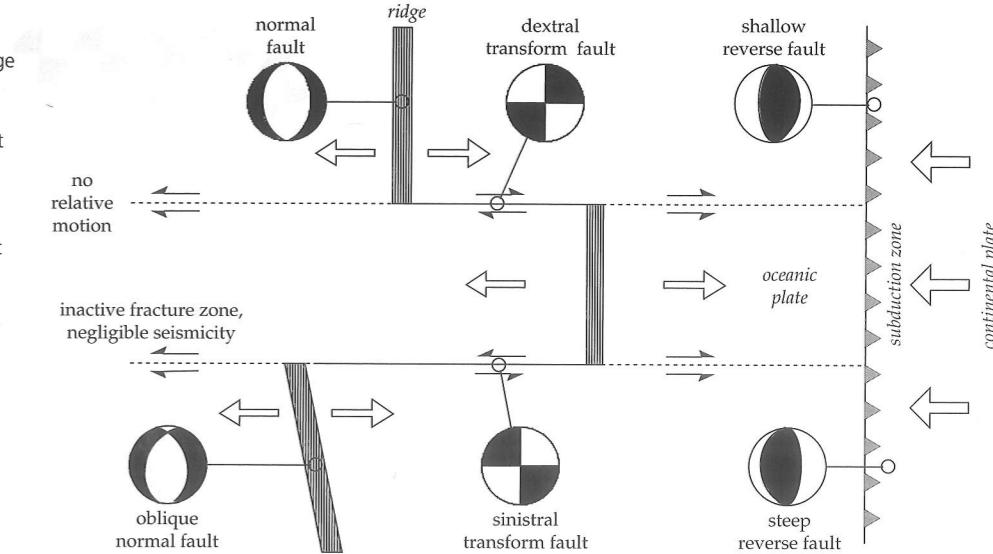
There are only three types of tectonic fault. These can be distinguished by the orientations of the principal axes of stress to the horizontal plane (Fig. 3.37). The focal solutions of earthquakes associated with each type of fault have characteristic geometries. When motion on the fault occurs up or down the fault plane it is called a *dip-slip fault*, and when the motion is horizontal, parallel to the strike of the fault, it is called a *strike-slip fault*.

Two classes of dip-slip fault are distinguished depending on the sense of the vertical component of motion. In a *normal fault*, the block on the upper side of the fault drops down an inclined plane of constant steepness relative to the underlying block (Fig. 3.37a). The corresponding fault-plane solution has regions of compression at the margins of the stereogram. The T-axis is horizontal and the P-axis is vertical. A special case of this type is the *listric fault*, in which the steepness of the fault surface is not constant but decreases with increasing depth.

In the second type of dip-slip fault, known as a *reverse fault* or *thrust fault*, the block on the upper side of the fault moves up the fault-plane, overriding the underlying block (Fig. 3.37b). The fault-plane solution is typified by a central compressional sector. The orientations of the axes of maximum tension and compression are the inverse of the case for a normal fault. The T-axis is now vertical and the P-axis is horizontal. When the fault-plane is inclined at a very flat angle, the upper block can be transported over large horizontal distances. This special type of *overthrust fault* is common in regions of continental collision, as for example in the Alpine–Himalayan mountain belts.

### 3.5 EARTHQUAKE SEISMOLOGY

**Fig. 3.38** Fault-plane solutions for hypothetical earthquakes at an ocean ridge and transform fault system. Note that the sense of movement on the fault is not given by the apparent offset of the ridge. The focal mechanisms of earthquakes on the transform fault reflect the relative motion between the plates. Note that in this and similar figures the sector with compressional first motions is shaded.



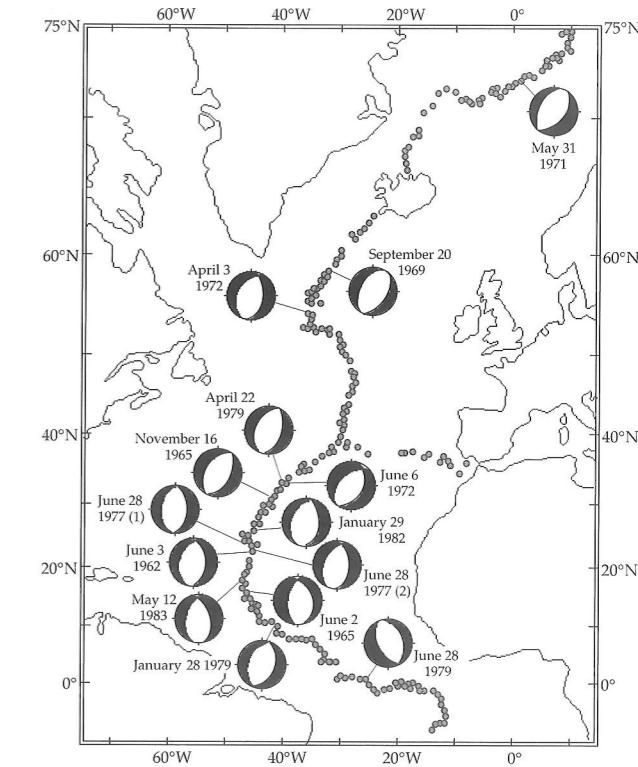
The simplest type of strike-slip fault is the *transcurrent fault*, in which the fault-plane is steep or vertical (Fig. 3.37c). To cause this type of fault the T- and P-axes must both lie in the horizontal plane. The fault-plane solution shows two compressional and two dilatational quadrants. Each side of the fault moves in opposite horizontal directions. If the opposite side of the fault to an observer is perceived to move to the left, the fault is said to be *sinistral*, or left-handed; if the opposite side moves to the right, the fault is *dextral*, or right-handed.

A variant of the strike-slip fault plays a very important role in the special context of plate tectonics. A *transform fault* allows horizontal motion of one plate relative to its neighbor. It joins segments of a spreading ocean ridge, or segments of a subducting plate margin. It is important in plate tectonics because it constitutes a conservative plate margin at which the tectonic plates are neither formed nor destroyed. The relative motion on the transform fault therefore reveals the direction of motion on adjacent segments of an active plate margin. The sense of motion is revealed by the pattern of compressional and dilatational sectors on the fault-plane solution.

### 3.5.4.4 Focal mechanisms at active plate margins

Some of the most impressive examples of focal mechanism solutions have been obtained from active plate margins. The results fully confirm the expectations of plate tectonic theory and give important evidence for the directions of plate motions. We can first ask what types of focal mechanism should be observed at each of the three types of active plate margin. In the theory of plate tectonics these are the *constructive* (or “accreting”), *conservative* (or “transform”) and *destructive* (or “consuming”) margins.

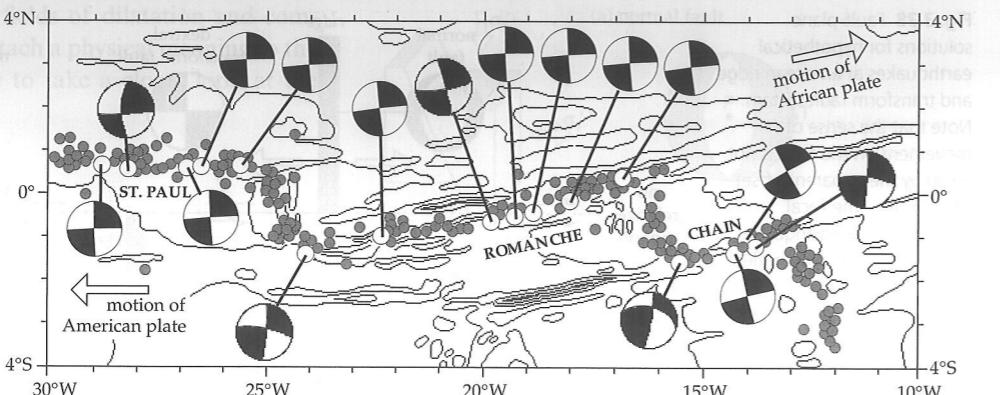
Oceanic spreading systems consist of both constructive and conservative margins. The seismicity at these plate margins forms narrow belts on the surface of the globe. The focal depths are predominantly shallow, generally less



**Fig. 3.39** Fault-plane solutions for earthquakes along the Mid-Atlantic Ridge, showing the prevalence of extensional tectonics with normal faulting in the axial zone of the spreading center (based on data from Huang et al., 1986).

than 10 km below the ocean bottom. Active ridge segments are separated by transform faults (Fig. 3.38). New oceanic lithosphere is generated at the spreading oceanic ridges; the separation of the plates at the spreading center is accompanied by extension. The plates appear to be pulled apart by the plate tectonic forces. The extensional nature of the ridge tectonics is documented by fault-plane solutions indicative of normal faulting, as is seen for some selected earthquakes along the Mid-Atlantic Ridge (Fig. 3.39). In

**Fig. 3.40** Fault-plane solutions for earthquakes on the St. Paul, Romanche and Chain transform faults in the Central Atlantic ocean (after Engeln *et al.*, 1986). Most focal mechanisms show right lateral (dextral) motions on these faults, corresponding to the relative motion between the African and American plates.



each case the fault-plane is oriented parallel to the strike of the ridge. On a ridge segment that is nearly normal to the nearest transform faults the focal mechanism solution is symmetric, with shaded compressional quadrants at the margins of the stereogram. Note that where the ridge is inclined to the strike of the transform fault the focal mechanism solution is not symmetric. This means that the plates are not being pulled apart perpendicular to the ridge. The fault-plane is still parallel to the strike of the ridge, but the slip-vector is oblique; the plate motion has a component perpendicular to the ridge and a component parallel to the ridge. We can understand why the direction of plate motion is not determined by the strike of the ridge axis by examining the motion on the adjacent transform faults.

A destructive (or consuming) plate margin is marked by a subduction zone, where a plate of oceanic lithosphere is destroyed by plunging under another plate of oceanic or continental lithosphere. Because this is a margin of convergence of the adjacent plates, the earthquake fault-plane solutions are typical of a compressional regime (Fig. 3.38). The regions of compressional first motion are in the center of the stereogram, indicating reverse faulting; the P-axes of maximum compressive stress are perpendicular to the strike of the subduction zone.

The type of focal mechanism observed at a subduction zone is dependent on the focal depth of the earthquake. This is because the state of stress varies within the subducting plate. At first the overriding plate is thrust at a shallow angle over the subducting plate. In the seismic zone at the contact between the two plates earthquake focal mechanisms are typical of low-angle reverse faulting. The focal mechanisms of earthquakes along the west coast of Mexico illustrate the first of these characteristics (Fig. 3.41). The selected earthquakes have large magnitudes ( $6.9 \leq M_s \leq 7.8$ ) and shallow focal depths. The strikes of the fault-planes follow the trend of the oceanic trench along the Mexican coastline. The focal mechanisms have a central sector of compressional first motions and the fault-plane is at a low angle to the northeast, typical of overthrusting tectonics. The seismicity pattern documents the subduction of the Cocos plate under the North American plate.

As a result of the interplate collision the subducting oceanic plate is bent downwards and its state of stress changes. Deeper than 60–70 km the seismicity does not arise from the contact between the converging plates. It is

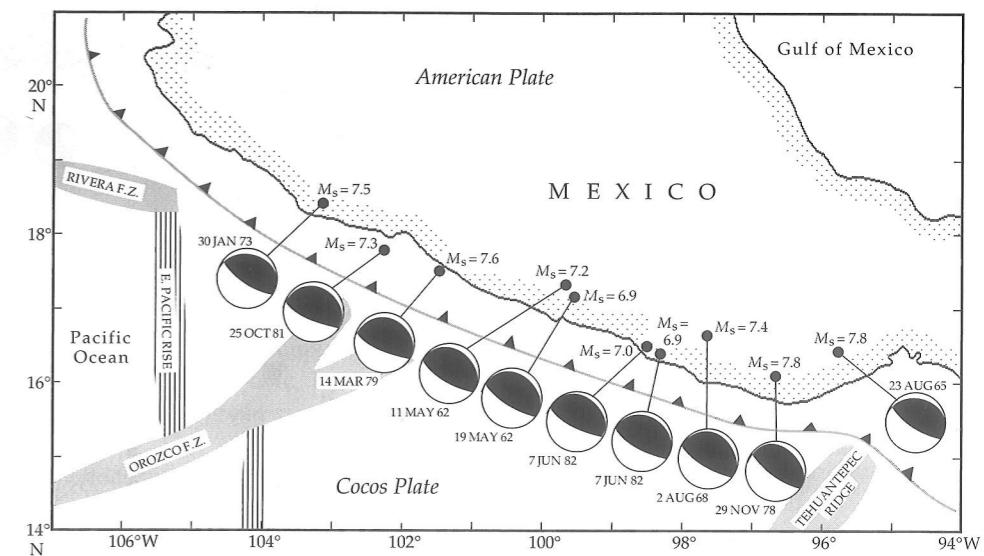
If we ignore changes in plate motion (which can, however, occur sporadically on some ridge systems), the offset of the neighboring ridge segments is a permanent feature that reflects how the plates first split apart. The orientations of the transform faults are very important, because the plates must move parallel to these faults. Thus the transform faults provide the key to determining the directions of plate motion. Where a ridge axis is not perpendicular to the transform fault the plate motion will have a component parallel to the ridge segment, which gives an oblique focal mechanism on the ridge.

The special class of strike-slip fault that joins active segments of a ridge or subduction zone is called a transform fault because it marks a conservative plate margin, where the lithospheric plates are being neither newly generated nor destroyed. The adjacent plates move past each other on the active fault. Relative plate motion is present only between the ridge segments. Almost the entire seismicity on the transform fault is concentrated in this region. On the parts of the fracture zone outside the segment of active faulting the plates move parallel to each other and there is little or no seismicity.

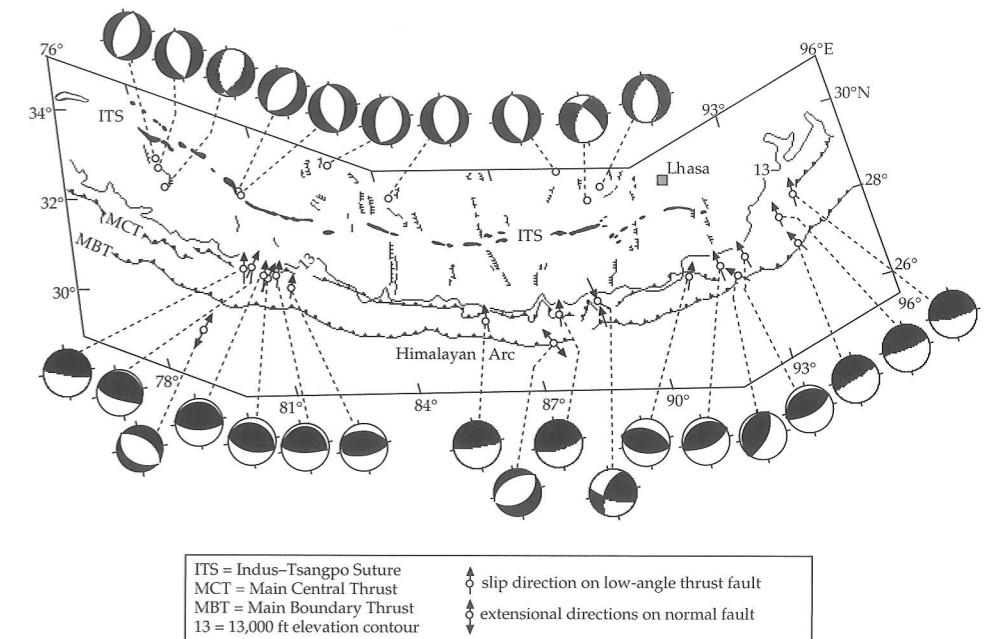
Because the relative motion is horizontal, the fault-plane solution is typical of a strike-slip fault. However, if the visible offset of the ridge segments were used to interpret the sense of motion on this fault, the wrong conclusion would be drawn. The conventional interpretation of this class of faults as a transcurrent fault was an early stumbling block to the development of plate tectonic theory. As indicated by arrows on the focal mechanism diagrams (Fig. 3.38), the relative motion on a transform fault is opposite to what one would expect for a transcurrent fault. It is determined by the opposite motions of the adjacent plates and not by the offset of the ridge segments. Hence, the focal mechanisms for a number of earthquakes on transform faults at the Mid-Atlantic Ridge in the central Atlantic reflect the eastwards motion of the African plate and westwards motion of the American plate (Fig. 3.40).

### 3.5 EARTHQUAKE SEISMOLOGY

**Fig. 3.41** Fault-plane solutions for selected large shallow earthquakes in the subduction zone along the west coast of Mexico (after Singh *et al.*, 1984). The focal mechanisms indicate low-angle overthrusting, as the Cocos plate is subducted to the northeast under Mexico.



**Fig. 3.42** Fault-plane solutions of earthquakes along the arcuate Himalayan mountain belt show normal faulting on north-south oriented fault planes in southern Tibet, and mainly low-angle thrust faults along the Lesser Himalaya mountain chain (after Ni and Barazangi, 1984).



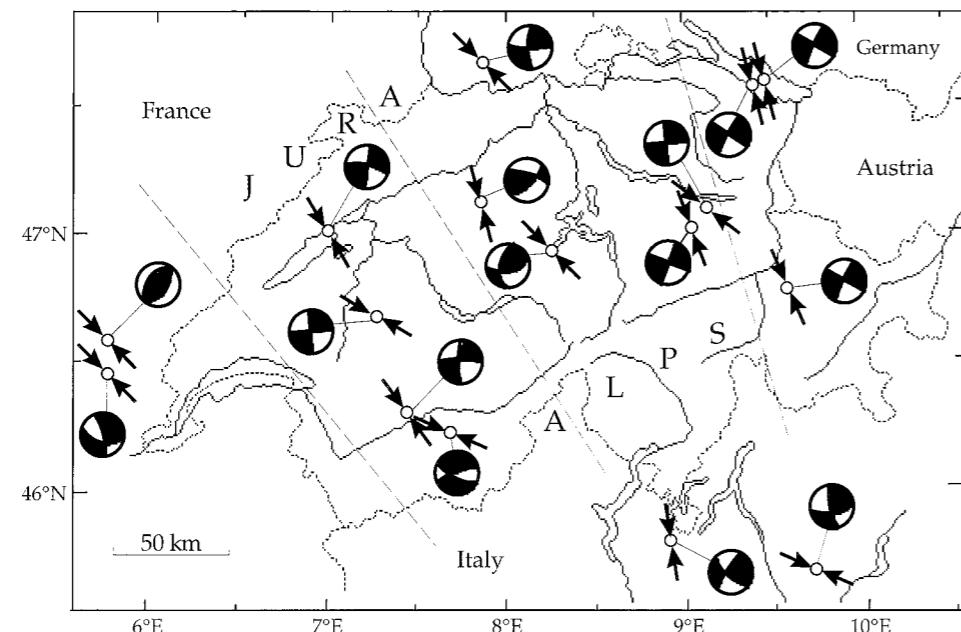
caused by the stress pattern within the subducted plate itself. The focal mechanisms of some intermediate-depth earthquakes (70–300 km) show down-dip extension (i.e., the T-axes are parallel to the surface of the dipping slab) but some show down-dip compression (i.e., the P-axes are parallel to the dip of the slab). At great depths in most Wadati–Benioff zones the focal mechanisms indicate down-dip compression.

#### 3.5.4.5 Focal mechanisms in continental collisional zones

When the continental portions of converging plates collide, they resist subduction into the denser mantle. The deformational forces are predominantly horizontal and lead to the formation of folded mountain ranges. The associated seismicity tends to be diffuse, spread out over a large geographic area. The focal mechanisms of earth-

quakes in the folded mountain belt reflect the ongoing collision. The collision of the northward-moving Indian plate with the Eurasian plate in the Late Tertiary led to the formation of the Himalayas. Focal mechanisms of earthquakes along the arcuate mountain belt show that the present style of deformation consists of two types (Fig. 3.42). Two fault-plane solutions south of the main mountain belt correspond to an extensional regime with normal faulting. To the north, in southern Tibet, the fault-plane solutions also show normal faulting on north-south oriented fault planes. Beneath the Lesser Himalaya seismic events distributed along the entire 1800 km length of the mountain chain indicate a deformational regime with some strike-slip faulting but predominantly low-angle thrust faulting. The Indian continental crust appears to be thrusting at a shallow angle under the Tibetan continental crust. This causes crustal thickening toward the north. In

**Fig. 3.43** Fault-plane solutions for earthquakes in or near the Swiss Alps in central Europe. The arrows show the horizontal components of the interpreted axes of maximum compression. They are approximately at right angles to the strike of the mountain ranges of the Jura and Alps (after Mayer Rosa and Müller, 1979, and Pavoni, 1977).



the main chain of the Lesser Himalaya the *minimum* compressive stress is vertical. The increased vertical load due to crustal thickening causes the directions of the principal stresses to change, so that under southern Tibet the *maximum* compressive stress is vertical.

A different type of collisional tectonics is shown by fault-plane solutions from the Alpine mountain belt in south-central Europe. The Alps were formed during the collision between the African and European plates, starting in the Early Tertiary. The focal mechanisms of many, mostly small earthquakes show that they are predominantly associated with strike-slip faults (Fig. 3.43). The horizontal projections of the compressional axes are oriented almost perpendicular to the strike of the Alps and rotate along the arc. The fault-plane solutions for the modern seismicity indicate that the Alpine fold-belt is a region of continuing interplate collision.

### 3.5.5 Secondary effects of earthquakes: landslides, tsunami, fires and fatalities

Before discussing methods of estimating the size of earthquakes, it is worth considering some secondary effects that can accompany large earthquakes: landslides, seismic sea waves and conflagrations. These rather exceptional effects cannot be included conveniently in the definition of earthquake *intensity*, because their consequences cannot be easily generalized or quantified. For example, once a major fire has been initiated, other factors not related directly to the size of the earthquake (such as aridity of foliage, combustibility of building materials, availability and efficiency of fire-fighting equipment) determine how it is extinguished.

A tsunami propagates throughout an ocean basin as a wave with period  $T$  of around 15–30 min. The entire water column participates in the motion. As a result, the velocity of the wave,  $v$ , is dependent on the water depth,  $d$ , and the acceleration due to gravity,  $g$ , and is given by:

$$v = \sqrt{gd} \quad (3.89)$$

Over an ocean basin with water depth greater than 4 km, the tsunami velocity is higher than  $200 \text{ m s}^{-1}$

### 3.5 EARTHQUAKE SEISMOLOGY

#### Box 3.3: Tsunami

When the ocean floor is abruptly lifted or dropped by a submarine earthquake or landslide, the entire water column is pushed up and down. More than 95% of the potential energy of the displaced water is gravitational and less than 5% is elastic energy resulting from compression of the ocean floor or water column. The potential energy of the vertical motion is converted to kinetic energy and propagates away from the source as a tsunami. The propagation of ocean waves is a complex non-linear problem that cannot be handled here, but approximate solutions are useful for the special case of a tsunami.

Ocean surface waves with periods shorter than 50 s are confined to the top kilometer of the ocean and cannot be excited by motions of the sea floor. The propagation of ocean waves is in general dispersive, i.e., the wave velocity depends on its wavelength and period (see Section 3.3.3.3). Figure B3.3 shows the computed wave velocities for different periods in water of different depths. The wave speeds are dispersive for periods shorter than about 300 s (5 minutes). These waves have very long wavelengths, much greater than the ocean depth.

For a wave with period  $T$  (angular velocity  $\omega = 2\pi/T$ ) and wavelength  $\lambda$  (wave number  $k = 2\pi/\lambda$ ), the dispersion in an ocean of depth  $d$  is governed by the relation

$$\omega^2 = gk \tanh(kd) \quad (1)$$

where  $g$  is the acceleration of gravity and  $\tanh(x)$  is the hyperbolic tangent function of  $x$ :

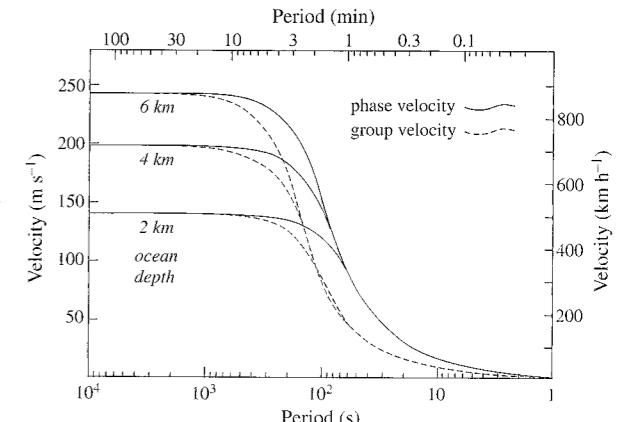
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \begin{cases} x, & \text{if } x \text{ small} \\ 1, & \text{if } x \text{ large} \end{cases} \quad (2)$$

If the wavelength  $\lambda$  is very much greater than the ocean depth,  $\tanh(kd)$  can be replaced in Eq. (1) by  $(kd)$  and the dispersion relation becomes

$$\begin{aligned} \omega^2 &= gk(kd) = k^2gd \\ \omega &= k\sqrt{gd} \end{aligned} \quad (3)$$

Analogously to surface waves, a tsunami travels across an ocean as a packet of waves of different period. The *phase velocity* ( $c$ ) of a wave is the speed of an individual phase in the packet. Using Eq. (3) and Eq. (3.62), the phase velocity for a tsunami is given by

$(720 \text{ km h}^{-1}, 450 \text{ m.p.h})$ , and the wavelength (equal to the product  $vT$ ) may measure 200 km (Box 3.3). The amplitude of a tsunami over the open ocean is comparatively small; the Sumatra tsunami measured about 80–100 cm from crest to trough in the open Indian Ocean. Despite the speed of the wave, an observer on a ship would be scarcely aware of the passage of such a low-amplitude, long-



**Fig. B3.3** Dependence of the phase-velocity ( $c$ ) and group-velocity ( $U$ ) of a tsunami on the period of the wave for different ocean depths (source: S. N. Ward, personal communication, 2006).

$$c = \frac{\omega}{k} = \sqrt{gd} \quad (4)$$

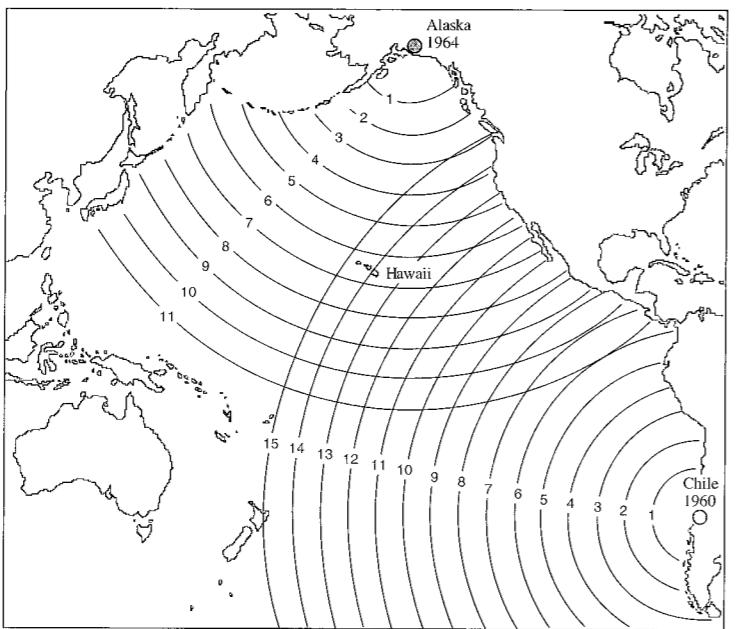
The *group velocity* ( $U$ ) is the propagation speed of the envelope of a wave packet, and thus is the speed with which the energy of the tsunami travels. Using Eq. (3) and Eq. (3.82), the group velocity for a tsunami is given by

$$U = \frac{\partial \omega}{\partial k} = \sqrt{gd} \quad (5)$$

For water depths that are much less than the wavelength the phase and group velocities are equal and the propagation of a tsunami is non-dispersive. These relations are valid for wavelengths greater than about 11 times the water depth. Over an ocean basin that is 4 km deep the propagation velocity is  $200 \text{ m s}^{-1}$  ( $720 \text{ km h}^{-1}$ ) and the wavelengths for periods of 1000–2000 s (around 15–30 min) are 200–400 km. When the tsunami approaches shore and the water depth decreases, its velocity slows. The advancing wave loses kinetic energy, part of which is lost as friction with the ocean bottom and part converted into potential energy. This is evident as an increase in wave height, from a few tens of centimeters over the open ocean to many meters close to land. The contact with land is less often as a ferocious breaking wave than as a rapid increase in sea-level – as during high tides – accompanied by fierce currents which may sweep far inland and withdraw in the same way.

wavelength disturbance. However, on approaching shallower water the leading part of the tsunami slows down and tends to be overridden by the following water mass, so that the height of the wave increases. The wave height may be amplified by the shapes of the sea-bottom and the coastline to several meters. In 1896 a large earthquake raised the ocean-bottom off the southern shore of Japan

**Fig. 3.44** Propagation of tsunami waves across the Pacific Ocean following the 1960 Chilean earthquake ( $M_w = 9.5$ ) and the 1964 Alaskan earthquake ( $M_w = 9.2$ ). Numbers on the wavefronts show the travel-times in hours.



and initiated a tsunami that raced ashore with an estimated wave height of more than 20 m, causing 26,000 fatalities. One of the best-studied tsunami was set off by a great earthquake in the Aleutian islands in 1946. It travelled across the Pacific and several hours later reached Hilo, Hawaii, where it swept ashore and up river estuaries as a wave 7 m high. A consequence of the devastation by this tsunami around the Pacific basin was the formation of the Tsunami Warning System. When a major earthquake is detected that can produce a tsunami, a warning is issued to alert endangered regions to the imminent threat. The system works well far from the source. It may take several hours for the tsunami to arrive at a distant location, as illustrated by records of tsunami propagation for the 1960 Chilean and 1964 Alaskan earthquakes (Fig. 3.44). This allows time to warn and evacuate the population in many places far from the epicenter. However, tsunami casualties may still occur near to the epicenter of the earthquake.

The tsunami warning system is currently effective only in the Pacific Ocean. The 2004 Sumatra earthquake triggered the worst tsunami in recorded history with probably more than 250,000 fatalities in Indonesia, Thailand, Sri Lanka and other countries bordering the Indian Ocean as far away as Somalia. As a consequence of this disaster tsunami warning systems are planned for the Indian and other marine basins.

In addition to causing direct damage to man-made structures, an earthquake can disrupt subterranean supply routes (e.g., telephone, electrical and gas lines) which in turn increases the danger of explosion and fire. Aqueducts and underground water pipelines may be broken, with serious consequences for the inhibition or suppression of fires. The San Francisco earthquake of 1906 was very powerful. The initial shock caused widespread damage, including the disruption of water supply lines. But a great fire followed the earthquake, and because the water supply

lines were broken by the tremor, it could not be extinguished. The greatest damage in San Francisco resulted from this conflagration.

### 3.5.6 Earthquake size

There are two methods of describing how large an earthquake is. The *intensity* of the earthquake is a subjective parameter that is based on an assessment of visible effects. It therefore depends on factors other than the actual size of the earthquake. The *magnitude* of an earthquake is determined instrumentally and is a more objective measure of its size, but it says little directly about the seriousness of the ensuing effects. Illogically, it is usually the magnitude that is reported in news coverage of a major earthquake, whereas the intensity is a more appropriate parameter for describing the severity of its effects on mankind and the environment.

#### 3.5.6.1 Earthquake intensity

Large earthquakes produce alterations to the Earth's natural surface features, or severe damage to man-made structures such as buildings, bridges and dams. Even small earthquakes can result in disproportionate damage to these edifices when inferior constructional methods or materials have been utilized. The intensity of an earthquake at a particular place is classified on the basis of the local character of the visible effects it produces. It depends very much on the acuity of the observer, and is in principle subjective. Yet, intensity estimates have proved to be a viable method of assessing earthquake size, including historical earthquakes.

The first attempt to grade earthquake severity was made in the late eighteenth century by Domenico Pignataro, an Italian physician, who classified more than 1000 earth-

### 3.5 EARTHQUAKE SEISMOLOGY

**Table 3.1** Abridged and simplified version of the European Macroseismic Scale 1998 (European Seismological Commission, 1998) for earthquake intensity

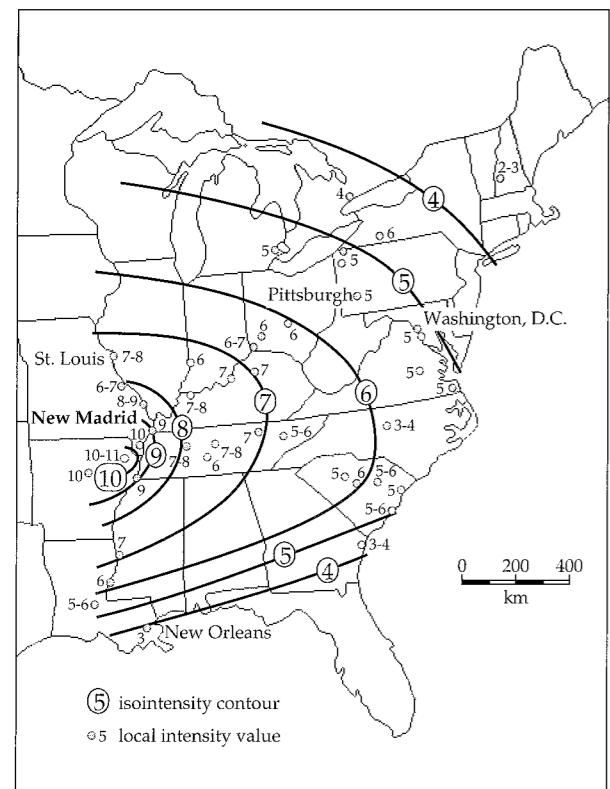
The scale focuses especially on the effects on people and buildings. It takes into account classifications of both the vulnerability of a structure (i.e., the materials and method of construction) and the degree of damage.

Intensity	Description of effects
<i>I–IV light to moderate earthquakes</i>	
<b>I Not felt</b>	
<b>II Scarcely felt</b> Felt only by a few individual people at rest in houses.	
<b>III Weak</b> Felt indoors by a few people. People at rest feel a swaying or light trembling.	
<b>IV Largely observed</b> Felt indoors by many people; outdoors by very few. A few people are awakened. Windows, doors and dishes rattle.	
<i>V–VIII moderate to severe earthquakes</i>	
<b>V Strong</b> Felt indoors by most, outdoors by few. Many sleeping people awake. A few are frightened. Buildings tremble throughout. Hanging objects swing considerably. Small objects are shifted. Doors and windows swing open or shut.	
<b>VI Slightly damaging</b> Many people are frightened and run outdoors. Some objects fall. Many houses suffer slight non-structural damage like hair-line cracks and fall of small pieces of plaster.	
<b>VII Damaging</b> Most people are frightened and run outdoors. Furniture is shifted and objects fall from shelves in large numbers. Many well built ordinary buildings suffer moderate damage: small cracks in walls, fall of plaster, parts of chimneys fall down; older buildings may show large cracks in walls and failure of fill-in walls.	
<b>VIII Heavily damaging</b> Many people find it difficult to stand. Many houses have large cracks in walls. A few well built ordinary buildings show serious failure of walls, while weak older structures may collapse.	
<i>IX–XII severe to destructive earthquakes</i>	
<b>IX Destructive</b> General panic. Many weak constructions collapse. Even well built ordinary buildings show very heavy damage: serious failure of walls and partial structural failure.	
<b>X Very destructive</b> Many ordinary well built buildings collapse.	
<b>XI Devastating</b> Most ordinary well built buildings collapse, even some with good earthquake resistant design are destroyed.	
<b>XII Completely devastating</b> Almost all buildings are destroyed.	

quakes that devastated the southern Italian province of Calabria in the years 1783–1786. His crude analysis classified the earthquakes according to whether they were very strong, strong, moderate or slight. In the mid-nineteenth century an Irish engineer, Robert Mallet, produced a list of 6831 earthquakes and plotted their estimated locations, producing the first map of the world's seismicity and establishing that earthquakes occurred in distinct zones. He also used a four-stage intensity scale to grade earthquake damage, and constructed the first isoseismal maps with lines that outlined areas with broadly equal grades of damage. The Rossi–Forel intensity scale, developed in the late nineteenth century by the Italian scientist M. S. de Rossi and the Swiss scientist F. Forel, incorporated ten stages describing effects of increasing damage. In 1902 an Italian seismologist, G. Mercalli, proposed a still more extensive, expanded intensity scale which reclassified earthquake severity in twelve stages. A variation, the Modified Mercalli (MM) scale, was developed in 1931 to suit building conditions in the United States, where a later modification is in common use. The Medvedev–Sponheuer–Karnik (MSK) scale, introduced in Europe in 1964, and modified in 1981, also has twelve stages and differs from the MM scale mainly in details. A new European Macroseismic Scale (EMS-98) was adopted in 1998; an abridged version

is shown in Table 3.1. The new 12-stage EMS scale is based on the MSK scale but takes into account the vulnerability of buildings to earthquake damage and incorporates more rigorous evaluation of the extent of damage to structures with different building standards.

In order to evaluate the active seismicity of a region questionnaires may be distributed to the population, asking for observations that can be used to estimate the intensity experienced. The questionnaires are evaluated with the aid of an intensity scale, and the intensity recorded at the location of each observer is plotted on a map. Continuous lines are then drawn to outline places with the same intensity (Fig. 3.45), in the same way that contour lines are used on topographic maps to show elevation. Comparison of the *isoseismal* maps with geological maps helps explain the response of the ground to the shake of an earthquake. This is valuable information for understanding earthquake risk. The foundation on which structures are erected plays a vital role in their survival of an earthquake. For example, soft sediments can amplify the ground motion, enhancing the damage caused. This is even more serious when the sediments have a high water content, in which case liquefaction of the sediments can occur, robbing structures built on them of support and promoting their collapse.



**Fig. 3.45** Isoseismal map with contours of equal intensity for the New Madrid, Missouri, earthquake of 1811 (after Nuttli, 1973).

There are numerous examples of this having occurred. In the great Alaskan earthquake in 1964 a section of Anchorage that was built on a headland underlain by a wet clay substratum collapsed and slid downslope to the sea. In 1985 a very large earthquake with magnitude 8.1 struck the Pacific coast of Mexico. About 350 km away in Mexico City, despite the large distance from the epicenter, the damage to buildings erected on the alluvium of a drained lakebed was very severe while buildings set on a hard rock foundation on the surrounding hills suffered only minor damage. In the Loma Prieta earthquake of 1989 severe damage was caused to houses built on landfill in San Francisco's Mission district, and an overhead freeway built on pillars on young alluvium north of Oakland collapsed dramatically. Both regions of destruction were more than 70 km from the epicenter in the Santa Cruz mountains. Similarly, in the San Francisco earthquake of 1906 worse damage occurred to structures built on landfill areas around the shore of the bay than to those with hard rock foundations in the hills of the San Francisco peninsula.

Intensity data play an important role in determining the historic seismicity of a region. An earthquake has dramatic consequences for a population; this was especially the case in the historic past, when real hazards were augmented by superstition. The date (and even the time) of occurrence of strong earthquakes and observations of their local effects have been recorded for centuries in church and civil documents. From such records it is sometimes possible to extract enough information for a given

earthquake to estimate the intensity experienced by the observer. If the population density is high enough, it may be possible to construct an isoseismal map from which the epicenter of the tremor may be roughly located. An interesting example of this kind of analysis is the study of the New Madrid earthquakes of 1811–1812, which caused devastation in the Mississippi valley and were felt as far away as the coastlines along the Atlantic and Gulf of Mexico (Fig. 3.45). There were probably three large earthquakes, but the events occurred before the invention of the seismograph so details of what happened are dependent on the subjective reports of observers. Historical records of the era allow development of an intensity map for the settled area east of the Mississippi, but the pioneering population west of the river was at that time too sparse to leave adequate records for intensity interpretation. On the basis of the available evidence these earthquakes are estimated to have had magnitudes of 7.8–8.1.

Earthquake intensity data are valuable for the construction of seismic risk maps, which portray graphically the estimated earthquake hazard of a region or country. The preparation of a seismic risk map is a lengthy and involved task, which typically combines a study of the present seismicity of a region with analysis of its historic seismicity. A map of the maximum accelerations experienced in recent seismicity helps to identify the areas that are most likely to suffer severe damage in a large earthquake. The likelihood of an earthquake happening in a given interval of time must also be taken into account. Even in a region where large earthquakes occur, they happen at irregular intervals, and so knowledge of local and regional earthquake frequency is important in risk estimation. A seismic risk map of the United States (Fig. 3.46) shows the peak accelerations that are likely to be experienced at least once in a 50-year period.

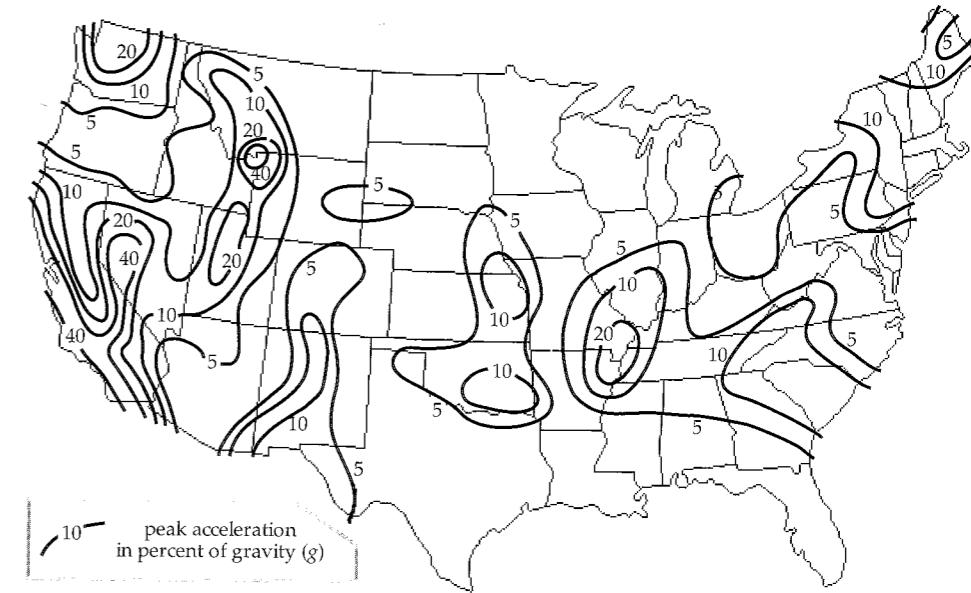
Devastating earthquakes can occur even in countries with relatively low seismic risk. Switzerland, in the center of Europe, is not regarded as a country prone to earthquakes. The seismic activity consists mostly of small earthquakes, mainly in the Alps in the collision zone of the African and European plates. Yet, in 1356 the northern Swiss town of Basel was destroyed by a major earthquake. Seismic risk maps are useful in planning safe sites for important edifices like nuclear power plants or high dams for hydroelectric power, which supply a substantial proportion of Switzerland's energy needs. Risk maps are also valuable to insurance companies, which must know the seismic risk of a region in order to assess the costs of earthquake insurance coverage for private and public buildings.

### 3.5.6.2 Earthquake magnitude

Magnitude is an experimentally determined measure of the size of an earthquake. In 1935 C. F. Richter attempted to grade the sizes of local earthquakes in Southern California on the basis of the amplitude of the ground vibrations they produced at a known distance from the

## 3.5 EARTHQUAKE SEISMOLOGY

**Fig. 3.46** Seismic risk map of the United States. The numbers on contour lines give the maximum acceleration (in percent of  $g$ ) that might be expected to be exceeded with a probability of 1 in 10 during a 50-year period (after Bolt, 1988).



epicenter. The vibrations were recorded by seismographs, which were standardized to have the same response to a given stimulus. Richter's original definition of magnitude was based on surface-wave amplitudes ( $A_s$ ) recorded by seismographs at an epicentral distance of 100 km. Because seismographs were located at various distances from the earthquake, an extra term was added to compensate for attenuation of the signal with increasing epicentral distance. Increasingly sensitive instruments allowed the recording of signals from distant earthquakes; those from events with epicentral distances greater than 20° are known as *teleseismic* signals. Originally, the magnitude was determined from the horizontal ground motion, because seismological stations were equipped mainly with horizontal-motion seismometers. However, the surface waves recorded by these instruments consist of superposed Love and Rayleigh waves, which complicates the theoretical interpretation of the records. Vertical-motion seismometers record only the Rayleigh waves (together with P- and SV-waves), and so progressively the definition of surface-wave magnitude has come to be based on the vertical component of motion. The majority of surface-wave magnitudes assigned to earthquakes worldwide are now based on vertical-motion records.

The International Association for Seismology and Physics of the Earth's Interior (IASPEI) has adopted the following definition of the surface-wave magnitude ( $M_s$ ) of an earthquake:

$$M_s = \log_{10}\left(\frac{A_s}{T}\right) + 1.66\log_{10}(\Delta) + 3.3 \quad (3.90)$$

where  $A_s$  is the vertical component of the ground motion in micrometers ( $\mu\text{m}$ ) determined from the maximum Rayleigh-wave amplitude,  $T$  is the period of the wave (18–22 seconds),  $\Delta$  is the epicentral distance in degrees ( $20^\circ \leq \Delta \leq 160^\circ$ ), and where the earthquake has a focal depth of less than 50 km. A similar equation to Eq. (3.90)

applies to broadband recordings; in this case  $(A_s/T)_{\max}$  corresponds to the maximum ground velocity. The surface-wave magnitudes of some important historical earthquakes are given in Table 3.2.

The depth of the source affects the nature of the seismic wave train, even when the same energy is released. An earthquake with a deep focus may generate only a small surface-wave train, while shallow earthquakes cause very strong surface waves. Equation (3.90) for  $M_s$  was derived from the study of shallow earthquakes, observed at a distance greater than 20°. Therefore, corrections must be made to the computed value of  $M_s$  to compensate for the effects of a focal depth greater than 50 km or epicentral distance less than 20°.

The amplitude of body waves is not sensitive to the focal depth. As a result, earthquake magnitude scales have also been developed for use with body waves. An equation, proposed by B. Gutenberg in 1945, can be used to calculate a *body-wave magnitude* ( $m_b$ ) from the maximum amplitude ( $A_p$ ) of the ground motion associated with P-waves having a period ( $T$ ) of less than 3 s:

$$m_b = \log_{10}\left(\frac{A_p}{T}\right) + Q(\Delta, h) \quad (3.91)$$

Where  $Q(\Delta, h)$  is an empirical correction for signal attenuation due to epicentral distance ( $\Delta$ ) and focal depth ( $h$ ) that is made by reading directly from a graph or table of values.

For some earthquakes both  $M_s$  and  $m_b$  can be calculated. Unfortunately, the different estimates of magnitude often do not agree well, except for small earthquakes. This is due to the way the ground responds to a seismic event, and to the different natures of body waves and surface waves. Body waves have a different dependence of amplitude on frequency than do surface waves.  $m_b$  is estimated from a high-frequency (1 Hz) phase whereas  $M_s$  is determined from

Table 3.2 Some important historical earthquakes, with their surface-wave magnitudes  $M_s$ , moment magnitudes  $M_w$ , and the numbers of fatalities

Year	Epicenter	Magnitude		
		$M_s$	$M_w$	Fatalities
1906	San Francisco, California	8.3	7.8	3,000
1908	Messina, Italy	7.2	—	70,000
1923	Kanto, Japan	8.2	7.9	143,000
1952	Kamchatka, Russia	8.2	9.0	
1957	Andreanof Islands, Alaska	8.1	8.6	
1960	Valdivia, Chile	8.5	9.5	5,700
1960	Agadir, Morocco	5.9	5.7	10,000
1964	Prince William Sound, Alaska	8.6	9.2	125
1970	Chimbote, Peru	7.8	7.9	66,000
1971	San Fernando, California	6.5	6.7	65
1975	Haicheng, China	7.4	7.0	≈300
1976	Tangshan, China	7.8	7.5	243,000
1980	El Asnam, Algeria	7.3	—	2,590
1985	Michoacan, Mexico	8.1	8.0	9,500
1989	Loma Prieta, California	7.1	6.9	62
1994	Northridge, California	6.8	6.7	60
1999	Ismi, Turkey	—	7.6	17,100
2004	Sumatra–Andaman Islands	—	9.0	250,000

low-frequency (0.05 Hz) vibrations. Above a certain size, each method becomes insensitive to the size of the earthquake, and exhibits *magnitude saturation*. This occurs for body-wave estimates at around  $m_b=6$ ; all larger earthquakes give the same body-wave magnitude. Similarly, surface-wave estimates of magnitude saturate at  $M_s=8$ . Thus, for very large earthquakes,  $M_s$  and  $m_b$  underestimate the energy released. An alternative definition of magnitude, based upon the long-period spectrum of the seismic wave, is preferred for very large earthquakes. It makes use of the physical dimensions of the focus.

As discussed in the elastic rebound model (Section 3.1), a tectonic earthquake arises from abrupt displacement of a segment of a fault. The area  $S$  of the fractured segment and the amount by which it slipped  $D$  can be inferred. Together with the rigidity modulus  $\mu$  of the rocks adjacent to the fault, these quantities define the *seismic moment*  $M_0$  of the earthquake. Assuming that the displacement and rigidity are constant over the area of the rupture:

$$M_0 = \mu S D \quad (3.92)$$

The seismic moment can be used to define a *moment magnitude* ( $M_w$ ). The definition adopted by the responsible commission of IASPEI is:

$$M_w = \frac{2}{3}(\log_{10} M_0 - 9.1) \quad (3.93)$$

In this equation  $M_0$  is in N m. If c.g.s. units are used instead of SI units,  $M_0$  is expressed in dyne cm and the corresponding equation is:

$$M_w = \frac{2}{3}(\log_{10} M_0 - 16.1) \quad (3.94)$$

$M_w$  is more appropriate for describing the magnitudes of very large earthquakes. It has largely replaced  $M_s$  in scientific evaluation of earthquake size, although  $M_s$  is often quoted in reports in the media. The moment magnitudes and surface-wave magnitudes of some historical earthquakes are listed in Table 3.2.

The magnitude scale is, in principle, open ended. Negative Richter magnitudes are possible, but the limit of sensitivity of seismographs is around –2. The maximum possible magnitude is limited by the shear strength of the crust and upper mantle, and since the beginning of instrumental recording none has been observed with a surface-wave magnitude  $M_s$  as high as 9. However, this is largely due to a saturation effect resulting from the method by which surface-wave magnitudes  $M_s$  are computed. Seismic moment magnitudes  $M_w$  of 9 and larger have been computed for some giant earthquakes (Table 3.2). The largest in recorded history was the 1960 Chile earthquake with  $M_w=9.5$ .

When an earthquake occurs on a fault, the ruptured area spreads in size from the point of initial failure, akin to opening a zip fastener. If the fault ruptures along a length  $L$ , and the fractured segment has a down-dip dimension  $w$  (referred to as the width of the faulted zone), the area  $S$  is equal to  $wL$ . Assuming that the aspect ratio of faults is constant (i.e.,  $w$  is proportional to  $L$ ) then the ruptured area  $S$  is proportional to  $L^2$ . This is a generalization, because different faults have different aspect ratios. Similarly, if the stress drop in an earthquake is constant, the displacement on the fault,  $D$ , can be assumed proportional to  $L$ . Together, these assumptions imply that the seismic moment  $M_0$  scales as  $L^3$ . Assuming that  $S \propto L^2$ , the seismic moment scales as  $S^{3/2}$ . This inference is supported

### 3.5 EARTHQUAKE SEISMOLOGY

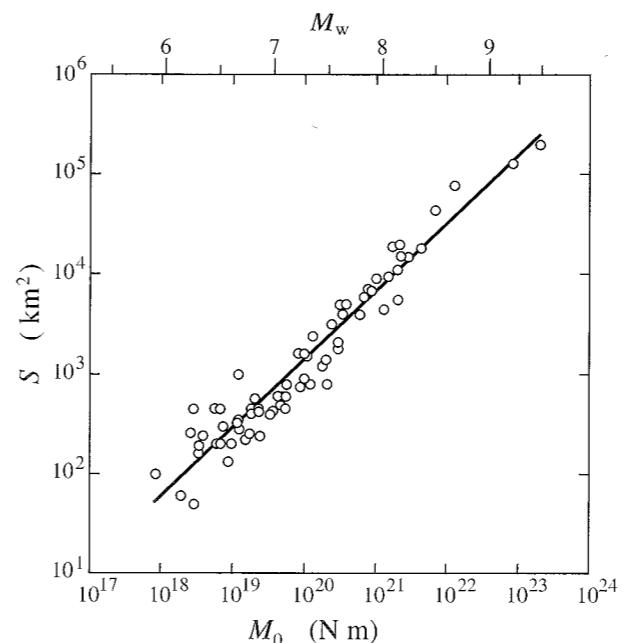


Fig. 3.47 Correlation of the ruptured area ( $S$ , in  $\text{km}^2$ ) with the seismic moment ( $M_0$ , in  $\text{N m}$ ) and moment magnitude ( $M_w$ ) for some shallow earthquakes (after Kanamori and Brodsky, 2004).

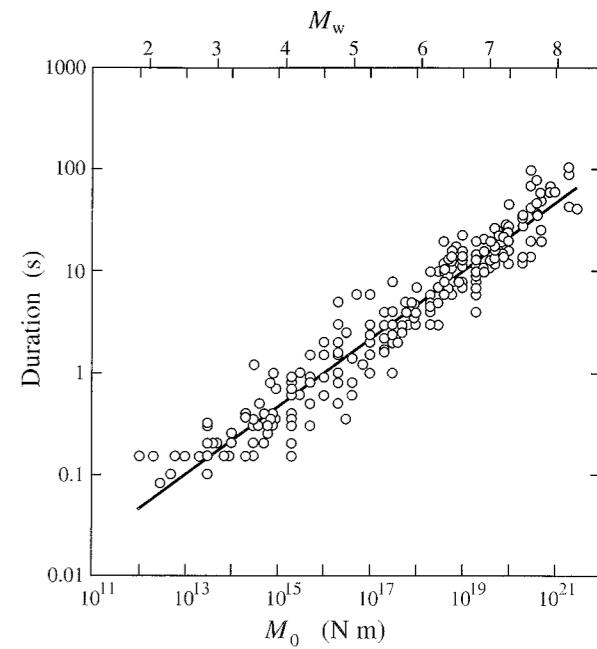


Fig. 3.48 Correlation of the source durations (in seconds) of shallow earthquakes and their seismic moments,  $M_0$  (in  $\text{N m}$ ) and moment magnitudes,  $M_w$  (after Kanamori and Brodsky, 2004).

by the correlation between seismic moment and the ruptured area  $S$  for many shallow earthquakes (Fig. 3.47).

The length  $L$  of the rupture zone determines how long the ground motion lasts in an earthquake. This factor has a significant bearing on the extent of damage to structures during moderate to strong earthquakes. Theory gives the speed of propagation of the rupture as 75–95% of the shear wave velocity. Assuming that the rupture propagates along the fault at about  $2.5\text{--}3.0 \text{ km s}^{-1}$ , the duration of the ground motion can be estimated for an earthquake of a given magnitude or seismic moment (Fig. 3.48). For example, close to the epicenter of a magnitude 5 earthquake the vibration may last a few seconds, whereas in a magnitude 8 earthquake it might last about 50 s. However, there is a lot of scatter about the ideal correlation. In the great Sumatra earthquake, with magnitude 9.0, the rupture continued for some 500 s.

The seismic moment  $M_0$  of an earthquake is determined in practice from the analysis of seismograms at a large number of seismic stations (currently 137) distributed world-wide and linked to form the Global Seismographic Network (GSN). The waveform of each seismogram is a product of the epicentral distance, focal depth and radiation pattern of the earthquake (Section 3.5.4.1). It is also affected by source parameters such as the area, orientation and amount of slip of the ruptured segment, factors which determine the seismic moment  $M_0$ . Inversion of the seismogram data leads to an understanding of the earthquake source. At large distances and for long-period components of the seismogram (e.g.,  $T>40 \text{ s}$ ) that correspond to wavelengths much longer than the dimensions of the faulted area, the source can be considered as a point source. Assuming a radiation model for the source and a velocity

model for the propagation, a synthetic seismogram (see also Section 3.6.5.4) can be computed for each station and compared with the observed waveform. Iterative adjustment of the source parameters to give a best fit to the waveforms at a number of receivers defines mathematically a *moment tensor*. The tensor can be visualized geometrically as an ellipsoid in which the lengths and orientations of the axes correspond to the moment of the earthquake and the directions of the tensional (T) axis and compressional (P) axis at the source. The location of the system of couples modelled by the moment tensor is the optimum point source location for the earthquake and is called its *centroid*. The *centroid moment tensor* (CMT) analyses of earthquakes are made rapidly and give the source parameters of an earthquake within a few hours. The centroid location may differ from the hypocenter of the earthquake (the place where rupture first occurred). The former is based on analysis of the full seismogram waveform, the latter only on the first arrival times. So, for example, the centroid of the 2004 Sumatra–Andaman earthquake was located about 160 km west of its epicenter.

#### 3.5.6.3 Relationship between magnitude and intensity

The intensity and magnitude scales for estimating the size of an earthquake are defined independently but they have some common features. Intensity is a measure of earthquake size based on the extent of local damage it causes at the location of an observer. The definition of magnitude is based on the amplitude of ground motion inferred from the signal recorded by the observer's seismograph, and of course it is the nature of the ground motion – its amplitude,

Table 3.3 Average number of earthquakes per year worldwide since 1990, except for  $M_s \geq 8$  which are averaged since 1900 (based on data from the US Geological Survey National Earthquake Information Center). The mean annual release of seismic energy is estimated using the energy-magnitude relation in Eq. (3.97)

Earthquake magnitude	Number per year	Annual energy [ $10^{15} \text{ J yr}^{-1}$ ]
$\geq 8.0$	$\approx 1$	$\approx 100$
7–7.9	17	190
6–6.9	134	45
5–5.9	1,319	14
4–4.9	$\approx 13,000$	4
3–3.9	$\approx 130,000$	1
2–2.9	$\approx 1,300,000$	0.4

velocity and acceleration – which produce the local damage used to classify intensity. However, in the definition of magnitude the ground-motion amplitude is corrected for epicentral distance and converted to a focal characteristic. Isoseismal maps showing the regional distribution of damage give the maximum intensity ( $I_{\max}$ ) experienced in an earthquake, which, although influenced by the geographic patterns of population and settlement, is usually near to the epicenter. A moderately strong, shallow-focus earthquake under a heavily populated area can result in higher intensities than a large deep focus earthquake under a wilderness area (compare, for example, the death tolls for the 1960 Agadir magnitude 5.7 and 1964 Alaskan magnitude 9.2 earthquakes in Table 3.2). However, for earthquakes with focal depth  $h < 50 \text{ km}$  the dependence of  $I_{\max}$  on the focal depth can be taken into account, and it is possible to relate the maximum intensity to the magnitude with an empirical equation (Karnik, 1969):

$$I_{\max} = 1.5M_s - 1.8\log_{10}h + 1.7 \quad (3.95)$$

This type of equation is useful for estimating quickly the probable damage that an earthquake causes. For example, it predicts that in the epicentral region of an earthquake with magnitude 5 and a shallow focal depth of 10 km, the maximum MSK intensity will be VII (moderately serious damage), whereas, if the focal depth is 100 km, a maximum intensity of only IV–V (minor damage) can be expected.

### 3.5.7 Earthquake frequency

Every year there are many small earthquakes, and only a few large ones. According to a compilation published by Gutenberg and Richter in 1954, the mean annual number of earthquakes in the years 1918–1945 with magnitudes 4–4.9 was around 6000, while there were only an average of about 100 earthquakes per year with magnitudes 6–6.9. The relationship between annual frequency ( $N$ ) and magnitude ( $M_s$ ) is logarithmic and is given by an equation of the form

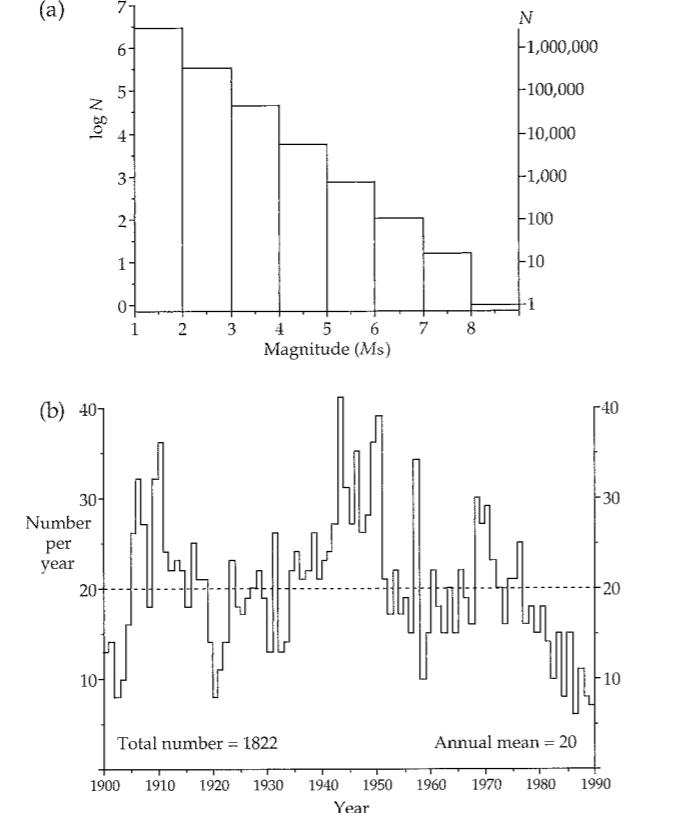


Fig. 3.49 Histograms of (a) the logarithm of the number ( $N$ ) of earthquakes per year with magnitude  $M_s$ , and (b) the annual number of earthquakes with magnitudes  $M_s \geq 7$  since 1900 (based on data from the US Geological Survey National Earthquake Information Center).

$$\log N = a - bM_s \quad (3.96)$$

The value of  $a$  varies between about 8 and 9 from one region to another, while  $b$  is approximately unity for regional and global seismicity. The mean annual numbers of earthquakes in different magnitude ranges are listed in Table 3.3; the frequency decreases with increasing magnitude (Fig. 3.49a), in accordance with Eq. (3.96). The annual number of large earthquakes with magnitude  $M_s \geq 7$  in the years 1900–1989 has varied between extremes of about 10 and 40, but the long-term average is about 20 per year (Fig. 3.49b).

### 3.5.8 Energy released in an earthquake

The definition of earthquake magnitude relates it to the logarithm of the amplitude of a seismic disturbance. Noting that the energy of a wave is proportional to the square of its amplitude it should be no surprise that the magnitude is also related to the logarithm of the energy. Several equations have been proposed for this relationship. An empirical formula worked out by Gutenberg (1956), relates the energy release  $E$  to the surface-wave magnitude  $M_s$ :

### 3.5 EARTHQUAKE SEISMOLOGY

$$\log_{10}E = 4.8 + 1.5M_s \quad (3.97)$$

where  $E$  is in joules. The logarithmic nature of the formula means that the energy release increases very rapidly with increasing magnitude. For example, when the magnitudes of two earthquakes differ by 1, their corresponding energies differ by a factor 32 ( $= 10^{1.5 \times 2}$ ). Hence, a magnitude 7 earthquake releases about 1000 ( $= 10^{1.5 \times 2}$ ) times the energy of a magnitude 5 earthquake. Another way of regarding this observation is that it takes 1000 magnitude 5 earthquakes to release the same amount of energy as a single large earthquake with magnitude 7. Multiplying the mean number of earthquakes per year by their estimated energy (using one of the energy-magnitude equations) gives an impression of the importance of very large earthquakes. Table 3.3 shows that the earthquakes with  $M_s \geq 7$  are responsible for most of the annual seismic energy. In a year in which a very large earthquake ( $M_s \geq 8$ ) occurs, most of the annual seismic energy is released in that single event.

It is rather difficult to appreciate the amount of energy released in an earthquake from the numerical magnitude alone. A few examples help illustrate the amounts of energy involved. Earthquakes with  $M_s = 1$  are so weak that they can only be recorded instrumentally; they are referred to as microearthquakes. The energy associated with one of these events is equivalent to the kinetic energy of a medium sized automobile weighing 1.5 tons which is travelling at  $130 \text{ km h}^{-1}$  (80 m.p.h.). The energy released by explosives provides another means of comparison, although the conversion of energy into heat, light and shock waves is proportionately different in the two phenomena. One ton of the explosive trinitrotoluene (TNT) releases about  $4.2 \times 10^9$  joules of energy. Equation (3.97) shows that the 11 kiloton atomic bomb which destroyed Hiroshima released about the same amount of energy as an earthquake with magnitude 5.9. The energy released in a 1 megaton nuclear explosion is equivalent to an earthquake with magnitude 7.2; the 2004 Sumatra earthquake (magnitude 9.0) released an amount of energy equivalent to the detonation of 475 megaton bombs.

### 3.5.9 Earthquake prediction

The problem of earthquake prediction is extremely difficult and is associated with sundry other problems of a sociological nature. To predict an earthquake correctly means deciding, as far in advance as possible, exactly where and when it will occur. It is also necessary to judge how strong it will be, which means realistically that people want to know what the likely damage will be, a feature expressed in the earthquake intensity. In fact the geophysicist is almost helpless in this respect, because at best an estimate of the predicted magnitude can be made. As seen above, even if it is possible to predict accurately the magnitude, the intensity depends on many factors (e.g., local geology, construction standards, secondary

effects like fires and floods) which are largely outside the influence of the seismologist who is asked to presage the seriousness of the event. The problem of prediction rapidly assumes sociological and political proportions. Even if the approximate time and place of a major earthquake can be predicted with reasonable certainty, the question then remains of what to do about the situation. Properly, the threatened area should be evacuated, but this would entail economic consequences of possibly enormous dimension.

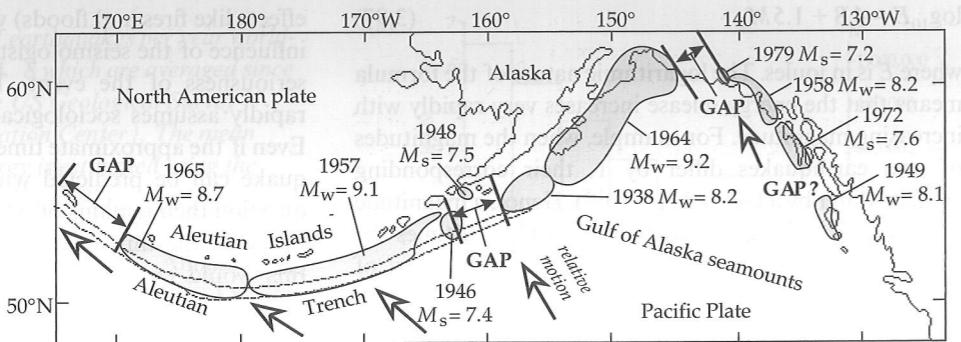
The difficulties are illustrated by the following possible scenario. Suppose that seismologists conclude that a very large earthquake, with probable magnitude 7 or greater, will take place sometime in a known month of a given year under a specific urban center. No more precise details are possible; in particular, the time of occurrence cannot be determined more exactly. Publication of this kind of prediction would cause great consternation, even panic, which for some of the population could be as disastrous as the earthquake itself. Should the warning be withheld? If the entire area, with its millions of inhabitants, were to be evacuated, the economic dislocation would be enormous. How long should the evacuation last, when every day is economically ruinous? Clearly, earthquake prediction is only useful in a practical sense when it is accurate in both place and time. The responsible authorities must also be provided with a reasonable estimate of the expected magnitude, from which the maximum intensity  $I_{\max}$  may be gauged with the aid of a relationship like that of Eq. (3.95). The problem then passes out of the domain of the scientist and into that of the politician, with the scientist retaining only a peripheral role as a consultant. But if the earthquake prediction is a failure, the consequences will rebound with certainty on the scientist.

#### 3.5.9.1 Prediction of the location of an earthquake

It is easier to predict where a major earthquake is likely to occur than when it will occur. The global seismicity patterns demonstrate that some regions are relatively aseismic. They are not completely free of earthquakes, many of which appear to occur randomly and without warning in these regions. Some intraplate earthquakes have a history of repeated occurrence at a known location, which can sensibly be expected to be the locus of a future shock. However, because most earthquakes occur in the seismically active zones at plate margins, these are the prime areas for trying to predict serious events. Predicting the location of a future earthquake in these zones combines a knowledge of the historical seismicity pattern with the elastic rebound model of what causes an earthquake.

The *seismic gap* theory is based on the simple idea that global plates move under the influence of forces which affect the plates as entities. The interactions at a plate margin therefore act along the entire length of the interplate boundary. Models of plate tectonic reconstructions

**Fig. 3.50** Seismic gaps along the Aleutian island arc. Shaded regions mark the areas of rupture of very large historic earthquakes ( $M_s$  or  $M_w > 7.4$ ). Three large gaps in seismicity are potential locations of a future large earthquake (after Sykes et al., 1981).



assume continuity of plate motions on a scale of millions of years. However, the global seismicity patterns show that on the scales of decades or centuries, the process is discontinuous both in time and place. This is because of the way individual earthquakes occur. According to the elastic rebound model, stress accumulates until it exceeds the local strength of the rocks, rupture produces motion on the fault, and an earthquake occurs. During the time of stress accumulation, the area experiences no major earthquake, and the regional pattern of seismicity shows a local gap (Fig. 3.50). This is the potential location for an earthquake that is in the process of accumulating the strain energy necessary to cause rupture. For example, the magnitude 7.6 earthquake at Izmit, Turkey, in 1999 resulted in more than 11,000 deaths. It occurred in a 100–150 km seismic gap between events that occurred more than three decades earlier, in 1967 and 1964, and which subsequently had been tranquil.

### 3.5.9.2 Prediction of the time and size of an earthquake

Seismic gap theory holds great promise as a means of determining where an earthquake is likely to occur. Unfortunately, it does not help to predict when it will occur or how large it will be. These factors depend on the local strength of the rocks and the rate at which strain accumulates. There are various ways to observe the effects of strain accumulation, but the largely unknown factor of local breaking strength of the rocks hinders prediction of the time of an earthquake.

The strain accumulation results in precursory indications of both sociological and scientific nature. The People's Republic of China has suffered terribly from the ravages of earthquakes, partly because of the unavoidable use of low-quality construction materials. In 1966 the highly disciplined society in the People's Republic of China was marshalled to report any strange occurrences associated with earthquakes. They noticed, for example, that wells and ponds bubbled, and sometimes gave off odors. Highly intriguing was the odd behavior of wild and domestic animals prior to many earthquakes. Dogs howled unaccountably, many creatures fell into panic, rats and mice left their holes, snakes abandoned their dens; even fish in ponds behaved in an agitated manner. It is not known how these creatures sense the imminent disaster, and the qualitative reports do not lend themselves to convenient

statistical evaluation. However, prediction by scientific methods is still uncertain, so the usefulness of alternative premonitory phenomena cannot be rejected out of hand.

Several scientific methods have been tested as possible ways of predicting the time of earthquake occurrence. They are based on detecting ground changes, or effects related to them, that accompany the progression of strain. For example, a geochemical method which has had some degree of success is the monitoring of radon. Some minerals in the Earth's crust contain discrete amounts of uranium. The gas radon is a natural product of the radioactive decay of uranium. It migrates through pores and cracks and because of its own radioactivity it is a known environmental health hazard in buildings constructed in some geographic areas. Radon gas can become trapped in the Earth's crust, and in many areas it forms a natural radioactive background. Prior to some earthquakes anomalous levels of radon have been detected. The enhanced leakage of radon from the crust may be due to porosity changes at depth in response to the accumulating strains.

The build-up of strain is manifest in horizontal and vertical displacements of the Earth's surface, depending on the type of faulting involved. These displacements can be measured geodetically by triangulation or by modern techniques of trilateration which employ laser-ranging devices to measure the time taken for a laser beam to travel to a reflecting target and back to its source. The travel-time of the beam is measured extremely accurately, and converted into the distance between emitter and reflector. A shift on the fault will change this distance. With laser techniques the constant creep of one side of a horizontal fault relative to the other can be observed accurately. In one method the source and receiver of a laser beam are placed on one side of a fault, with a reflector on the opposite side. In an alternative method a laser source and receiver are placed on each side of the fault. Pulsed signals are beamed from each unit to an orbiting reflecting satellite, the position of which is known accurately. For each ground station, differences in the elapsed time of the pulse are converted by computer into ground movement, and into differential motion on the fault.

Several methods are suited to detecting differential vertical motion. Sensitive gravimeters on opposite sides of a fault can detect vertical displacement of as little as one centimeter. One instrument relative to the other. Distension

## 3.5 EARTHQUAKE SEISMOLOGY

of the Earth can be monitored with a tiltmeter, which is an instrument designed on the principle of a water-level. It consists of a long tube about 10 m in length, connecting two water-filled containers, in which the difference in water levels is monitored electronically. A tiltmeter is capable of determining tilt changes of the order of  $10^{-7}$  degrees. Tiltmeters and gravimeters installed near to active faults have shown that episodes of ground uplift and tilt can sometimes precede major earthquakes.

Geodetic and geophysical observations, such as changes in the local geomagnetic field or the electrical resistivity of the ground, are of fundamental interest. However, the most promising methods of predicting the time and size of an imminent earthquake are based on seismic observations. According to the elastic rebound model, the next earthquake on an active fault (or fault segment) will occur when the stress released in the most recent earthquake again builds up to the local breaking point of the rocks. Thus the probability at any time that an earthquake will occur on the fault depends on the magnitude of the latest earthquake, the time elapsed since it occurred and the local rate of accumulation of stress. A symptom of the stress build-up before a major earthquake can be an increase in foreshock activity, and this was evidently a key parameter in the successful prediction of a large earthquake in Liaoning province, China, in February, 1975. The frequency of minor shocks increased, at first gradually but eventually dramatically, and then there was an ominous pause in earthquake activity. Chinese seismologists interpreted this "time gap" as an indication of an impending earthquake. The population was ordered to vacate their homes several hours before the province was struck by a magnitude 7.4 earthquake that destroyed cities and communities. Because of the successful prediction of this earthquake, the epicenter of which was near to the city of Haicheng, the death toll among the 3,000,000 inhabitants of the province was very low.

A technique, no longer in favor, but which at one time looked promising for predicting the time and magnitude of an earthquake is the *dilatancy hypothesis*, based upon systematic variations in the ratio of the travel-times of P-waves and S-waves which originated in the focal volume of larger shocks. Russian seismologists noticed that prior to an earthquake the travel-time ratio  $t_s/t_p$  changed systematically: at first it decreased by up to 5%, then it returned to normal values just before the earthquake.

The observations have been attributed to changes in the dilatancy of the ground. Laboratory experiments have shown that, before a rock fractures under stress, it develops minute cracks which cause the rocks to dilate or swell. This dilatancy alters the P-wave velocity, which drops initially (thereby increasing  $t_p$ ) as, instead of being water filled, the new volume of the dilated pores at first fills with air. Later, water seeps in under pressure, replaces the air and the  $t_s/t_p$  ratio returns to normal values. At this point an earthquake is imminent. The time for which the ratio remains low is a measure of the strain energy that is stored and therefore a guide to the magnitude of the earthquake that is eventually

unleashed. Initial success in predicting small earthquakes with the dilatancy model led to a period of optimism that an ultimate solution to earthquake prediction had been found. Unfortunately, it has become apparent that the dilatancy effect is not universal, and its importance appears to be restricted only to certain kinds of earthquakes.

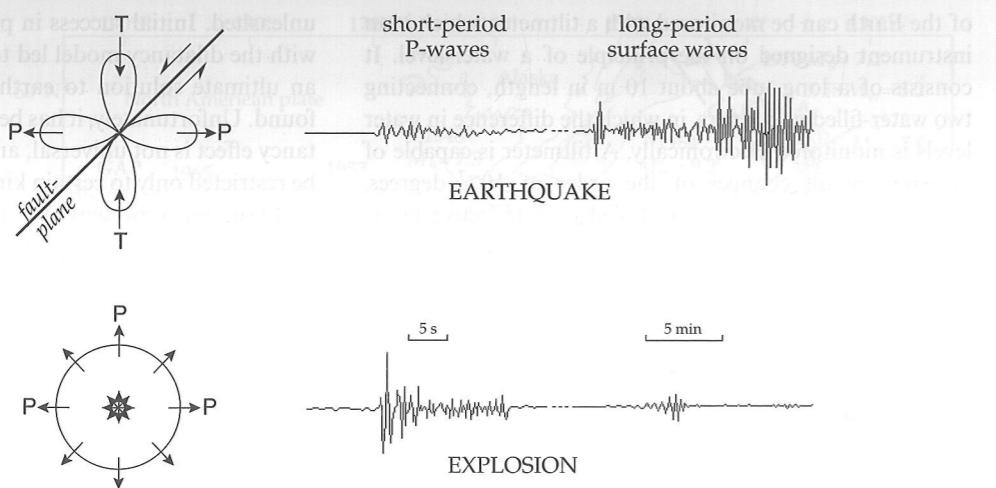
In summary, present seismicity patterns, in conjunction with our knowledge of where historic earthquakes have occurred, permit reasonable judgements of where future earthquakes are most likely to be located. However, despite years of effort and the investigation of various scientific methods, it is still not possible to predict reliably when an earthquake is likely to happen in an endangered area.

### 3.5.10 Earthquake control

Earthquakes constitute a serious natural environmental hazard. Despite great efforts by scientists in various countries, successful prediction is not yet generally possible. Consequently, the protection of people against seismic hazard depends currently on the identification of especially perilous areas (such as active faults), the avoidance of these as the sites of constructions, the development and enforcement of appropriate building codes, and the education and training of the population in emergency procedures to be followed during and in the aftermath of a shock. Unfortunately, many densely populated regions are subject to high seismic risk. It is impossible to prevent the cumulation of strain in a region subject to tectonic earthquakes; the efforts of the human race are not likely to have much effect on the processes of plate tectonics! However, it may be possible to influence the manner in which the strain energy is released. The catastrophic earthquakes are those in which a huge amount of strain energy that has accumulated over a long period of time is suddenly released in a single event. If the energy could be released progressively over a longer period of time in many smaller shakes, the violence and disastrous consequences of a major earthquake might be avoided. The intriguing possibility of this type of earthquake control has been investigated in special situations.

In 1962 the US Army began to dispose of liquid toxic waste from the manufacture of chemical weapons by injection into a well, more than 3 km deep, near Denver, Colorado. Although the region had been devoid of earthquake activity for the preceding 80 years, earthquakes began to occur several weeks after pumping started. Until 1965, when waste injection was halted, more than 1000 earthquakes were recorded. They were mostly very small, of the microearthquake category, but some had Richter magnitudes as high as 4.6. When pumping was halted, the seismicity ceased; when pumping was resumed, the earthquake activity started anew. It was conjectured that the liquid waste had seeped into old faults, and by acting as a kind of lubricant or by increasing the pore pressure, had repeatedly permitted slippage, with an accompanying small earthquake. This incident suggested that it might be

**Fig. 3.51** Comparison of P-wave radiation patterns and the relative amplitudes of long-period and short-period surface waves for an earthquake and a nuclear explosion (after Richards, 1989).



possible to control fault motion, either by injecting fluids to lubricate sections of the fault plane, or by pumping out fluid to lock the fault. This opened the intriguing possibility that, by making alternate use of the two processes, the slippage on a fault might be controlled so that it took place by a large number of small earthquakes rather than by a few disastrous earthquakes.

In 1969 the US Geological Survey carried out a test of this lubrication effect in the depleted Rangely oil field in western Colorado. There were many disused wells in the oil field, through which fluids were pumped in and out of the ground over a considerable area. Meanwhile the local seismicity was monitored. This controlled test agreed with the observations at the Denver site, confirming that the earthquake activity correlated with the amount of fluid injected. Moreover, it was established that the earthquake activity increased when the pore pressure exceeded a critical threshold value, and it ceased when the pressure dropped below this value as the fluid was withdrawn. Despite the apparent success of this experiment, it was agreed that further testing was necessary to explore the validity of the method. An obvious difficulty of testing the modification of seismic activity on critical faults is that the tests must be made in remote areas so as to avoid costly damage caused by the testing. The conditions under which the method may be applicable have not been established definitively.

### 3.5.11 Monitoring nuclear explosions

Since 1963 most nuclear explosions have been conducted underground to prevent the dangerous radioactive fallout that accompanies nuclear explosions conducted underwater or in the atmosphere. The detection and monitoring of such testing activity became important tasks for seismologists. The bilateral Threshold Test Ban Treaty of 1974 between the former Soviet Union and the USA prohibited underground testing of nuclear devices with a yield greater than 150 kilotons of TNT equivalent, which corresponds roughly to an earthquake with magnitude

about 6. Current efforts are underway to establish a global monitoring system to verify compliance with a future Comprehensive Test Ban Treaty. The system should detect nuclear explosions with a yield as low as one kiloton (a well-coupled kiloton explosion has a magnitude of around 4). Detection of these events at distances of several thousand kilometers, and discriminating them from the approximately 7000 earthquakes that occur annually with magnitudes of 4 or above (Table 3.3) poses a monumental challenge to seismologists.

In order to achieve the high detection capability needed to monitor underground testing, many so-called seismic arrays have been set up. An array consists of several individual seismometers, with spacing on the order of a kilometer or less, that feed their output signals in parallel into a centralized data-processing center. By filtering, delaying and summing the signals of the individual instruments, the incoherent noise is reduced and the coherent signal is increased, thus improving the signal-to-noise ratio significantly over that for a single sensor. A local seismic disturbance arrives at the array on an almost horizontal path and triggers the individual seismometers at successively different times, whereas a teleseismic arrival from a very distant source reaches all seismometers in the array at nearly the same time along a steeply inclined path. The development of seismic arrays permitted the analysis of distant weak events. The enhanced sensitivity led to several advances in seismology. Features of the deep structure of the Earth (e.g., the inner core) could be investigated, earthquake location became more accurate and the analysis of focal mechanisms received a necessary impetus. These improvements were essential because of the need to identify correctly the features that distinguish underground nuclear explosions from small earthquakes.

An earthquake is the result of sudden motion of crustal blocks on opposite sides of a fault-plane. The radiation pattern of P-wave amplitude has four lobes of alternating compression and dilatation (Fig. 3.51). The first motions at the surface of the Earth are either pushes away from the source or tugs toward it, depending on the geometry of the

## 3.6 SEISMIC WAVE PROPAGATION

focal mechanism. In contrast, an underground explosion causes outward pressure around the source. The first motions at the surface are all pushes away from the source. Hence, focal mechanism analysis provides an important clue to the nature of the recorded event. Moreover, an explosion produces predominantly P-waves, while earthquakes are much more efficient in also generating surface waves. Consequently, the relative amplitudes of the long-period surface-wave part of the record and of the short-period P-wave part are much higher for an earthquake than for an explosion (Fig. 3.51).

Further discrimination criteria are the epicentral location and the focal depth. Intraplate earthquakes are much less common than earthquakes at active plate margins, so an intraplate event might be suspected to be an explosion. If the depth of a suspicious event is determined with a high degree of confidence to be greater than about 15 km, one can virtually exclude that it is an explosion. Deeper holes have not been drilled due to the great technical difficulty, e.g., in dealing with the high temperatures at such depths.

### 3.6 SEISMIC WAVE PROPAGATION

#### 3.6.1 Introduction

A seismic disturbance is transmitted by periodic elastic displacements of the particles of a material. The progress of the seismic wave through a medium is determined by the advancement of the wavefront. We now have to consider how the wave behaves at the boundary between two media. Historically, two separate ways of handling this problem developed independently in the seventeenth century. One method, using Huygens' principle, describes the behavior of wavefronts; the other, using Fermat's principle, handles the geometry of ray paths at the interface. The eikonal equation (Section 3.3.2.5) establishes that these two methods of treating seismic wave propagation are equivalent.

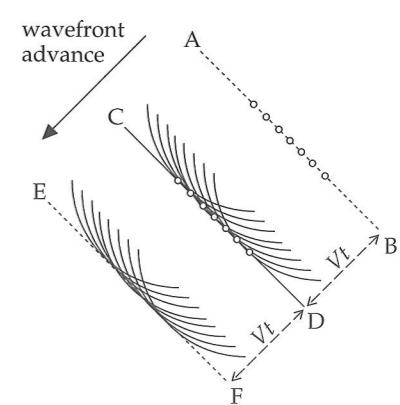
In the Earth's crust the velocities of P- and S-waves are often proportional to each other. This follows from Eqs. (3.39) and (3.47), which give the body-wave velocities in terms of the Lamé constants  $\lambda$  and  $\mu$ . For many rocks, Poisson's relation  $\lambda = \mu$  applies (see Box 3.1), and so

$$\frac{\alpha}{\beta} = \sqrt{\frac{\lambda + 2\mu}{\mu}} = \sqrt{3} \quad (3.98)$$

For brevity, the following discussion handles P-waves only, which are assumed to travel with velocities  $\alpha_1$  and  $\alpha_2$  in the two media. However, we can equally apply the analyses to S-waves, by substituting the appropriate shear-wave velocities  $\beta_1$  and  $\beta_2$  for the media.

#### 3.6.2 Huygens' principle

The passage of a wave through a medium and across interfaces between adjacent media was first explained by the seventeenth century Dutch mathematician and physicist,



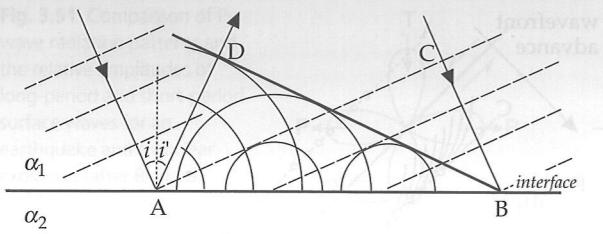
**Fig. 3.52** Application of Huygens' principle to explain the advance of a plane wavefront. The wavefront at CD is the envelope of wavelets set up by particle vibrations when the wavefront was at the previous position AB. Similarly, the envelope of wavelets set up by vibrating particles in the wavefront CD forms the wavefront EF.

Christiaan Huygens, who formulated a principle for the propagation of light as a wave, rather than as the stream of particles visualized by his great and influential contemporary, Sir Isaac Newton. Although derived for the laws of optics, Huygens' principle (1678) can be applied equally to any kind of wave phenomenon. The theory is based on simple geometrical constructions and permits the future position of a wavefront to be calculated if its present position is known. Huygens' principle can be stated: *'All points on a wavefront can be regarded as point sources for the production of new spherical waves; the new wavefront is the tangential surface (or envelope) of the secondary wavelets.'*

This principle can be illustrated simply for a plane wavefront (Fig. 3.52), although the method also applies to curved wavefronts. Let the wavefront initially occupy the position AB and let the open circles represent individual particles of the material in the wavefront. The particles are agitated by the arrival of the wavefront and act as sources of secondary wavelets. If the seismic velocity of the material is  $V$ , the distance travelled by each wavelet after time  $t$  is  $Vt$  and it describes a small sphere around its source particle. If the original wavefront contained numerous closely spaced particles instead of a discrete number, the plane CD tangential to the small wavelets would represent the new position of the wavefront. It is also planar, and lies at a perpendicular distance  $Vt$  from the original wavefront. In their turn the particles in the wavefront CD act as sources for new secondary wavelets, and the process is repeated. This principle can be used to derive the laws of reflection and refraction of seismic waves at an interface, and also to describe the process of diffraction by which a wave is deflected at a corner or at the edge of an object in its path.

##### 3.6.2.1 The law of reflection using Huygens' principle

Consider what happens to a plane P-wave travelling in a medium with seismic velocity  $\alpha_1$  when it encounters the boundary to another medium in which the P-wave velocity is  $\alpha_2$  (Fig. 3.53). At the boundary part of the energy of



**Fig. 3.53** The reflection of a plane P-wave at an interface between two media with different seismic velocities: incident plane waves (e.g., AC); spherical wavelets set up in the upper medium by vibrating particles in the segment AB of the interface; and the reflected plane wave BD, which is the envelope of the wavelets.

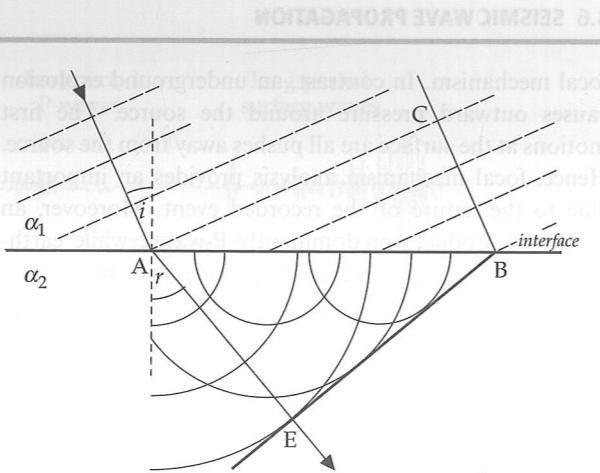
the incident wave is transferred to the second medium, and the remainder is reflected back into the first medium. If the incident wavefront AC first makes contact with the interface at A it agitates particles of the first medium at A and simultaneously the particles of the second medium in contact with the first medium at A. The vibrations of these particles set up secondary waves that travel away from A, back into the first medium as a reflected wave with velocity  $\alpha_1$  (and onward into the second medium as a refracted wave with velocity  $\alpha_2$ ).

By the time the incident wavefront reaches the interface at B all particles of the wavefront between A and B have been agitated. Applying Huygens' principle, the wavefront of the reflected disturbance is the tangent plane to the secondary wavelet in the first medium. In Fig. 3.53 this is represented by the tangent BD from B to the circle centered at A, the first point of contact with the boundary. In the time  $t$  that elapses between the arrival of the plane wave at A and its arrival at B, the incident wavefront travels a distance CB and the secondary wavelet from A travels the equal distance AD. The triangles ABC and ABD are congruent. It follows that the reflected wavefront makes the same angle with the interface as the incident wave.

It is customary to describe the orientation of a plane by the direction of its normal. The angle between the normal to the interface and the normal to the incident wavefront is called the angle of incidence ( $i$ ); the angle between the normal to the interface and the normal to the reflected wavefront is called the angle of reflection ( $i'$ ). This application of Huygens' principle to plane seismic waves shows that the angle of reflection is equal to the angle of incidence ( $i = i'$ ). This is known as the *law of reflection*. Although initially developed for light-beams, it is also valid for the propagation of seismic waves.

### 3.6.2.2 The law of refraction using Huygens' principle

The discussion of the interaction of the incident wave with the boundary can be extended to cover the part of the disturbance that travels into the second medium (Fig. 3.54). This disturbance travels with the velocity  $\alpha_2$  of the second medium. Let  $t$  be the time taken for the incident wavefront in the first medium to advance from C to B; then  $BC = \alpha_1 t$ . In this time all particles of the second



**Fig. 3.54** The refraction of a plane P-wave at an interface between two media with different seismic velocities  $\alpha_1$  and  $\alpha_2$  ( $> \alpha_1$ ): incident plane waves (e.g., AC); spherical wavelets set up in the lower medium by vibrating particles in the segment AB of the interface; and the refracted plane wave BE, which is the envelope of the wavelets. The angles of incidence ( $i$ ) and refraction ( $r$ ) are defined between the normal to the interface and the respective rays.

medium between A and B have been agitated and now act as sources for new wavelets in the second medium. When the incident wave reaches B, the wavelet from A in the second medium has spread out to the point E, where  $AE = \alpha_2 t$ . The wavefront in the second medium is the tangent BE from B to the circle centered at A. The angle of incidence ( $i$ ) is defined as before; the angle between the normal to the interface and the normal to the transmitted wavefront is called the angle of refraction ( $r$ ). Comparison of the triangles ABC and ABE shows that  $BC = AB \sin i$ , and  $AE = AB \sin r$ . Consequently,

$$\frac{AB \sin i}{AB \sin r} = \frac{BC}{AE} = \frac{\alpha_1 t}{\alpha_2 t} \quad (3.99)$$

$$\frac{\sin i}{\sin r} = \frac{\alpha_1}{\alpha_2} \quad (3.100)$$

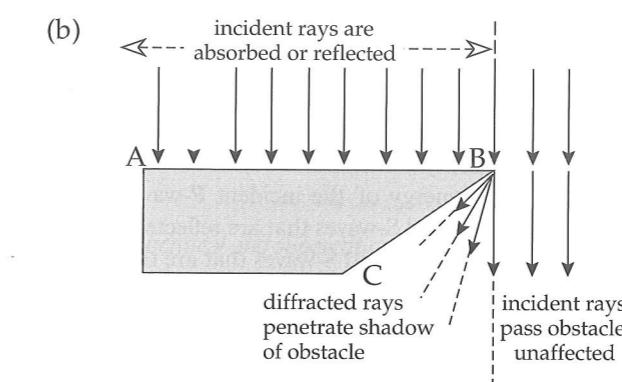
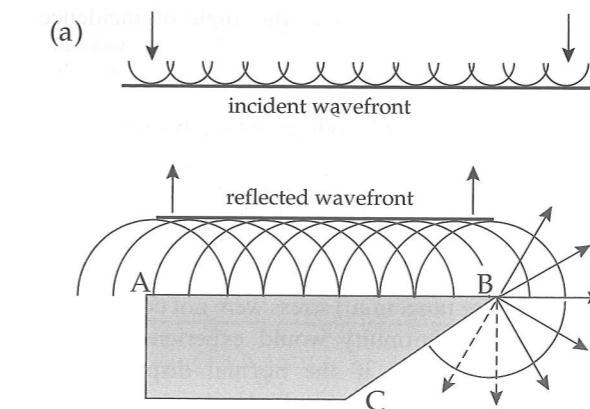
Equation (3.100) is called the *law of refraction* for plane seismic waves. Its equivalent in optics is often called *Snell's law*, in recognition of its discoverer, the Dutch mathematician Willebrod Snellius (or Snell).

### 3.6.2.3 Diffraction

The laws of reflection and refraction derived above with the aid of Huygens' principle apply to the behavior of plane seismic waves at plane boundaries. When a plane or spherical seismic wave encounters a pointed obstacle or discontinuous surface, it experiences *diffraction*. This phenomenon allows the wave to bend around the obstacle, penetrating what otherwise would be a shadow zone for the wave. It is the diffraction of sound waves, for example, that allows us to hear the voices of people who are still invisible to us around a corner, or on the other side of a high fence. Huygens' principle also gives an explanation for diffraction, as illustrated by the following simple case.

Consider the normal incidence of a plane wave on a straight boundary that ends at a sharp corner B

## 3.6 SEISMIC WAVE PROPAGATION

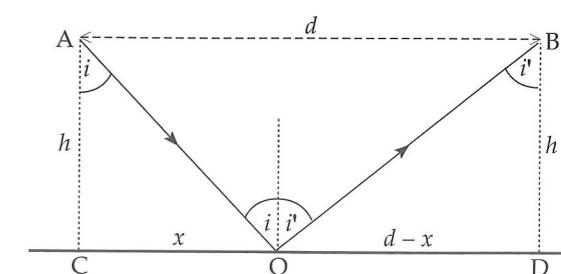


**Fig. 3.55** Explanation of diffraction at an edge with the aid of Huygens' principle. (a) The incident and reflected plane wavefronts are the envelopes to Huygens wavelets, which are able to carry the incident disturbance around a sharp corner. (b) The incident rays are absorbed, reflected or pass by the obstacle, but some rays related to the wavelets generated at the point of the obstruction are diffracted into its shadow.

(Fig. 3.55). The incident wave is reflected along the entire length AB, with each particle of AB acting as a secondary source according to Huygen's principle. Beyond the edge B the incident wavefronts cannot be reflected. The plane wavefront passes by the edge B, so that the point C should lie in the shadow of AB. However, the corner B also acts as a source of secondary wavelets, part of which contribute to the reflected wavefront and part pass into the shadow zone. The intensity of the wave diffracted into the shadow zone is weaker than in the main wavefront, and it decreases progressively with increasing angle away from the direction of travel of the incident wavefront.

### 3.6.3 Fermat's principle

The behavior of seismic ray paths at an interface is explained by another principle of optics that was formulated – also in the seventeenth century – by the French mathematician Pierre de Fermat. As applied to seismology, Fermat's principle states that, of the many possible paths between two points A and B, the seismic ray follows the path that gives the shortest travel-time between



**Fig. 3.56** Geometry of incident and reflected rays for derivation of the law of reflection with the aid of Fermat's principle.

the points. If  $ds$  is the element of distance along a ray path and  $c$  is the seismic velocity over this short distance, then the travel-time  $t$  between A and B is minimum. Thus,

$$t = \int_A^B \frac{ds}{c} = \text{minimum} \quad (3.101)$$

Generally, when the velocity varies continuously with position, the determination of the ray path is intricate. In the case of a layered medium, in which the velocity is constant in each layer, Fermat's principle provides us with an independent method for determining the laws of reflection and refraction.

### 3.6.3.1 The law of reflection using Fermat's principle

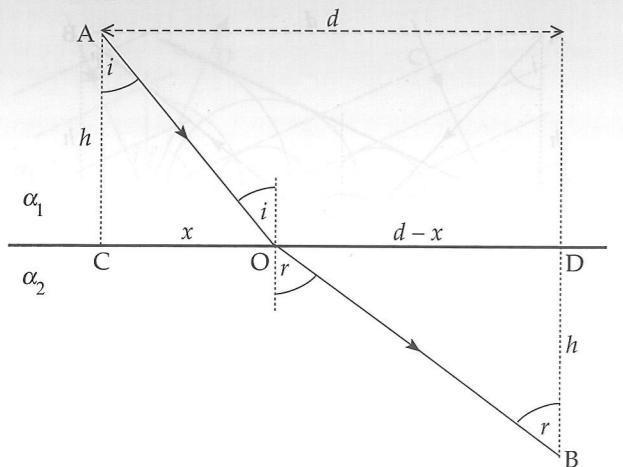
Consider the reflection of a seismic ray in a medium with constant P-wave velocity  $\alpha_1$  at the boundary to another medium (Fig. 3.56). For convenience we take the boundary to be horizontal. Let A be a point on the incident ray at a vertical distance  $h$  from the boundary and let B be the corresponding point on the reflected ray. Let C and D be the nearest points on the boundary to A and B, respectively. Further, let  $d$  be the horizontal separation AB, and let O be the point of reflection on the interface at a horizontal distance  $x$  from C; then OD is equal to  $(d - x)$  and we can write for the travel-time  $t$  from A to B:

$$t = \frac{AO}{\alpha_1} + \frac{OB}{\alpha_1} = \frac{1}{\alpha_1} [\sqrt{h^2 + x^2} + \sqrt{h^2 + (d - x)^2}] \quad (3.102)$$

According to Fermat's principle the travel-time  $t$  must be a minimum. The only variable in Eq. (3.102) is  $x$ . To find the condition that gives the minimum travel-time we differentiate  $t$  with respect to  $x$  and set the result equal to zero:

$$\frac{\partial t}{\partial x} = \frac{1}{\alpha_1} \left[ \frac{x}{\sqrt{h^2 + x^2}} - \frac{(d - x)}{\sqrt{h^2 + (d - x)^2}} \right] = 0 \quad (3.103)$$

By inspection of Fig. 3.56 the relationships of these expressions to the angle of incidence ( $i$ ) and the angle of reflection ( $i'$ ) are evident. The first expression inside the brackets is  $\sin i$  and the second is  $\sin i'$ . The condition for the minimum travel-time is again  $i' = i$ ; the angle of reflection equals the angle of incidence.



**Fig. 3.57** Geometry of incident and refracted rays for derivation of the law of refraction with the aid of Fermat's principle.

### 3.6.3.2 The law of refraction using Fermat's principle

We can use a similar approach to determine the law of refraction. This time we study the passage of the seismic ray from a medium with velocity  $\alpha_1$  into a medium with higher velocity  $\alpha_2$  (Fig. 3.57). Let A again be a point on the incident ray at a vertical distance  $h$  from a point C on the interface. The ray traverses the boundary at O, a horizontal distance  $x$  from C. Let B now be a point on the ray in the second medium at a distance  $h$  from D, the closest point on the interface. The distance CD is  $d$ , so that again OD is equal to  $(d - x)$ . The travel-time  $t$  which we have to minimize is given by

$$t = \frac{AO}{\alpha_1} + \frac{OB}{\alpha_2} = \frac{\sqrt{h^2 + x^2}}{\alpha_1} + \frac{\sqrt{h^2 + (d - x)^2}}{\alpha_2} \quad (3.104)$$

Differentiating Eq. (3.104) with respect to  $x$  and setting the result equal to zero gives us the condition for the minimum value of  $t$ :

$$\frac{x}{\alpha_1 \sqrt{h^2 + x^2}} - \frac{(d - x)}{\alpha_2 \sqrt{h^2 + (d - x)^2}} = 0 \quad (3.105)$$

By reference to Fig. 3.57 we can write this expression in terms of the sines of the angles of incidence ( $i$ ) and refraction ( $r$ ). This application of Fermat's principle to the seismic ray paths gives again the law of refraction that we derived by applying Huygens' principle to the wavefronts (Eq. (3.100)). It can also be stated as

$$\frac{\sin i}{\alpha_1} = \frac{\sin r}{\alpha_2} \quad (3.106)$$

In this example we have assumed that  $\alpha_2 > \alpha_1$ . As it passes from the medium with lower velocity into the medium with higher velocity the refracted ray is bent away from the normal to the boundary, giving an angle of refraction that is greater than the angle of incidence ( $r > i$ ). Under the opposite conditions, if  $\alpha_2 < \alpha_1$ , the refracted ray is bent back toward the normal and the

angle of refraction is less than the angle of incidence ( $r < i$ ).

### 3.6.4 Partitioning of seismic body waves at a boundary

The conditions that must be fulfilled at a boundary are that the normal and tangential components of stress, as well as the normal and tangential components of the displacements, must be continuous across the interface. If the normal (or tangential) stress were not continuous, the point of discontinuity would experience infinite acceleration. Similarly, if the normal displacements were not continuous, a gap would develop between the media or parts of both media would overlap to occupy the same space; discontinuous tangential displacements would result in relative motion between the media across the boundary. These anomalies are impossible if the boundary is a fixed surface that clearly separates the media.

As a result of the conditions of continuity, a P-wave incident on a boundary energizes the particles on each side of the boundary at the point of incidence, and sets up four waves. The energy of the incident P-wave is partitioned between P- and S-waves that are reflected from the boundary, and other P- and S-waves that are transmitted into the adjacent layer. The way in which this takes place may be understood by considering the particle motion that is induced at the interface.

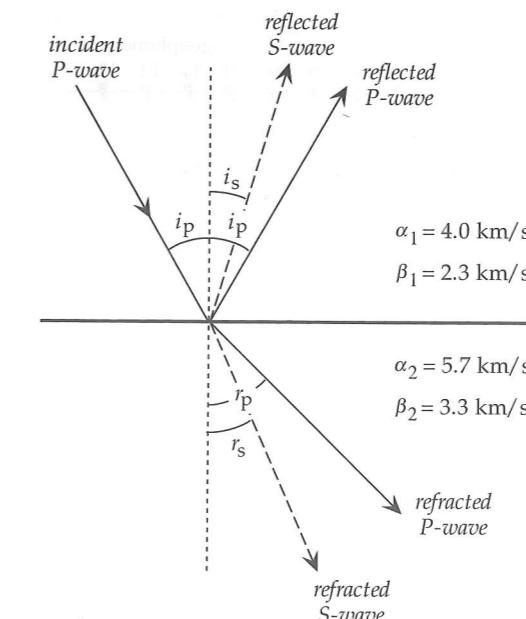
The particle motion in the incident P-wave is parallel to the direction of propagation. At the interface the vibration of particles of the lower layer can be resolved into a component perpendicular to the interface and a component parallel to it in the vertical plane containing the incident P-wave. In the second layer each of these motions can in turn be resolved into a component parallel to the direction of propagation (a refracted P-wave) and a component perpendicular to it in the vertical plane (a refracted SV-wave). Because of continuity at the interface, similar vibrations are induced in the upper layer, corresponding to a reflected P-wave and a reflected SV-wave, respectively.

Let the angles between the normal to the interface and the ray paths of the P- and S-waves in medium 1 be  $i_p$  and  $i_s$ , respectively, and the corresponding angles in medium 2 be  $r_p$  and  $r_s$  (Fig. 3.58). Applying Snell's law to both the reflected and refracted P- and S-waves gives

$$\frac{\sin i_p}{\alpha_1} = \frac{\sin r_p}{\alpha_2} = \frac{\sin i_s}{\beta_1} = \frac{\sin r_s}{\beta_2} \quad (3.107)$$

By similar reasoning it is evident that an incident SV-wave also generates vibrations that have components normal and parallel to the interface, and will set up refracted and reflected P- and SV-waves. The situation is different for an incident SH-wave, which has no component of motion normal to the interface. In this case only refracted and reflected SH-waves are created.

## 3.6 SEISMIC WAVE PROPAGATION



**Fig. 3.58** The generation of reflected and refracted P- and S-waves from a P-wave incident on a plane interface.

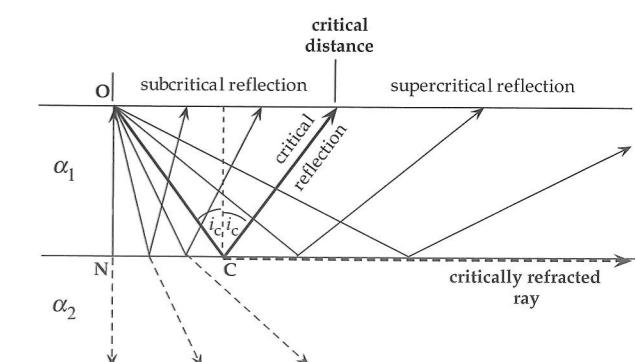
### 3.6.4.1 Subcritical and supercritical reflections, and critical refraction

Let O be a seismic source near the surface of a uniformly thick horizontal layer with P-wave velocity  $\alpha_1$  that lies on top of a layer with higher velocity  $\alpha_2$  (Fig. 3.59). Consider what happens to seismic rays that leave O and arrive at the boundary with all possible angles of incidence. The most simple ray is that which travels vertically to meet the boundary with zero angle of incidence at the point N. This *normally incident ray* is partially reflected back along its track, and partially transmitted vertically into the next medium without change of direction. As the angle of incidence increases, the point of incidence moves from N towards C. The transmitted ray experiences a change of direction according to Snell's law of refraction, and the ray reflected to the surface is termed a *subcritical reflection*.

The ray that is incident on the boundary at C is called the *critical ray* because it experiences *critical refraction*. It encounters the boundary with a *critical angle of incidence*. The corresponding refracted ray makes an angle of refraction of  $90^\circ$  with the normal to the boundary. As a result, it travels parallel to the boundary in the top of the lower layer with faster velocity  $\alpha_2$ . The sine of the angle of refraction of the critical ray is unity, and we can calculate the critical angle,  $i_c$ , by applying Snell's law:

$$\sin i_c = \frac{\alpha_1}{\alpha_2} \quad (3.108)$$

The critical ray is accompanied by a *critical reflection*. It reaches the surface at a *critical distance* ( $x_c$ ) from the source at O. The reflections that arrive inside the critical

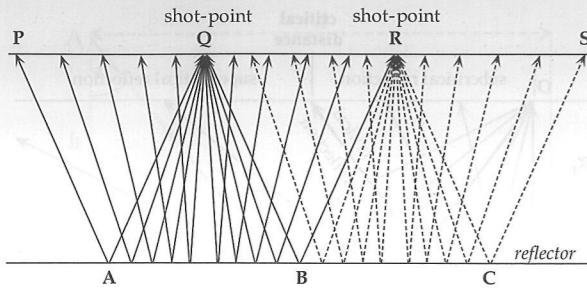


**Fig. 3.59** The critical reflection defines two domains, corresponding to regions of subcritical and supercritical reflection.

distance are called *subcritical reflections*. At angles up to the critical angle, refracted rays pass into the lower medium, but for rays incident at angles greater than the critical angle, refraction is no longer possible. The seismic rays that are incident more obliquely than the critical angle are reflected almost completely. These reflections are termed *supercritical reflections*, or simply *wide-angle reflections*. They lose little energy to refraction, and are thus capable of travelling large distances from the source in the upper medium. Supercritical reflections are recorded with strong amplitudes on seismograms at distant stations.

### 3.6.5 Reflection seismology

Reflection seismology is directed primarily at finding the depths to reflecting surfaces and the seismic velocities of subsurface rock layers. The techniques of acquiring and processing reflection seismology data have been developed and refined to a very high degree of sophistication as a result of the intensive application of this method in the search for petroleum. The principle is simple. A seismic signal (e.g., an explosion) is produced at a known place at a known time, and the echoes reflected from the boundaries between rock layers with different seismic velocities and densities are recorded and analyzed. Compactly designed, robust, electromagnetic seismometers – called “geophones” in industrial usage – are spread in the region of subcritical reflection, within the critical distance from the shot-point, where no refracted arrivals are possible. Within this distance the only signals received are the wave that travels directly from the shot-point to the geophones and the waves reflected at subsurface interfaces. Surface waves are also recorded and constitute an important disturbing “noise,” because they interfere with the reflected signal. The closer the geophone array is located to the shot-point, the more nearly the paths of the reflected rays travel vertically. Reflection seismic data are most usually acquired along profiles that cross geological structures as nearly as possible normal to the strike of the structure. The travel-times recorded at the geophones along a profile are plotted as a two-dimensional cross-section of the structure. In recent years, three-dimensional surveying,



**Fig. 3.60** The split-spread method of obtaining continuous subsurface coverage of a seismic reflector.

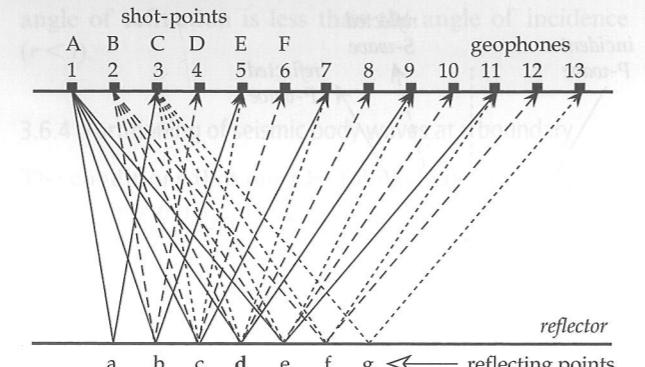
which covers the entire subsurface, has become more important.

Several field procedures are in common use. They are distinguished by different layouts of the geophones relative to the shot-point. The most routine application of reflection seismology is in *continuous profiling*, in which the geophones are laid out at discrete distances along a profile through the shot-point. To reduce seismic noise, each recording point is represented by a group of interconnected geophones. After each shot the geophone layout and shot-point are moved a predetermined distance along the profile, and the procedure is repeated. Broadly speaking, there are two main variations of this method, depending on whether each reflection point on the reflector is sampled only once (*conventional coverage*) or more than once (*redundant coverage*).

The most common form of conventional coverage is a *split-spread* method (Fig. 3.60), in which the geophones are spread symmetrically on either side of the shot-point. If the reflector is flat-lying, the point of reflection of a ray recorded at any geophone is below the point midway between the shot-point and the geophone. For a shot-point at Q the rays QAP and QBR that are reflected to geophones at P and R represent extreme cases. The two-way travel-time of the ray QAP gives the depth of the reflection point A, which is plotted below the mid-point of QP. Similarly, B is plotted below the mid-point of QR. The split-spread layout around the shot-point Q gives the depths of reflection points along AB, which is half the length of the geophone spread PR. The shot-point is now moved to the point R, and the geophones between P and Q are moved to cover the segment RS. From the new shot-point R the positions of reflection points in the segment BC of the reflector are obtained. The ray RBQ from shot-point R to the geophone at Q has the same path as the ray QBR from shot-point Q to the geophone at R. By successively moving the shot-point and half of the split-spread geophone layout a continuous coverage of the subsurface reflector is obtained.

### 3.6.5.1 Reflection at a horizontal interface

The simplest case of seismic reflection is the two-dimensional reflection at a horizontal boundary (Fig. 3.62). Let the reflecting bed be at depth  $d$  below the shot-point S. The ray that strikes the boundary at R is reflected to the surface and recorded by a geophone at the point G, so that the angles of incidence and reflection are equal. Let G be at a horizontal distance  $x$  from the shot-point. If the P-wave velocity is  $V$ , the first signal received at G is from the direct wave that travels directly along SG. Its travel-time is given by  $t_d = x/V$ . It is important to keep in mind that the direct wave is not a surface wave but a body wave that travels parallel to and just below the surface of the top layer. The travel-time  $t$  of the reflected ray SRG is  $(SR+RG)/V$ . However, SR and RG are equal and therefore



**Fig. 3.61** Common-mid-point method of seismic reflection shooting, showing rays from successive shot-points at A, B and C and the repeated sampling of the same point on the reflector (e.g., d) by rays from each shot-point.

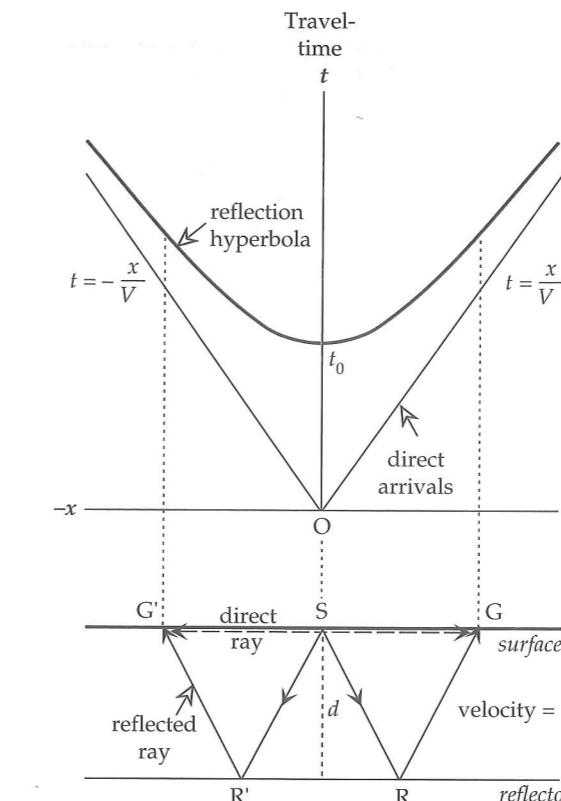
is fired at A, the signals received at geophones 3–11 give subsurface coverage of the reflector between points a and e. The shot-point is now moved to B, which coincides with the position occupied by geophone 2 for the first shot, and the geophone array is moved forward correspondingly along the direction of the profile to positions 4–12. From shot-point B the subsurface coverage of the reflector is between points b and f. The reflector points b to e are common to both sets of data. By repeatedly moving the shot-point and geophone array in the described manner, each reflecting point of the interface is sampled multiply. For example, in Fig. 3.61 the reflecting point d is sampled multiply by the rays Ad9, Bd8, Cd7, etc. The lengths of these ray paths are different. During subsequent data-processing the reflection travel-times are corrected for *normal moveout*, which is a geometrical effect related to geophone distance from the shot-point. The records are then *stacked*, which is a procedure for enhancing the signal-to-noise ratio.

**3.6.5.1 Reflection at a horizontal interface**

The simplest case of seismic reflection is the two-dimensional reflection at a horizontal boundary (Fig. 3.62). Let the reflecting bed be at depth  $d$  below the shot-point S. The ray that strikes the boundary at R is reflected to the surface and recorded by a geophone at the point G, so that the angles of incidence and reflection are equal. Let G be at a horizontal distance  $x$  from the shot-point. If the P-wave velocity is  $V$ , the first signal received at G is from the direct wave that travels directly along SG. Its travel-time is given by  $t_d = x/V$ . It is important to keep in mind that the direct wave is not a surface wave but a body wave that travels parallel to and just below the surface of the top layer. The travel-time  $t$  of the reflected ray SRG is  $(SR+RG)/V$ . However, SR and RG are equal and therefore

$$t = \frac{2}{V} \sqrt{d^2 + \frac{x^2}{4}} \quad (3.109)$$

## 3.6 SEISMIC WAVE PROPAGATION



**Fig. 3.62** The travel-time versus distance curve for reflections from a horizontal boundary is a hyperbola. The vertical reflection time  $t_0$  is the intercept of the hyperbola with the travel-time axis.

$$t = \frac{2d}{V} \sqrt{1 + \frac{x^2}{4d^2}} = t_0 \sqrt{1 + \frac{x^2}{4d^2}} \quad (3.110)$$

At  $x=0$  the travel-time corresponds to the vertical echo from the reflector; this ‘echo-time’ is given by  $t_0 = 2d/V$ . The quantity under the square root in Eq. (3.110) determines the curvature of the  $t-x$  curve and is called the *normal moveout* factor. It arises because the ray reaching a geophone at a horizontal distance  $x$  from the shot-point has not travelled vertically between it and the reflector. Squaring both sides of Eq. (3.110) and rearranging terms gives

$$\frac{t^2}{t_0^2} - \frac{x^2}{4d^2} = 1 \quad (3.111)$$

This is the equation of a hyperbola (Fig. 3.62) that is symmetrical about the vertical time axis, which it intersects at  $t_0$ . For large distances from the shot-point ( $x \gg d$ ) the travel-time of the reflected ray approaches the travel-time of the direct ray and the hyperbola is asymptotic to the two lines  $t = \pm x/V$ .

A principle goal of seismic reflection profiling is usually to find the vertical distance ( $d$ ) to a reflecting interface. This can be determined from  $t_0$ , the two-way reflection travel-time recorded by a geophone at the shot-point, once the velocity  $V$  is known. One way of determining the velocity is by comparing  $t_0$  with the travel-time  $t_x$  to a geophone at distance  $x$ . In reflection seismology the geophones are

laid out close to the shot-point and the assumption is made that the geophone distance is much less than the depth of the reflector ( $x \ll d$ ). Equation (3.110) becomes

$$\begin{aligned} t_x &= t_0 \left( 1 + \left( \frac{x}{2d} \right)^2 \right)^{1/2} = t_0 \left( 1 + \frac{1}{2} \left( \frac{x}{2d} \right)^2 + \dots \right) \\ &= t_0 \left( 1 + \frac{1}{2} \left( \frac{x}{Vt_0} \right)^2 \right) \end{aligned} \quad (3.112)$$

The difference between the travel-time  $t_x$  and the shot-point travel-time  $t_0$  is the *normal moveout*,  $\Delta t_n = t_x - t_0$ . By rearranging Eq. (3.112) we get

$$\Delta t_n = \frac{x^2}{2V^2 t_0} \quad (3.113)$$

The echo time  $t_0$  and the normal moveout time  $\Delta t_n$  are found from the reflection data. The distance  $x$  of the geophone from the shot-point is known and therefore the layer velocity  $V$  can be determined. The depth  $d$  of the reflecting horizon can then be found by using the formula for the echo time.

An alternative way of interpreting reflection arrival times becomes evident when Eq. (3.111) is rearranged in the form

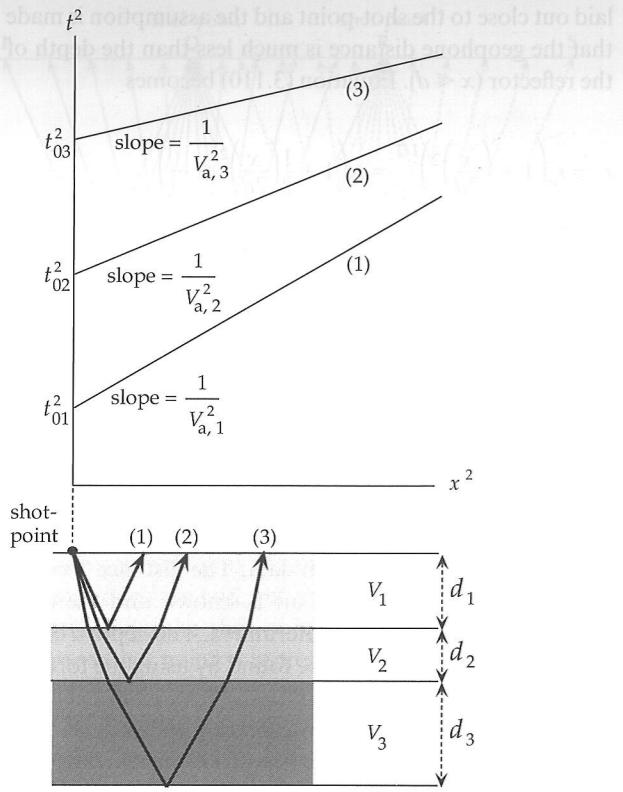
$$t^2 = t_0^2 + \frac{x^2}{V^2} \quad (3.114)$$

A plot of  $t^2$  against  $x^2$  is a straight line that has slope  $1/V^2$ . Its intercept with the  $t^2$ -axis gives the square of the echo time,  $t_0^2$ , from which the depth  $d$  to the reflector can be found once the velocity  $V$  is known. The record at each geophone will contain reflections from several reflectors. For the first reflector the velocity determined by the  $t^2-x^2$  method is the true *interval velocity* of the uppermost layer,  $V_1$ , which, in conjunction with  $t_{01}$ , the first echo time, gives the thickness  $d_1$  of the top layer. However, the ray reflected from the second interface has travelled through the first layer with interval velocity  $V_1$  and the second layer with interval velocity  $V_2$ . The velocity interpreted in the  $t^2-x^2$  method for this reflection, and for reflections from all deeper interfaces, is an *average velocity*. If the incident and reflected rays travel nearly vertically, the average velocity  $V_{a,n}$  for the reflection from the  $n$ th reflector is given by

$$V_{a,n} = \frac{d_1 + d_2 + d_3 + \dots + d_n}{t_1 + t_2 + t_3 + \dots + t_n} = \frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n t_i} \quad (3.115)$$

where  $d_i$  is the thickness and  $t_i$  the interval travel-time for the  $i$ th layer.

The  $t^2-x^2$  method is a simple way of estimating layer thicknesses and average velocities for a multi-layered Earth (Fig. 3.63). The slope of the second straight line gives  $V_{a,2}$ , which is used with the appropriate echo time  $t_{02}$  to find the combined depth  $D_2$  to the second interface, given by  $D_2 = d_1 + d_2 = (V_{a,2})t_{02}$ ;  $d_1$  is known, and so  $d_2$



**Fig. 3.63** Illustration of a “ $t^2-x^2$  plot” for near-vertical reflections from three horizontal reflectors;  $V_{a,1}$  is the true velocity  $V_1$  of layer 1, but  $V_{a,2}$  and  $V_{a,3}$  are “average” velocities that depend on the true velocities and the layer thicknesses.

can be calculated. The two-way travel-time in the second layer is  $(t_{02} - t_{01})$  and thus the interval velocity  $V_2$  can be found. In this way the thicknesses and interval velocities of deeper layers can be determined successively.

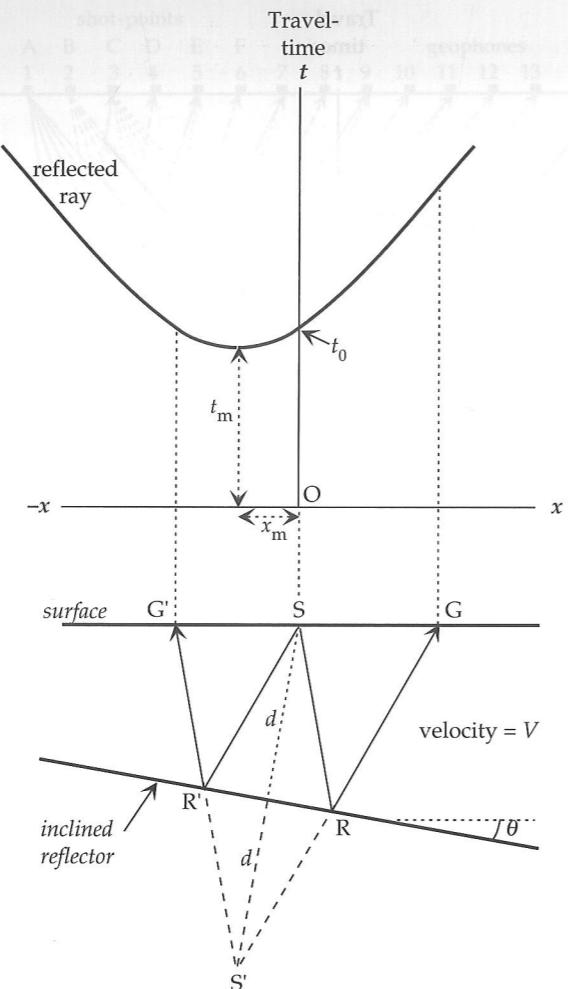
In fact, of course, the rays do not travel vertically but are bent as they pass from one layer to another (Fig. 3.63). Moreover, the elastic properties of a layer are rarely homogeneous so that the seismic velocity is variable and the ray path in the layer is curved. Exploration seismologists have found it possible to compensate for these effects by replacing the average velocity with the root-mean-square velocity  $V_{rms}$  defined by

$$V_{rms}^2 = \frac{\sum_i^n V_i^2 t_i}{\sum_i^n t_i} \quad (3.116)$$

where  $V_i$  is the interval velocity and  $t_i$  the travel-time for the  $i$ th layer.

#### 3.6.5.2 Reflection at an inclined interface

When the reflecting interface is inclined at an angle  $\theta$  to the horizontal, as in Fig. 3.64, the shortest distance  $d$  between the shot-points and the reflector is the perpendicular distance to the inclined plane. The paths of reflected rays on the down-dip side of the shot-point are



**Fig. 3.64** The travel-time versus distance curve for an inclined reflector is also a hyperbola with vertical axis (cf. Fig. 3.62), but the minimum travel-time ( $t_m$ ) is measured at distance  $x_m$  from the shot-point.

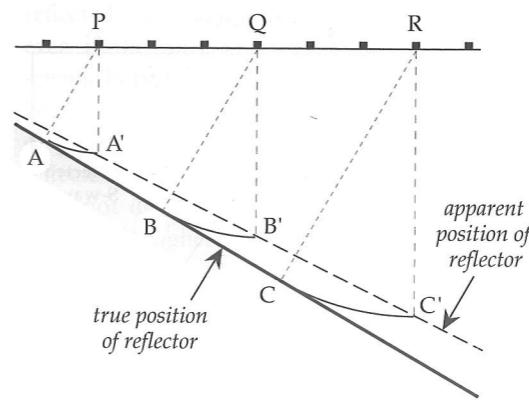
longer than those on the up-dip side; this has corresponding effects on the travel-times. The rays obey the laws of reflection optics and appear to return to the surface from the point  $S'$ , which is the image point of the shot-point with respect to the reflector. The travel-time  $t$  through the layer with velocity  $V$  is readily found with the aid of the image point. For example for the ray SRG, recorded by a geophone on the surface at G, we get

$$t = \frac{SR + RG}{V} = \frac{S'R + RG}{V} = \frac{S'G}{V} \quad (3.117)$$

The image point  $S'$  is as far behind the reflector as the shot-point is in front:  $S'S = 2d$ . In triangle  $S'SG$  the side  $SG$  equals the geophone distance  $x$  and the obtuse angle  $S'SG$  equals  $(90^\circ + \theta)$ . If we apply the law of cosines to the triangle  $S'SG$  we can solve for  $S'G$  and substitute the answer in Eq. (3.117). This gives the travel-time of the reflection from the inclined boundary:

$$t = \frac{1}{V} \sqrt{(x^2 + 4xd \sin \theta + 4d^2)} \quad (3.118)$$

## 3.6 SEISMIC WAVE PROPAGATION



**Fig. 3.65** When the reflector is inclined and depths are plotted vertically under geophone positions, the true reflecting points A, B and C are mapped at A', B' and C', falsifying the position of the reflector.

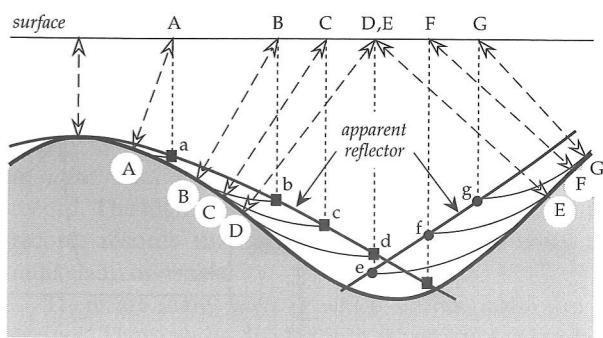
Equation (3.118) is the equation of a hyperbola whose axis of symmetry is vertical, parallel to the  $t$ -axis. For a flat reflector the hyperbola was symmetric about the  $t$ -axis (Fig. 3.62) and the minimum travel-time (echo time) corresponded to the vertical reflection below the shot-point ( $x = 0$ ). For an inclined reflector the minimum travel-time  $t_m$  is no longer the perpendicular path to the reflector, which would give the travel-time  $t_0$  in Fig. 3.64. Although the perpendicular path is the shortest distance from shot-point to reflector, it is not the shortest path of a reflected ray between the shot-point and a geophone. The shortest travel-time is recorded by the geophone at a horizontal distance  $x_m$  on the up-dip side of the shot-point. The coordinates  $(x_m, t_m)$  of the minimum point of the travel-time hyperbola are

$$x_m = -2d \sin \theta \quad (3.119)$$

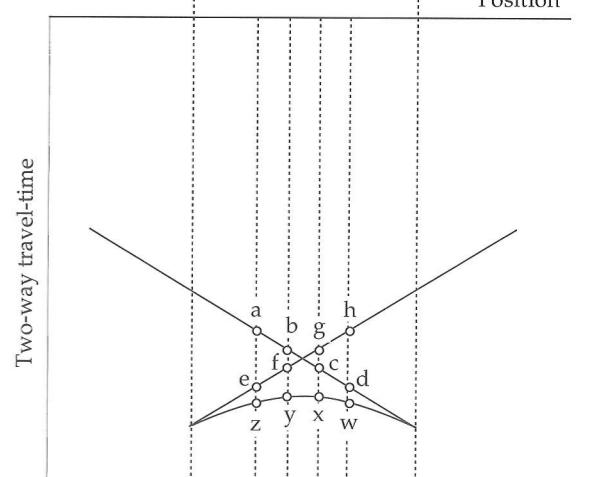
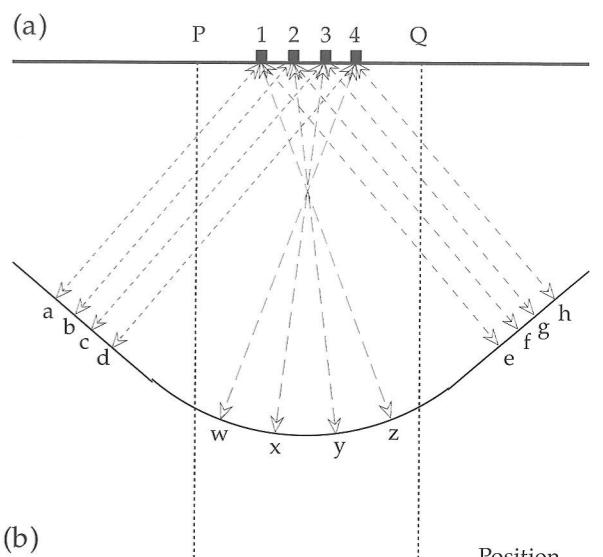
$$t_m = \frac{2d \cos \theta}{V}$$

In practice it is not known a priori whether a reflector is horizontal or inclined. If reflection records are not corrected for the effect of layer-dip, an error results in plotting the positions of dipping beds. The shot-point travel-time  $t_0$  gives the direct distance to a reflector, but the path along which the echo has travelled is not known. Consider the geometry of the inclined boundary in Fig. 3.65. First arrival reflections recorded for shot-points P, Q, and R come from the true reflection points A, B, and C. If the computed reflector depths are plotted directly below the shot-points at A', B' and C', the dipping boundary will appear to lie at a shallower depth than its true position, and the apparent dip of the reflector will be less steep than the true dip. This leads to a distorted picture of the underground structure. For example, an anticline appears broader and less steep-sided than it is. Similarly, if the limbs of a syncline dip steeply enough, the first arrivals from the dipping limbs can conceal the true structure (Fig. 3.66).

This happens when the radius of curvature of the bottom of the syncline is less than the subsurface depth of



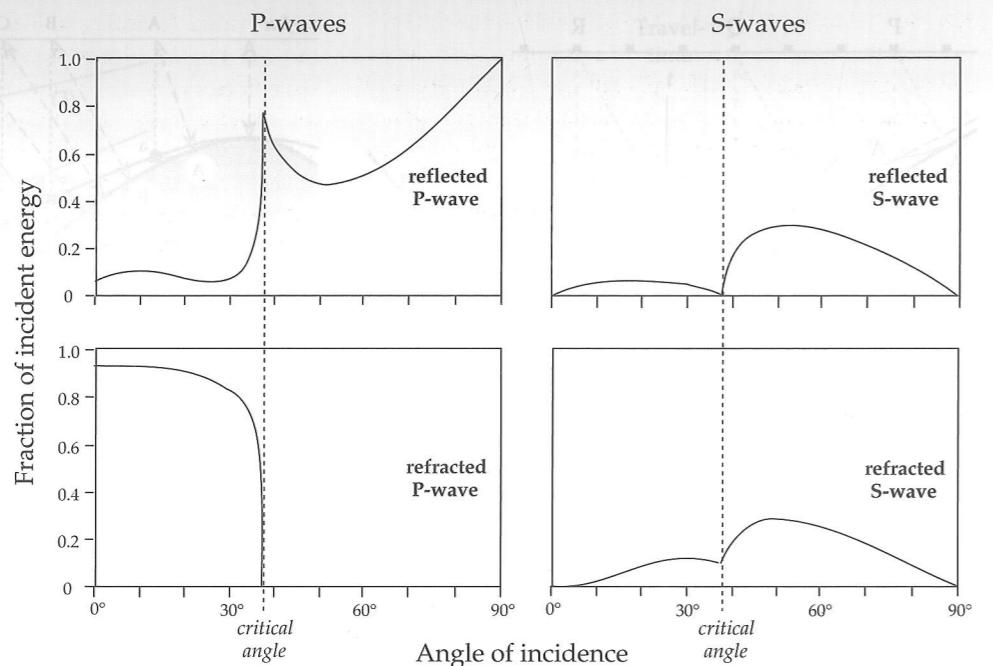
**Fig. 3.66** Paths of reflected rays over an anticline and syncline, showing the false apparent depth to the reflecting surface. True reflection points A-G are wrongly mapped at locations a-g beneath the corresponding shot-points.



**Fig. 3.67** (a) Paths of rays reflected from both flanks and the trough of a tightly curved syncline. (b) Appearance of the corresponding reflection record; the letters on the cusped feature refer to the reflection points in (a).

its axis. Over the axis of the syncline, rays reflected from the dipping flanks may be the first to reach the shot-point geophone. The bottom of the syncline is seen as an upwardly convex reflection between two cusps (Fig. 3.67).

**Fig. 3.68** Partitioning of the energy of an incident P-wave between refracted and reflected P- and S-waves (after Dobrin, 1976 and Richards, 1961).



On an uncorrected reflection record, the appearance of a tight syncline resembles a diffraction.

Reflection seismic records must be corrected for non-vertical reflections. The correctional process is called *migration*. It is an essential part of a reflection seismic study. When the reflection events on seismic cross-sections are plotted vertically below control points on the surface (e.g., as the two-way vertical travel-time to a reflector below the shot-point), the section is said to be *unmigrated*. As discussed above, an unmigrated section misrepresents the depth and dip of inclined reflectors. A *migrated* section is one which has been corrected for non-vertical reflections. It gives a truer picture of the positions of subsurface reflectors.

The process of migration is complex, and requires prior knowledge of the seismic velocity distribution, which in an unexplored or tectonically complicated region is often inadequately known. Several techniques, mostly computer-based, can be used but to treat them adequately is beyond the scope of this text.

### 3.6.5.3 Reflection and transmission coefficients

The partitioning of energy between refractions and reflections at different angles of incidence on a boundary is rather complex. For example, an incident P-wave may be partially reflected and partially refracted, or it may be totally reflected, depending how steeply the incident ray encounters the boundary. The fraction of the incident P-wave energy that is partitioned between reflected and refracted P- and S-waves depends strongly on the angle of incidence (Fig. 3.68). In the case of oblique incidence at less than the critical angle, the amplitudes of the different waves are given by complicated functions of the wave velocities and the angles of incidence, reflection and

refraction. The relative amounts of energy in the refracted and reflected P- and S-waves do not change much for angles of incidence up to about 15°. Beyond the critical angle, the refracted P-wave ceases, so that the incident energy is partly reflected as a P-wave and partly converted to refracted and reflected S-waves.

In practice, reflection seismology is carried out at comparatively small angles of incidence. At *normal incidence* on an interface a P-wave excites no tangential stresses or displacements, and no shear waves are induced. The partitioning of energy between the reflected and refracted P-waves then becomes much simpler. It depends on a property of each medium known as its *acoustic impedance*,  $Z$ , which is defined as the product of the density  $\rho$  of the medium and its P-wave velocity  $\alpha$ ; thus  $Z = \rho\alpha$ . The solution of the equations for the amplitudes  $A_1$  and  $A_2$  of the reflected and refracted P-waves, respectively, in terms of the amplitude  $A_0$  of the incident wave are given by:

$$\begin{aligned} RC &= \frac{A_1}{A_0} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\rho_2\alpha_2 - \rho_1\alpha_1}{\rho_2\alpha_2 + \rho_1\alpha_1} \\ TC &= \frac{A_2}{A_0} = \frac{2Z_1}{Z_2 + Z_1} = \frac{2\rho_1\alpha_1}{\rho_2\alpha_2 + \rho_1\alpha_1} \end{aligned} \quad (3.120)$$

The amplitude ratios RC and TC are called the reflection coefficient and the transmission coefficient, respectively. As shown earlier (see Eq. (3.70)), the energy of a wave is proportional to the square of its amplitude. The fraction  $E_r$  of the incident energy that is reflected is given by the square of RC, the fraction  $E_t$  that is transmitted is equal to  $(1 - E_r)$ .

When the incident wave is reflected at the surface of a medium with higher seismic impedance ( $Z_2 > Z_1$ ), the reflection coefficient RC is positive. This means that the

### 3.6 SEISMIC WAVE PROPAGATION

reflected wave is in phase with the incident wave. However, if the wave is incident on a medium with lower seismic impedance ( $Z_2 < Z_1$ ), the reflection coefficient will be negative. This implies that the reflected wave is 180° out of phase with the incident wave. The fraction of energy reflected from an interface is equal to  $RC^2$ , and therefore does not depend on whether the incidence is from the medium of higher or lower seismic impedance.

#### 3.6.5.4 Synthetic seismograms

The travel-time of a reflection from a deep boundary in a multi-layered Earth is determined by the thicknesses and seismic velocities of the layers the seismic ray traverses. The amplitude of the recorded reflection is determined by the transmission and reflection coefficients for the subsurface interfaces. If the densities and seismic velocities of subsurface layers are known (for example, from sonic and density logs in conveniently located boreholes), it is possible to reconstruct what the seismogram should look like. Sometimes the density variations are ignored, and reflection and transmission coefficients are calculated simply on the basis of seismic velocities. This approximation can often be very useful, for example in the exploration of the deep structure of the Earth's crust with the seismic reflection method. The densities of deep layers are inaccessible directly, although they can be inferred from seismic velocities. A vertical model of seismic velocities may be available from a related refraction study. These data can be incorporated in a deep seismic reflection study to calculate a *synthetic seismogram*. The comparison of real and synthetic seismograms is useful for correlating reflection events and for separating real reflections from noise signals such as multiples.

The principle is simple, but the construction is laborious. A vertically incident wave on the first boundary is resolved into reflected and transmitted components, with amplitudes corresponding to the seismic impedances above and below the interface. The transmitted wave is further subdivided at the next deeper interface into other reflected and transmitted components, and this is repeated at each subsequent boundary. Each wave is followed as it is reflected and refracted at subsurface interfaces until it eventually returns to the surface. The theoretical record consists of the superposition of the numerous events, and represents the total travel-time and amplitude of each event. Whether in a high-resolution reflection study of near-surface sedimentary layering for petroleum exploration or in an analysis of deep crustal structure, the construction of a synthetic seismogram is an exacting chore that requires the use of fast modern computers.

#### 3.6.5.5 Seismic noise

Controlled-source seismology allows fine resolution of a layered underground through analysis of the seismic

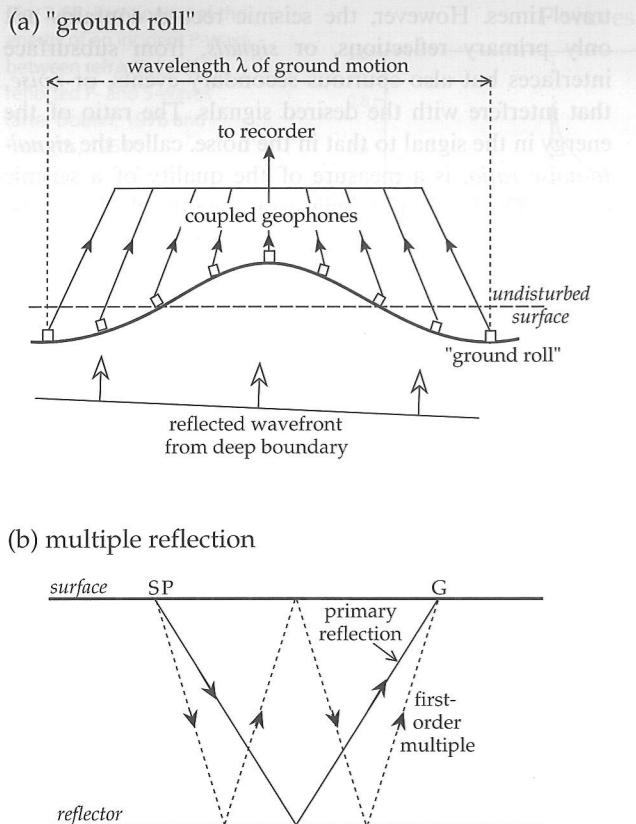
travel-times. However, the seismic record contains not only primary reflections, or *signals*, from subsurface interfaces but also spurious secondary events, or *noise*, that interfere with the desired signals. The ratio of the energy in the signal to that in the noise, called the *signal-to-noise ratio*, is a measure of the quality of a seismic record. The higher the signal-to-noise ratio the better the record; records that have a ratio less than unity are unlikely to be usable.

There are many ways in which seismic noise can be excited. They can be divided into *incoherent* (or *random*) *noise* and *coherent noise*. Incoherent noise is local in origin and is caused by shallow inhomogeneities such as boulders, roots or other non-uniformities that can scatter seismic waves locally. As a result of its local nature incoherent noise is different on the records from adjacent geophones unless they are very close. In reflection seismology it is reduced by arranging the geophones in groups or *arrays*, of typically 16 geophones per group, and combining the individual outputs to produce a single record. When  $n$  geophones form a group, this practice enhances the signal-to-noise ratio by the factor  $\sqrt{n}$ .

Coherent noise is present on several adjacent traces of a seismogram. Two common forms result from surface waves and multiple reflections, respectively.

A near-surface explosion excites surface waves (especially Rayleigh waves) that can have strong amplitudes. They travel more slowly than P-waves but often reach the geophones together with the train of subsurface reflections. The resultant "ground roll" can obscure the reflections, particularly if they are weak. The problem can be minimized by the geometry of the geophone layout (Fig. 3.69a). For example, if the 16 geophones in a group are laid out at equal distances to cover a complete wavelength of the Rayleigh wave, the signals of individual geophones are effectively integrated to a low value. This procedure is only partially effective. Although most of the ground roll is related to Rayleigh waves, part is thought to have more complex near-surface origins. Another method of reducing the effects of ground roll is frequency filtering. The frequency of the ground roll is often lower than that of the reflected P-waves, which allows attenuation of this type of coherent noise by including a high-pass filter in the geophone circuit or in the subsequent processing to cut out low frequencies.

Multiple reflections are a very common source of coherent noise in a layered medium. They can originate in several ways, the most serious of which is the surface multiple (Fig. 3.69b). The reflection coefficient at the free surface of the Earth is high (in principle,  $RC \approx -1$ ), and multiple reflections can occur between the surface and a reflecting interface. The travel-time for a first-order multiple at near-vertical incidence is double that of the primary signal. A copy of the reflector is observed on the seismogram at twice the real depth (or travel-time). Higher-order multiples produce additional apparent reflectors. A further advantage of using the common-mid-point method of



**Fig. 3.69** Examples of seismic noise: (a) "ground roll" due to surface wave, and (b) multiple reflections between a reflector and the free surface.

reflection profiling is that it is effective in attenuating the surface multiples.

### 3.6.5.6 Reflection seismic section

After migration, the cleaned records from all the seismometers are plotted side-by-side to form a sort of cross-section of the underground structure beneath the reflection profile. Strong reflectors can be followed across the section, and where they are interrupted, faults can be deduced. The top part of Fig. 3.70 shows the results of a 113 km long, nearly north-south crustal reflection profile across western Lake Superior at the northern end of the North American Mid-Continent Rift System. This is an aborted Precambrian (age  $\sim 1000$  Ma) rift that has prominent gravity and magnetic expressions. The profile (Line C) was carried out with closely spaced shot-points using the common-mid-point method. The records were subsequently stacked and migrated. The lower part of the figure shows the interpreted subsurface structures above a rather steeply dipping Moho, which under this profile increases in depth from about 32 km in the north to about 50 km in the south. The section shows the presence of some inferred major faults, such as the southward dipping Douglas fault and the northward dipping Keweenawan fault, which appears to truncate other steeply dipping faults.

### 3.6.6 Refraction seismology

The method of seismic refraction can be understood by applying Huygens' principle to the critical refraction at the interface between two layers. The seismic disturbance travels immediately below the interface with the higher velocity of the lower medium. It is called a *head wave* (or *Mintrop wave*, after the German seismologist who patented its use in seismic exploration in 1919). The upper and lower media are in contact at the interface and so the upper medium is forced to move in phase with the lower medium. The vibration excited at the boundary by the passage of the head wave acts as a moving source of secondary waves in the upper layer. The secondary waves interfere constructively (in the same way as a reflected wave is formed) to build plane wavefronts; the ray paths return to the surface at the critical angle within the region of supercritical reflections (Fig. 3.71). The doubly refracted waves are especially important for the information they reveal about the layered structure of the deep interior of the Earth.

#### 3.6.6.1 Refraction at a horizontal interface

The method of refraction seismology is illustrated for the case of the flat interface between two horizontal layers in Fig. 3.71. Let the depth to the interface be  $d$  and the seismic velocities of the upper and lower layers be  $V_1$  and  $V_2$  respectively ( $V_1 < V_2$ ). The *direct ray* from the shot-point at S is recorded by a geophone G at distance  $x$  on the surface after time  $x/V_1$ . The travel-time curve for the direct ray is a straight line through the origin with slope  $m_1 = 1/V_1$ . The hyperbolic  $t-x$  curve for the *reflected ray* intersects the time axis at the two-way vertical reflection ("echo") time  $t_0$ . At great distances from the shot-point the reflection hyperbola is asymptotic to the straight line for the direct ray.

The *doubly refracted ray* travels along the path SC with the velocity  $V_1$  of the upper layer, impinges with critical angle  $i_c$  on the interface at C, passes along the segment CD with velocity  $V_2$  of the lower layer, and returns to the surface along DG with velocity  $V_1$ . The segments SC and DG are equal,  $CD = x - 2SA$  and the travel-time for the path SCDG can be written

$$t = \frac{2SC}{V_1} + \frac{CD}{V_2} \quad (3.121)$$

i.e.,

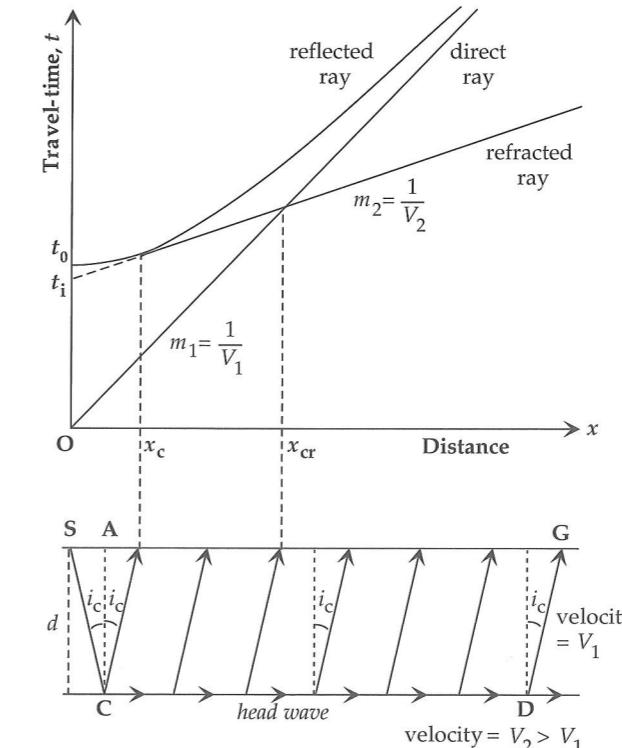
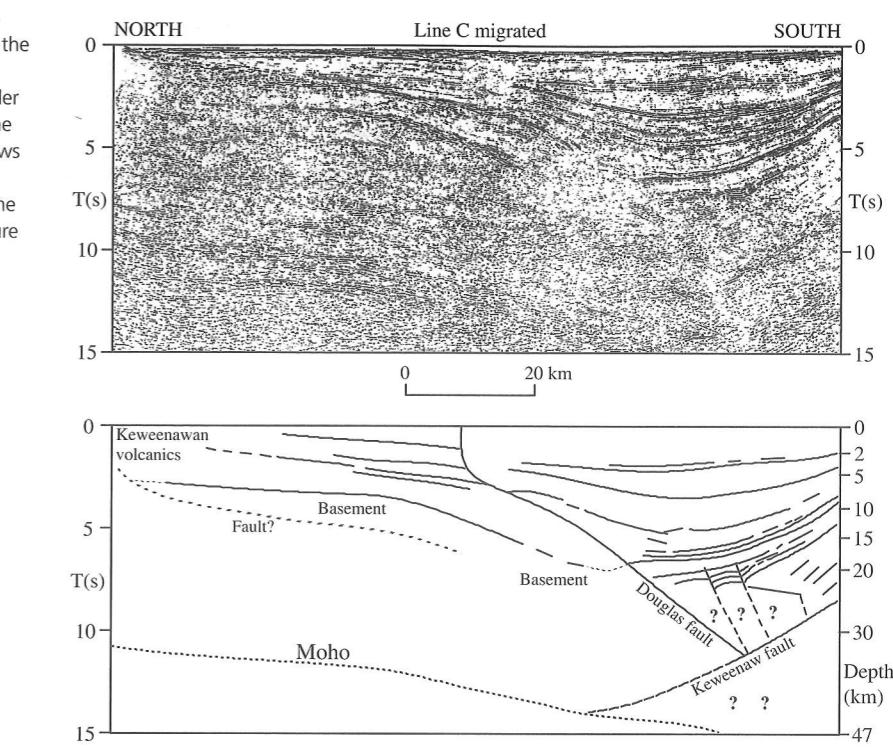
$$t = \frac{2d}{V_1 \cos i_c} + \frac{(x - 2d \tan i_c)}{V_2} \quad (3.122)$$

Rearranging terms and using Snell's law,  $\sin i_c = V_1/V_2$ , we get for the travel-time of the doubly refracted ray

$$t = \frac{x}{V_2} + \frac{2d}{V_1} \cos i_c \quad (3.123)$$

### 3.6 SEISMIC WAVE PROPAGATION

**Fig. 3.70** Results of deep reflection profiling across the North American Mid-continent Rift System under western Lake Superior. The top part of the figure shows the migrated reflection record, the bottom part the interpreted crustal structure (courtesy of A. G. Green).



**Fig. 3.71** Travel-time versus distance curves for the direct ray and the reflected and refracted rays at a horizontal interface between two layers with seismic velocities  $V_1$  and  $V_2$  ( $V_1 > V_2$ ).

The equation represents a straight line with slope  $m_2 = 1/V_2$ . The doubly refracted rays are only recorded at distances greater than the critical distance  $x_c$ . The first arrival recorded at  $x_c$  can be regarded as both a doubly

refracted ray and a reflection; the travel-time line for the head wave is tangential to the reflection hyperbola at  $x_c$ . By backward extrapolation the refraction  $t-x$  curve is found to intersect the time axis at the *intercept time*  $t_i$ , given by

$$t_i = \frac{2d}{V_1} \cos i_c = 2d \frac{\sqrt{V_2^2 - V_1^2}}{V_1 V_2} \quad (3.124)$$

Close to the shot-point the direct ray is the first to be recorded. However, the doubly refracted ray travels part of its path at the faster velocity of the lower layer, so that it eventually overtakes the direct ray and becomes the first arrival. The straight lines for the direct and doubly refracted rays cross each other at this distance, which is accordingly called the *crossover distance*,  $x_{cr}$ . It is computed by equating the travel-times for the direct and refracted rays:

$$\frac{x}{V_1} = \frac{x}{V_2} + 2d \frac{\sqrt{V_2^2 - V_1^2}}{V_1 V_2} \quad (3.125)$$

$$x_{cr} = 2d \sqrt{\frac{V_2 + V_1}{V_2 - V_1}} \quad (3.126)$$

Refraction seismology gives the velocities of subsurface layers directly from the reciprocal slopes of the straight lines corresponding to the direct and doubly refracted rays. Once these velocities have been determined it is possible to compute the depth  $d$  to the interface by using either the intercept time  $t_i$  or the crossover distance  $x_{cr}$ , which can be read directly from the  $t-x$  plot:

$$d = \frac{1}{2} t_i \sqrt{V_2^2 - V_1^2} \quad (3.127)$$

$$d = \frac{1}{2} x_{cr} \sqrt{\frac{V_2 - V_1}{V_2 + V_1}} \quad (3.128)$$

### 3.6.6.2 Refraction at an inclined interface

In practice, the refracting interface is often not horizontal. The assumption of flat layers then leads to errors in the velocity and depth estimates. When the refractor is suspected to have a dip, the velocities of the beds and the dip of the interface can be obtained by shooting a second complementary profile in the opposite direction. Suppose a refractor dips at an angle  $\theta$  as in Fig. 3.72. Shot-points A and B are located at the ends of a geophone layout that covers AB. The ray ACDB from the shot-point A strikes the interface at the critical angle  $i_c$  at C, runs as a head wave with velocity  $V_2$  along the dipping interface, and the ray emerging at D eventually reaches a geophone at the end of the profile at B. During reverse shooting, the ray from the shot-point at B to a geophone at A traverses the same path in the reverse direction. However, the  $t-x$  curves are different for the up-dip and down-dip shots. Let  $d_A$  and  $d_B$  be the perpendicular distances from the shot-points A and B to the interface at P and Q, respectively. For the down-dip shot at A the travel-time to distance  $x$  is given by

$$t_d = \frac{AC + DB}{V_1} + \frac{CD}{V_2} \quad (3.129)$$

The geometry of Fig. 3.72 gives the following trigonometric relationships:

$$\begin{aligned} AC &= \frac{d_A}{\cos i_c} & DB &= \frac{d_B}{\cos i_c} \\ PC &= d_A \tan i_c & DQ &= d_B \tan i_c \\ CD &= x \cos \theta - (PC + DQ) \end{aligned} \quad (3.130)$$

we have

$$\begin{aligned} t_d &= \frac{(d_A + d_B)}{V_1 \cos i_c} + \frac{x \cos \theta - (d_A + d_B) \tan i_c}{V_2} \\ &= \frac{x \cos \theta}{V_2} + \frac{(d_A + d_B)}{V_1} \cos i_c \end{aligned} \quad (3.131)$$

Equation (3.131) can be simplified by noting that

$$d_B = d_A + x \sin \theta \quad \text{and} \quad \frac{1}{V_2} = \frac{\sin i_c}{V_1} \quad (3.132)$$

After substitution and gathering terms the down-dip travel-time is given by

$$\begin{aligned} t_d &= \frac{x \sin i_c \cos \theta}{V_1} + \frac{x \cos i_c \sin \theta}{V_1} + \frac{2d_A \cos i_c}{V_1} \\ &= \frac{x}{V_1} \sin(i_c + \theta) + t_{id} \end{aligned} \quad (3.133)$$

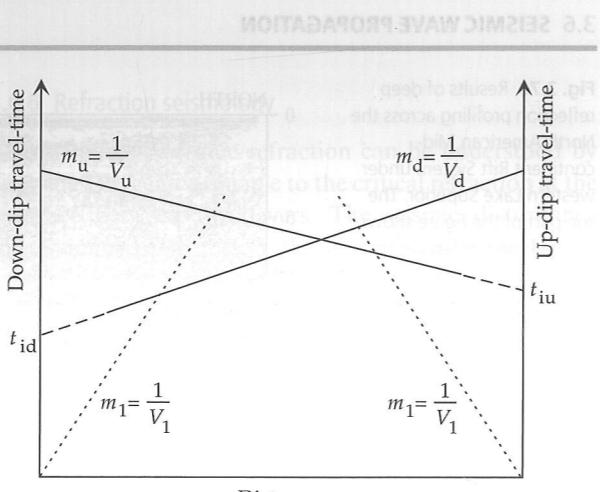


Fig. 3.72 Travel-time versus distance curves of direct and refracted rays for up-dip and down-dip profiles when the refracting boundary dips at angle  $\theta$ .

where  $t_{id}$  is the intercept time for the down-dip shot:

$$t_{id} = \frac{2d_A}{V_1} \cos i_c \quad (3.134)$$

The analysis for shooting in the up-dip direction is analogous and gives

$$t_u = \frac{x}{V_1} \sin(i_c - \theta) + t_{iu} \quad (3.135)$$

where  $t_{iu}$  is the intercept time for the up-dip shot:

$$t_{iu} = \frac{2d_B}{V_1} \cos i_c \quad (3.136)$$

If the upper layer is homogeneous, the segments for the direct ray will have equal slopes, the reciprocals of which give the velocity  $V_1$  of the upper layer. The segments of the  $t-x$  curves corresponding to the doubly refracted ray are different for up-dip and down-dip shooting. The total travel-times in either direction along ACDB must be equal, but the  $t-x$  curves have different intercept times. As these are proportional to the perpendicular distances to the refractor below the shot-points, the up-dip intercept time  $t_{iu}$  is larger than the down-dip intercept time  $t_{id}$ . This means that the slope of the up-dip refraction in Fig. 3.72 is flatter than the down-dip slope. If we interpret the reciprocal of the slope as the velocity of the lower medium, we get two apparent velocities,  $V_d$  and  $V_u$ , given by

### 3.6 SEISMIC WAVE PROPAGATION

$$\begin{aligned} \frac{1}{V_d} &= \frac{1}{V_1} \sin(i_c + \theta) \\ \frac{1}{V_u} &= \frac{1}{V_1} \sin(i_c - \theta) \end{aligned} \quad (3.137)$$

Once the real velocity  $V_1$  and the apparent velocities  $V_d$  and  $V_u$  have been determined from the  $t-x$  curves, the dip of the interface  $\theta$  and the critical angle  $i_c$  (and from it the true velocity  $V_2$  of the lower layer) can be computed:

$$\theta = \frac{1}{2} \left[ \sin^{-1} \left( \frac{V_1}{V_d} \right) - \sin^{-1} \left( \frac{V_1}{V_u} \right) \right] \quad (3.138)$$

$$i_c = \frac{1}{2} \left[ \sin^{-1} \left( \frac{V_1}{V_d} \right) + \sin^{-1} \left( \frac{V_1}{V_u} \right) \right] \quad (3.139)$$

If the reciprocal apparent velocities (Eq. (3.137)) are added, a simple approximation for the true velocity of the lower layer is obtained:

$$\begin{aligned} \frac{1}{V_d} + \frac{1}{V_u} &= \frac{1}{V_1} (\sin(i_c + \theta) + \sin(i_c - \theta)) \\ &= \frac{2}{V_2} \sin i_c \cos \theta \\ &= \frac{2}{V_2} \cos \theta \end{aligned} \quad (3.140)$$

If the refractor dip is small,  $\cos \theta \approx 1$  (for example, if  $\theta < 15^\circ$ ,  $\cos \theta > 0.96$ ) and an approximate formula for the true velocity of the second layer is

$$\frac{1}{V_2} \approx \frac{1}{2} \left( \frac{1}{V_d} + \frac{1}{V_u} \right) \quad (3.141)$$

### 3.6.6.3 Refraction with continuous change of velocity with depth

Imagine the Earth to have a multi-layered structure with numerous thin horizontal layers, each characterized by a constant seismic velocity, which increases progressively with increasing depth (Fig. 3.73). A seismic ray that leaves the surface with angle  $i_1$  will be refracted at each interface until it is finally refracted critically. The ray that finally returns to the surface will have an emergence angle equal to  $i_1$ . Snell's law applies to each successive refraction (e.g., at the top surface of the  $n$ th layer, which has velocity  $V_n$ )

$$\frac{\sin i_1}{V_1} = \frac{\sin i_2}{V_2} = \dots = \frac{\sin i_n}{V_n} = \text{constant} = p \quad (3.142)$$

The constant  $p$  is called the *ray parameter*. It is characteristic for a particular ray with emergence angle  $i_1$  and velocity  $V_1$  in the surface layer. If  $V_m$  is the velocity of the deepest layer, along whose surface the ray is eventually critically refracted ( $\sin i_m = 1$ ), then the value of  $p$  must be equal to  $1/V_m$ .

As the number of layers increases and the thickness of each layer decreases, the situation is approached in which

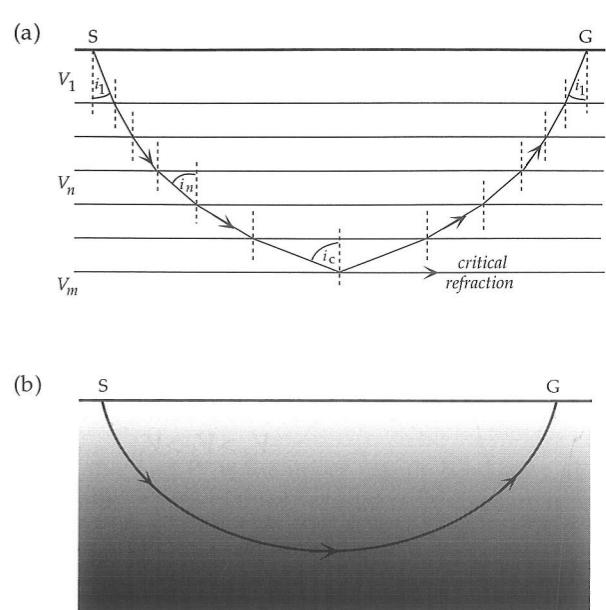


Fig. 3.73 (a) The path of a seismic wave through a horizontally layered medium, in which the seismic velocity is constant in each layer and increases with increasing depth, becomes ever flatter until critical refraction is reached; the return path of each emerging ray mirrors the incident path. (b) When the velocity increases continuously with depth, the ray is a smooth curve that is concave upward.

the velocity increases constantly with increasing depth. Each ray then has a smoothly curved path. If the vertical increase in velocity is linear with depth, the curved rays are circular arcs.

In the above we have assumed that the refracting interfaces are horizontal. This type of analysis is common in seismic prospecting, where only local structures and comparatively shallow depths are evaluated. The passage of seismic body waves through a layered spherical Earth can be treated to a first approximation in the same way. We can represent the vertical (radial) velocity structure by subdividing the Earth into concentric shells, each with a faster body-wave velocity than the shell above it (Fig. 3.74). Snell's law of refraction applies to the interface between each pair of shells. For example, at point A we can write

$$\frac{\sin i_1}{V_1} = \frac{\sin a_1}{V_2} \quad (3.143)$$

Multiplying both sides by  $r_1$  gives

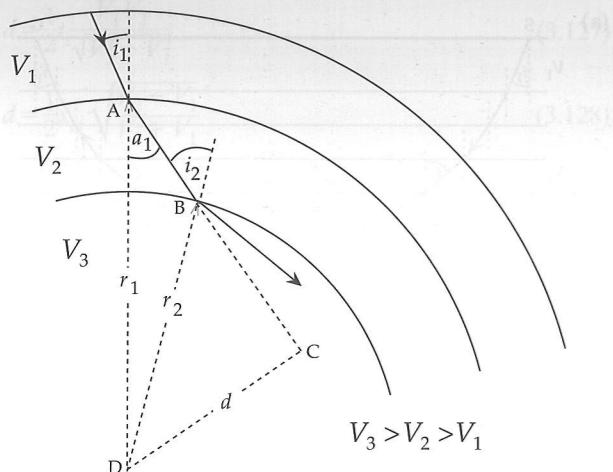
$$\frac{r_1 \sin i_1}{V_1} = \frac{r_1 \sin a_1}{V_2} \quad (3.144)$$

In triangles ACD and BCD, respectively, we have

$$d = r_1 \sin a_1 = r_2 \sin a_2 \quad (3.145)$$

Combining Eqs. (3.143), (3.144) and (3.145) gives the result

$$\frac{r_1 \sin i_1}{V_1} = \frac{r_2 \sin i_2}{V_2} = \dots = \frac{r_n \sin i_n}{V_n} = p \quad (3.146)$$



**Fig. 3.74** Refraction of a seismic ray in a spherically layered Earth, in which the seismic velocity is constant in each layer and the layer-velocity increases with depth.

The constant  $p$  is again called the *ray parameter*, although it has different dimensions than in Eq. (3.142) for flat horizontal layers. Here the seismic ray is a straight line within each spherical layer with constant velocity. If the velocity increases continuously with depth, the seismic ray is refracted continuously and its shape is curved concavely upward. It reaches its deepest point when  $\sin i = 1$ , at radius  $r_0$  where the velocity is  $V_0$ ; these parameters are related by the *Benndorf relationship*:

$$\frac{r \sin i}{V} = \frac{r_0}{V_0} = p \quad (3.147)$$

Determination of the ray parameter is the key to determining the variation of seismic velocity inside the Earth. Access to the Earth's interior is provided by analysis of the travel-times of seismic waves that have traversed the various internal regions and emerge at the surface, where they are recorded. We will see in Section 3.7.3.1 that the travel-time ( $t$ ) of a seismic ray to a known epicentral distance ( $\Delta$ ) can be mathematically inverted to give the velocity  $V_0$  at the deepest point of the path. The theory applies for P- and S-waves, the general velocity  $V$  being replaced by the appropriate velocity  $\alpha$  or  $\beta$ , respectively.

### 3.7 INTERNAL STRUCTURE OF THE EARTH

#### 3.7.1 Introduction

It is well known that the Earth has a molten core. What is now general knowledge was slow to develop. In order to explain the existence of volcanoes, some nineteenth century scientists postulated that the Earth must consist of a rigid outer crust around a molten interior. It was also known in the last century that the mean density of the Earth is about 5.5 times that of water. This is much larger than the known specific density of surface rocks, which is about 2.5–3. From this it was inferred that density

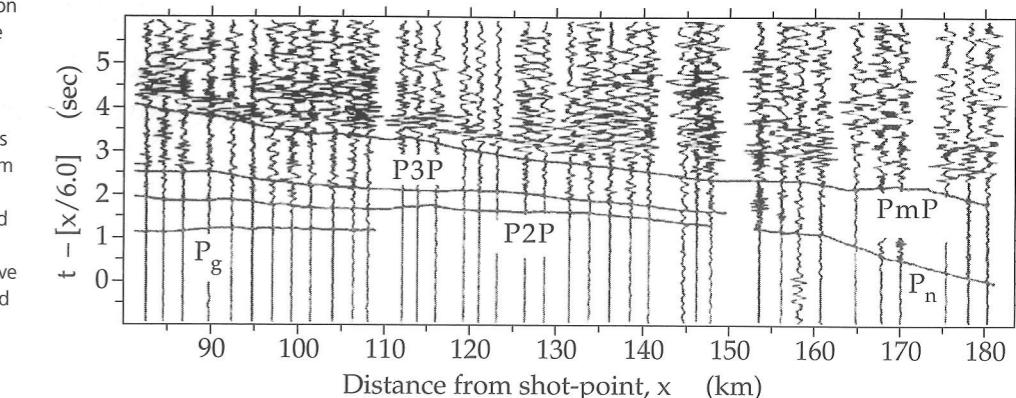
increased towards the Earth's center under the effect of gravitational pressure. The density at the Earth's center was estimated to be comparatively high, greater than  $7000 \text{ kg m}^{-3}$  and probably in the range  $10,000\text{--}12,000 \text{ kg m}^{-3}$ . It was known that some meteorites had a rock-like composition, while others were much denser, composed largely of iron. In 1897 E. Wiechert, who subsequently became a renowned German seismologist, suggested that the interior of the Earth might consist of a dense metallic core, cloaked in a rocky outer cover. He called this cloak the "Mantel," which later became anglicized to mantle.

The key to modern understanding of the interior of the Earth – its density, pressure and elasticity – was provided by the invention of the Milne seismograph. The progressive refinement of this instrument and its systematic employment world-wide led to the rapid development of the modern science of seismology. Important results were obtained early in the twentieth century. The Earth's fluid core was first detected seismologically in 1906 by R. D. Oldham. He observed that, if the travel-times of P-waves observed at epicentral distances of less than  $100^\circ$  were extrapolated to greater distances, the expected travel-times were less than those observed. This meant that the P-waves arriving at large epicentral distances were delayed in their passage through the Earth. Oldham inferred from this the existence of a central core in which the P-wave velocity was reduced. He predicted that there would be a region of epicentral distances (a "shadow zone") in which P-waves could not arrive. About this time it was found that P- and S-waves passed through the mantle but that no S-waves arrived beyond an epicentral distance of  $105^\circ$ . In 1914, B. Gutenberg verified the existence of a shadow zone for P-waves in the range of epicentral distances between  $105^\circ$  and  $143^\circ$ . Gutenberg also located the depth of the core–mantle boundary with impressive accuracy at about 2900 km. A modern estimate of the radius of the core is  $3485 \pm 3$  km, giving a mantle 2885 km thick. Gutenberg also predicted that P-waves and S-waves would be reflected from the core–mantle boundary. These waves, known today as  $P_{cP}$  and  $S_{cS}$  waves, were not observed until many years later. In honor of Gutenberg the core–mantle boundary is known today as the *Gutenberg seismic discontinuity*.

While studying the P-wave arrivals from an earthquake in Croatia in 1909 Andrija Mohorovičić found only a single arrival ( $P_g$ ) at distances close to the epicenter. Beyond about 200 km there were two arrivals; the  $P_g$  event was overtaken by another arrival ( $P_n$ ) which had evidently travelled at higher speed. Mohorovičić identified  $P_g$  as the direct wave from the earthquake and  $P_n$  as a doubly refracted wave (equivalent to a head wave) that travelled partly in the upper mantle. Mohorovičić calculated velocities of  $5.6 \text{ km s}^{-1}$  for  $P_g$  and  $7.9 \text{ km s}^{-1}$  for  $P_n$  and estimated that the sudden velocity increase occurred at a depth of 54 km. This seismic discontinuity is now called the *Mohorovičić discontinuity*, or *Moho* for short. It represents the boundary between the crust and mantle.

### 3.7 INTERNAL STRUCTURE OF THE EARTH

**Fig. 3.75** A SW-NE refraction seismic profile parallel to the strike of the Swiss Alps. The seismograms from adjacent stations have been modified to show reduced travel-times as a function of distance from the shot-point. The layered subsurface structure is traced by connecting main arrivals such as the crustal direct wave  $P_g$ , mantle head wave  $P_n$ , and mantle reflection  $PmP$  (after Maurer and Ansorge, 1992).



The crustal thickness is known to be very variable. It averages about 33 km, but measures as little as 5 km under oceans and as much as 60–80 km under some mountain ranges.

The *seismological Moho* is commonly defined as the depth at which the P-wave velocity exceeds  $7.6 \text{ km s}^{-1}$ . This seismic definition is dependent on the density and elastic properties of crustal and mantle rocks, and need not correspond precisely to a change of rock type. An alternative definition of the Moho as the depth where the rock types change is called the *petrological Moho*. For most purposes the two definitions of the crust–mantle boundary are equivalent.

It is now known that the crust is not homogeneous but has a vertically layered structure. In 1925 V. Conrad separated arrivals from a Tauern (Eastern Alps) earthquake of 1923 into  $P_g$  and  $S_g$  waves in an upper crustal layer and faster  $P^*$  and  $S^*$  waves that travelled with velocities  $6.29 \text{ km s}^{-1}$  and  $3.57 \text{ km s}^{-1}$ , respectively, in a deeper layer. Because the  $P^*$  and  $S^*$  velocities are significantly slower than corresponding upper mantle velocities, Conrad deduced that they were head waves from a lower crustal layer. The interface separating the continental crust into an upper crustal layer and a lower crustal layer is called the *Conrad discontinuity*. Influenced by early petrological models of crustal composition and by comparison with seismic velocities in known materials, seismologists referred to the upper and lower crustal layers as the granitic layer and the basaltic layer, respectively. This petrological separation is now known to be overly simplistic. In contrast to the Moho, which is found everywhere, the Conrad discontinuity is poorly defined or absent in some areas.

The appearance of some of these arrivals on refraction seismic records from the European continental crust is illustrated in Fig. 3.75. The vertical axis in this figure shows a reduced travel-time, obtained by dividing the shot-point to receiver distance by a representative crustal velocity ( $6 \text{ km s}^{-1}$  in this case) and subtracting this from the observed travel-time. This method of displaying the data prevents the plot from becoming unwieldy at large distances. The crustal direct wave  $P_g$  is represented as a nearly horizontal arrival,  $PmP$  is a P-wave reflected from

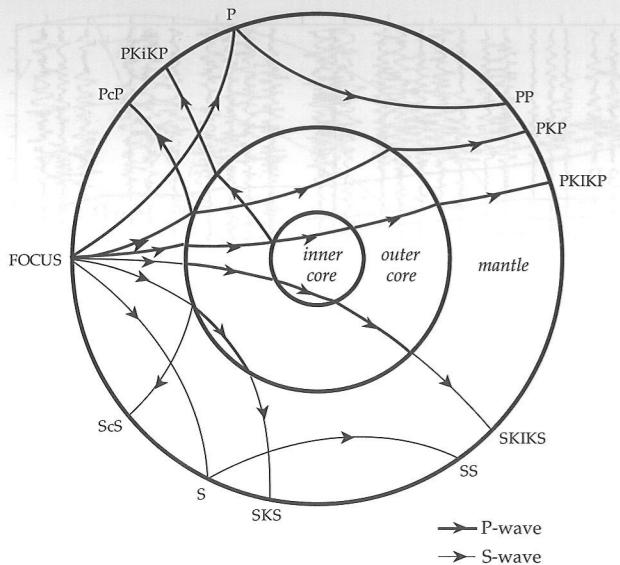
the Moho, and  $P_n$  is the upper mantle head wave along the Moho. Additional intracrustal reflections are labeled  $P2P$ ,  $P3P$ , etc.

Similar designations are used for events in seismic sections of oceanic crust. The oceanic crustal layers consist of sedimentary Layer 1, basaltic Layer 2, and gabbroic Layer 3 (Section 3.7.5.1). The direct wave in Layer 2 is called  $P_2$ , the head wave at the basalt–gabbro interface is referred to as  $P_3$ , and the Moho head wave is  $P_n$ .

The core shadow-zone and its interpretation in terms of a fluid core were well established in 1936 when Inge Lehmann, a Danish seismologist, reported weak P-wave arrivals within the shadow zone. She interpreted these in terms of an inner core with higher seismic velocity. However, the existence of the inner core remained controversial for many years. Improved seismometer design, digital signal treatment and the setting up of seismic arrays have provided corroborating evidence. The existence of a solid inner core is also supported by analyses of the Earth's natural vibrations.

The gross internal structure of the Earth is modelled as a set of concentric shells, corresponding to the inner core, outer core and mantle (Fig. 3.76). An important step in understanding this layered structure has been the development of travel-time curves for seismic rays that pass through the different shells. To facilitate identification of the arrivals of these rays on seismograms, a convenient shorthand notation is used. A P- or S-wave that travels from an earthquake directly to a seismometer is labelled with the appropriate letter P or S; until the margin of the core shadow-zone, the P- and S-waves follow identical curved paths. (The curvature, as explained in Section 3.6.6.3, arises from the increase in seismic velocity with depth.) A wave that reaches the seismometer after being reflected once from the crust is labelled PP (or SS), as its path consists of two identical P- or S-segments.

The energy of an incident P- or S-wave is partitioned at an interface into reflected and refracted P- and S-waves (see Section 3.6.4). A P-wave incident on the boundary between mantle and fluid outer core is refracted towards the normal to the interface, because the P-wave velocity drops from about  $13 \text{ km s}^{-1}$  to about  $8 \text{ km s}^{-1}$  at the boundary. After a second refraction it emerges beyond

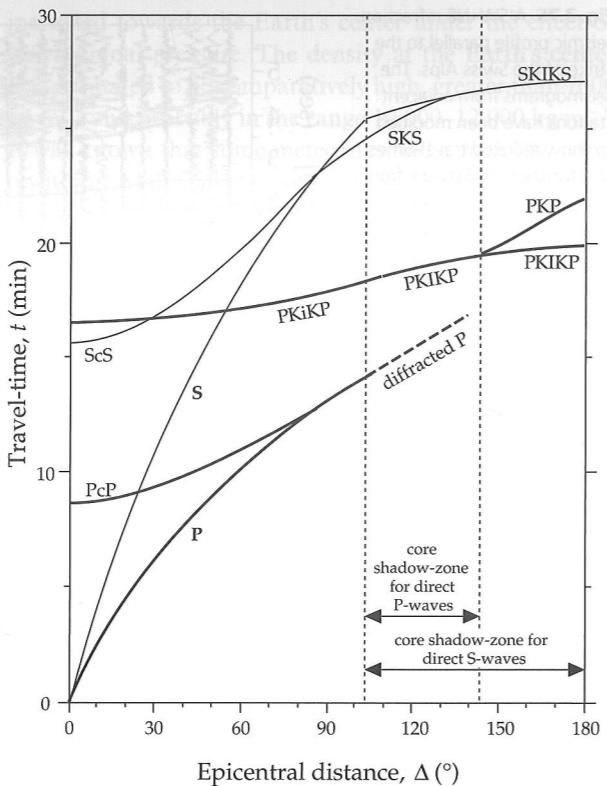


**Fig. 3.76** Seismic wave paths of some important refracted and reflected P-wave and S-wave phases from an earthquake with focus at the Earth's surface.

the shadow zone and is called a PKP wave (the letter K stands for Kern, the German word for core). An S-wave incident at the same point has a lower mantle velocity of about  $7 \text{ km s}^{-1}$ . Part of the incident energy is converted to a P-wave in the outer core, which has a higher velocity of  $8 \text{ km s}^{-1}$ . The refraction is away from the normal to the interface. After a further refraction the incident S-wave reaches the surface as an SKS phase. A P-wave that travels through mantle, fluid core and inner core is labelled PKIKP. Each of these rays is refracted at an internal interface. To indicate seismic phases that are reflected at the outer core boundary the letter c is used, giving rise, for example, to Pcp and ScS phases (Fig. 3.76). Reflections from the inner core are designated with the letter i, as for example in the phase PKiKP.

### 3.7.2 Refractions and reflections in the Earth's interior

If it possesses sufficient energy, a seismic disturbance may be refracted and reflected – or converted from a P-wave to an S-wave, or vice versa – many times at the several seismic discontinuities within the Earth and at its free surface. As a result, the seismogram of a large earthquake contains numerous overlapping seismic signals and the identification of individual phases is a difficult task. Late-arriving phases that have been multiply reflected or that have travelled through several regions of the Earth's interior are difficult to resolve from the disturbance caused by earlier arrivals. In the period 1932–1939 H. Jeffreys and K. E. Bullen analyzed a large number of good records of earthquakes registered at a world-wide, though sparse, distribution of seismic stations. In 1940 they published a set of tables giving the travel-times of P- and S-waves through the Earth. A slightly different set of tables was reported by B. Gutenberg and C. F. Richter. The good



**Fig. 3.77** Travel-time versus epicentral distance ( $t$ - $\Delta$ ) curves for some important seismic phases (modified from Jeffreys and Bullen, 1940).

agreement of the independent analyses confirmed the reliability of the results. The Jeffreys–Bullen seismological tables were used by the international seismic community as the standard of reference for many years.

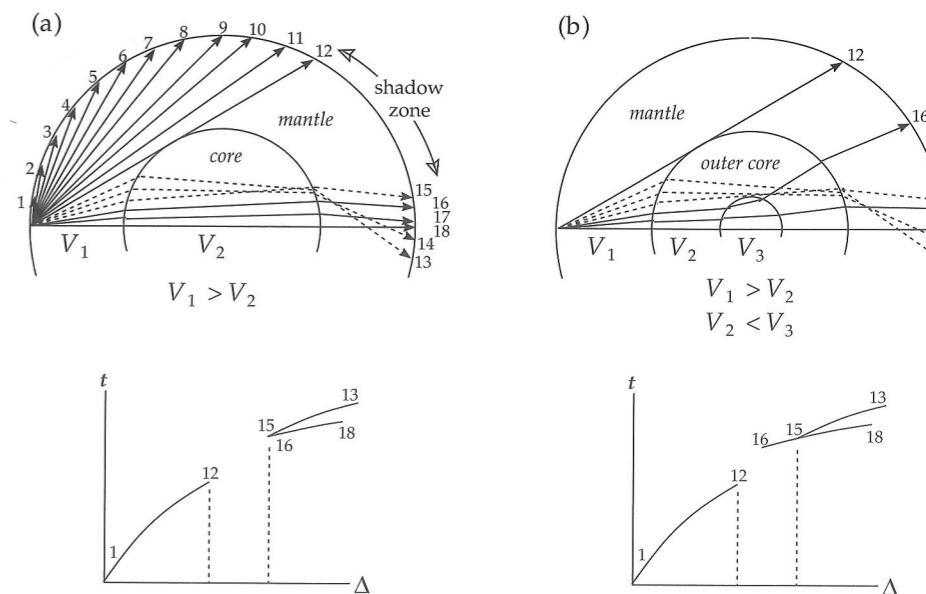
The travel-time of a seismic wave to a given epicentral distance is affected by the focal depth of the earthquake, which may be as much as several hundred kilometers. The travel-time versus distance curves of some important phases are shown in Fig. 3.77 for an earthquake occurring at the Earth's surface. The model assumes that the Earth is spherically symmetric, with the same vertical structure underneath each place on the surface. This assumption works fairly well, although it is not quite true. Lateral variations of seismic velocity have been found at many depths within the Earth. For example, there are lateral differences in seismic velocity between oceanic and continental crust, and between oceanic and continental lithosphere. At even greater depths significant lateral departures from the spherical model have been detected. These discrepancies form the basis of the branch of seismology called *seismic tomography*, which we will examine later.

#### 3.7.2.1 Seismic rays in a uniformly layered Earth

It is important to understand clearly the relationship between the travel-time ( $t$ ) versus epicentral distance ( $\Delta$ ) curves and the paths of seismic waves in the Earth, like those shown in Fig. 3.76. Consider first an Earth that

### 3.7 INTERNAL STRUCTURE OF THE EARTH

**Fig. 3.78** Seismic wave paths and their  $t$ - $\Delta$  curves for P-waves passing through a spherical Earth with constant velocities in the mantle, outer core and inner core, respectively. (a) Development of a shadow zone when the mantle velocity ( $V_1$ ) is higher than the outer core velocity ( $V_2$ ). (b) Penetration of the shadow zone by rays refracted in an inner core with higher velocity than the outer core ( $V_3 > V_2$ ).



consists of two concentric shells representing the mantle and core (Fig. 3.78a). The P-wave velocity in each shell is constant, and is faster in the mantle than in the core ( $V_1 > V_2$ ). The figure shows the paths of 18 rays that leave a surface source at angular intervals of  $5^\circ$ . Rays 1–12 travel directly through the mantle as P-waves and emerge at progressively greater epicentral distances. The convex upwards shape of the  $t$ - $\Delta$  curve is here due to the curved outer surface, the layer velocity being constant. Ray 13 is partially refracted into the core, and partially reflected (not shown in the figure). Because  $V_2 < V_1$  the refracted ray is bent towards the Earth's radius, which is normal to the refracting interface. This ray is further bent on leaving the core and reaches the Earth's surface as a PKP phase at an epicentral distance greater than  $180^\circ$ . Rays 14 and 15 impinge more directly on the core and are refracted less severely; their epicentral distances become successively smaller and their travel-times become shorter than that of ray 13, as indicated by branch 13–15 of the  $t$ - $\Delta$  curve. This branch is offset in time from the extrapolation of branch 1–12 because of the lower velocity in the core. The paths of rays 16, 17 and 18 (which is a straight line through mantle and core and emerges at an epicentral distance of  $180^\circ$ ) become progressively longer, and a second branch 16–18 develops on the  $t$ - $\Delta$  curve. The two branches meet at a sharp point, or cusp. No P-waves reach the surface in the gap between rays 12 and 15 in this simple model. There is a *shadow zone* between the last P-wave that just touches the core and the PKP-wave with the smallest epicentral distance. The existence of a shadow zone for P-waves is evidence for a core with lower P-wave velocities than the mantle. S-waves in the mantle follow the same ray paths as the P-waves. However, no direct S-waves arrive in the shadow zone, which indicates that the core must be fluid.

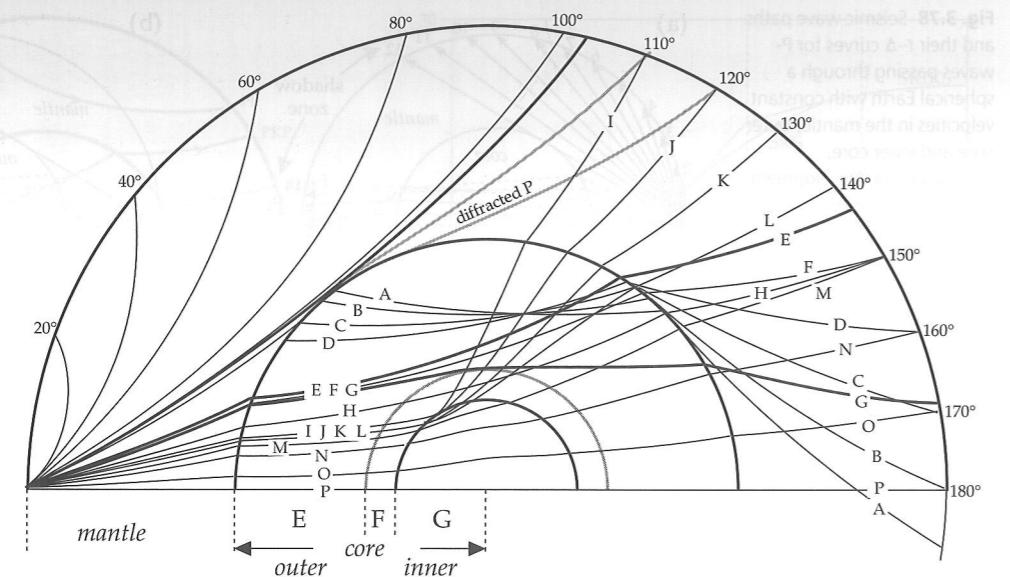
Now suppose an inner core with a constant velocity  $V_3$  that is higher than the velocity  $V_2$  in the outer core

(Fig. 3.78b). The paths of rays 1–15 are the same as before, through the mantle and outer core. The segments 1–12 and 13–15 of the  $t$ - $\Delta$  curve are the same as previously. Ray 16 impinges on the inner core and is sharply refracted away from the Earth's radius; on returning to the outer core it is again refracted, back towards the radius. After further refraction at the core–mantle interface this ray emerges at the Earth's surface at a smaller epicentral distance than ray 15, within the P-wave shadow-zone, as a PKIPK event. Successive PKIPK rays are bent less strongly. The PKIPK rays map out a new branch 16–18 of the  $t$ - $\Delta$  curve (see also Fig. 3.77).

#### 3.7.2.2 Travel-time curves for P-, PKP- and PKIKP-waves

In general the velocities of P-waves and S-waves increase with depth. As described in Section 3.6.6.3 and illustrated in Fig. 3.73a, the ray paths are curved lines, concave towards the surface. However, the explanation of the paths of P, PKP, and PKIKP phases shown in Fig. 3.79 closely follows the preceding discussion. There is a shadow zone for direct P-waves between about  $103^\circ$  and  $143^\circ$ , and no direct S-waves are found beyond  $103^\circ$ . The shallowest PKP ray (A in Fig. 3.79) is deviated the furthest, emerging at an epicentral distance greater than  $180^\circ$ . Successively deeper PKP rays (B–E) emerge at ever-smaller epicentral distances until about  $143^\circ$ , after which the epicentral distance increases again to almost  $170^\circ$  (rays F, G). It was long believed that the boundary between inner and outer core was a transitional region (called region F in standard Earth models) with higher P-wave velocity, and that PKP rays traversing this region would again emerge at smaller epicentral distances (ray H). The first rays penetrating the inner core are sharply refracted and emerge in the P-wave shadow zone. The most strongly deviated (ray I) is observed at an epicentral distance of about  $110^\circ$ ; deeper rays (J–P) arrive at

**Fig. 3.79** The wave paths of some P, PKP, and PKIKP rays (after Gutenberg, 1959).



ever-greater distances up to 180°. There are at least two branches of the  $t-\Delta$  curve for  $\Delta > 143^\circ$ , corresponding to the PKP and PKIKP phases, respectively (Fig. 3.77). In fact, depending how the transitional region F is modelled, the  $t-\Delta$  curve near 143° can have several branches.

The edges of the shadow zone defined by P and PKP phases are not sharp. One reason is the intrusion of PKIKP phases at the 143° edge. Another is the effect of diffraction of P-waves at the 103° edge (Fig. 3.79). The bending of plane waves at an edge into the shadow of an obstruction was described in Section 3.6.2.3, and explained with the aid of Huygen's principle. The diffraction of plane waves is called *Fraunhofer diffraction*. When their source is not at infinity, waves must be handled as spherical waves. Spherical wavefronts that pass an obstacle are also diffracted. This type of behavior is called *Fresnel diffraction*, and it is also explainable with Huygen's principle as the product of interference between the primary wavefront and secondary waves generated at the obstacle. Wave energy penetrates into the shadow of the obstacle, as though the wavefront were bent around the edge. In this way very deep P-waves are diffracted around the core and into the shadow zone. The intensity of the diffracted rays falls off with increasing angular distance from the diffracting edge, in this case the core-mantle boundary. Modern instrumentation enables detection of long-period diffracted P-waves to large epicentral distances (Figs. 3.77, 3.79). The velocity structure above the core-mantle boundary, in particular in the D''-layer (Section 3.7.5.3), has a strong influence on the ray paths, travel-times and waveforms of the diffracted waves.

### 3.7.3 Radial variations of seismic velocities

Models of the radial variations of physical parameters inside the Earth implicitly assume spherical symmetry. They are therefore "average" models of the Earth that do

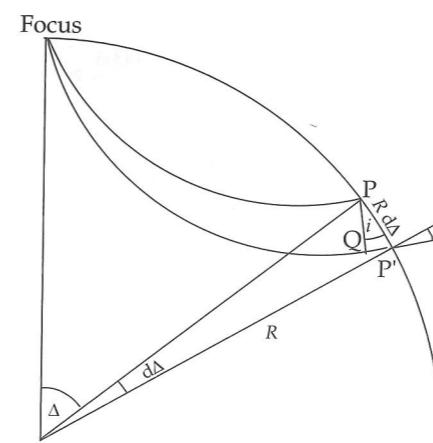
not take into account lateral variations (e.g., of velocity or density) at the same depth. This is a necessary first step in approaching the true distributions as long as lateral variations are relatively small. This appears to be the case; although geophysically significant, the lateral variations in physical properties remain within a few percent of the average value at any depth.

There are two main ways to determine the distributions of body-wave velocities in a spherically symmetric Earth. They are referred to as forward and inverse modelling. Both methods have to employ the same sets of observations, which are the travel-times of different seismic phases to known epicentral distances. The forward technique starts with a known or assumed variation of seismic velocities and calculates the corresponding travel-times. The inversion method starts with the observed  $t-\Delta$  curves and computes a model of the velocity distributions that could produce the curves. The inversion method is the older one, in use since the early part of the twentieth century, and forms an important branch of mathematical theory. Forward modelling is a more recent method that has been successfully employed since the advent of powerful computers.

#### 3.7.3.1 Inversion of travel-time versus distance curves

In 1907 the German geophysicist E. Wiechert, building upon an evaluation of the Benndorf problem (Eq. (3.147)) by the mathematician G. Herglotz, developed an analytical method for computing the internal distributions of seismic velocities from observations made at the Earth's surface. The technique is called inversion of travel-times, and it is considered one of the classical methods of geophysics. The observational data consist of the  $t-\Delta$  curves for important seismic phases (Fig. 3.77). The clues to deciphering the velocity distributions were the Benndorf relationship for the ray

## 3.7 INTERNAL STRUCTURE OF THE EARTH



**Fig. 3.80** Paths of two rays that leave an earthquake focus at infinitesimally different angles, reaching the surface at points P and P' at epicentral distances  $\Delta$  and  $\Delta + d\Delta$ , respectively.

parameter  $p$  (Eq. (3.147)) and the recognition that the value of  $p$  can be obtained from the slope of the travel-time curve at the epicentral distance where the ray returns to the surface.

Consider two rays that leave an earthquake focus at infinitesimally different angles, reaching the surface at points P and P' at epicentral distances  $\Delta$  and  $\Delta + d\Delta$ , respectively (Fig. 3.80). The distance PP' is  $R d\Delta$  (where  $R$  is the Earth's radius) and the difference in arc-distances along the adjacent rays is equal to  $V dt$ , where  $V$  is the velocity in the surface layer and  $dt$  is the difference in travel-times of the two rays. In the small triangle PP'Q the angle QPP' is equal to  $i$ , the angle of emergence (incidence). Therefore,

$$\sin i = \frac{V dt}{R d\Delta} \quad (3.148)$$

$$R \sin i = p = \frac{dt}{d\Delta} \quad (3.149)$$

This means that the value of  $p$  for the ray emerging at epicentral distance  $\Delta$  can be obtained by calculating the slope ( $dt/d\Delta$ ) of the travel-time curve for that distance. This is an important step in finding the velocity  $V_0$  at the deepest point of the ray, at radius  $r_0$ , because  $V_0 = r_0/p$ . However, before we can find the velocity we need to know the value of  $r_0$ . The continuation of the analysis – known as Herglotz–Wiechert inversion – is an intricate mathematical procedure, beyond the scope of this text, which fortunately results in a fairly simple formula:

$$\ln \frac{R}{r_0} = \frac{1}{\pi} \int_0^{\Delta_1} \cosh^{-1} \left( \frac{p(\Delta)}{p(\Delta_1)} \right) d\Delta \quad (3.150)$$

where  $p(\Delta_1)$  is the slope of the  $t-\Delta$  curve at  $\Delta_1$ , the epicentral distance of emergence, and  $p(\Delta)$  is the slope at any intermediate epicentral distance  $\Delta$ . Equation (3.150) can be used to integrate numerically along the ray to give the value of  $r_0$  for the ray, and  $V_0 = r_0/p$ .

The Herglotz–Wiechert inversion is valid for regions of the Earth in which  $p$  varies monotonically with  $\Delta$ . It cannot be used in the Earth's crust, because conditions are too inhomogeneous. Seismic velocity distributions in the crust are deduced empirically from long seismic refraction profiles. The Herglotz–Wiechert method cannot be used where a low-velocity zone is present, because the ray does not bottom in the zone. It works well for the Earth's mantle, but care must be taken where a seismic discontinuity is present. In the Earth's core the refraction of P-waves at the core–mantle boundary means that no PKP waves reach their deepest point in the outer layers of the core (Fig. 3.79). However, SKS-waves bottom in the outer core. Inversion of SKS-wave travel-times complements the inversion of PKP-wave data to give the P-wave velocity distribution in the core.

#### 3.7.3.2 Forward modelling: polynomial parametrization

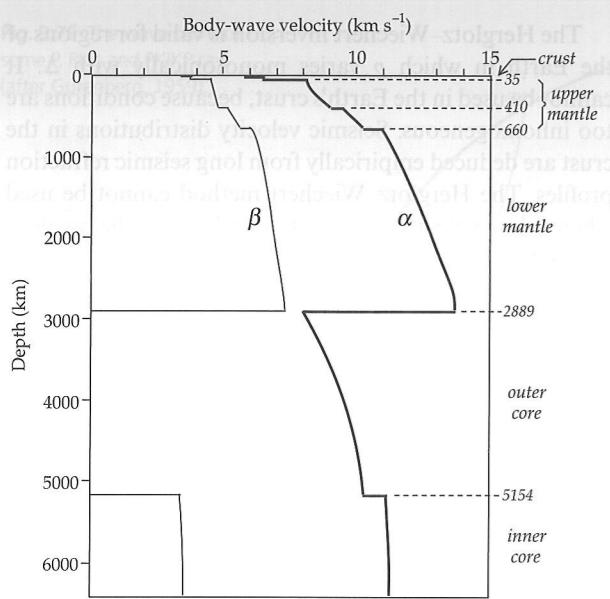
The forward modelling method starts with a presupposed dependence of seismic velocity with depth. The method assumes that the variation of velocity can be expressed by a smooth polynomial function of radial distance within limited depth ranges. This procedure is called polynomial parametrization and in constructing models of the Earth's interior it is applied to the P-wave and S-wave velocities, the seismic attenuation, and the density.

The travel-times of P- and S-waves to any epicentral distance are calculated on the basis of the spherically symmetric, layered model. The computed travel-times are compared with the observed  $t-\Delta$  curves, and the model is adjusted to account for differences. The procedure is repeated as often as necessary until an acceptable agreement between the computed and real travel-times is achieved. The method requires good travel-time data for many seismic phases and involves intensive computation.

In 1981 A. M. Dziewonski and D. L. Anderson constructed a *Preliminary Reference Earth Model* (acronym: PREM) in which the distributions of body-wave velocities in important layers of the Earth were represented by cubic or quadratic polynomials of normalized radial distance; in thin layers of the upper mantle linear relationships were used. A similar, revised parametrized velocity model (*iasp91*) was proposed by B. L. N. Kennett and E. R. Engdahl in 1991. The variations of P- and S-wave velocities with depth in the Earth according to the *iasp91* model are shown in Fig. 3.81.

#### 3.7.4 Radial variations of density, gravity and pressure

In order to determine density, gravity and pressure in the Earth's interior several simplifying assumptions must be made, which appear to be warranted by experience. The Earth is assumed to be spherically symmetric and composed of concentric homogeneous shells or layers (e.g., inner core, outer core, mantle, etc.). Possible effects of chemical and phase changes within a shell are not taken



**Fig. 3.81** The variations with depth of longitudinal- and shear-wave velocities,  $a$  and  $b$ , respectively, in the Earth's interior, according to the Earth model iasp91 (data source: Kennett and Engdahl, 1991).

into account. Pressure and density are assumed to increase purely hydrostatically. If the distributions of seismic body-wave velocities  $\alpha$  and  $\beta$  are known, an important seismic parameter  $\Phi$  can be defined:

$$\Phi = \alpha^2 - \frac{4}{3}\beta^2 \quad (3.151)$$

By comparing Eq. (3.151) and Eq. (3.48) we see that  $\Phi$  is equal to  $K/\rho$ , where  $K$  is the bulk modulus and  $\rho$  the density.

#### 3.7.4.1 Density inside the Earth

Consider a vertical prism between depths  $z$  and  $z + dz$  (Fig. 3.82). The hydrostatic pressure increases from  $p$  at depth  $z$  to  $(p + dp)$  at depth  $(z + dz)$  because of the extra pressure due to the extra material in the small prism of height  $dz$ . The pressure increase  $dp$  is equal to the weight  $w$  of the prism divided by the area  $A$  of the base of the prism, over which the weight is distributed.

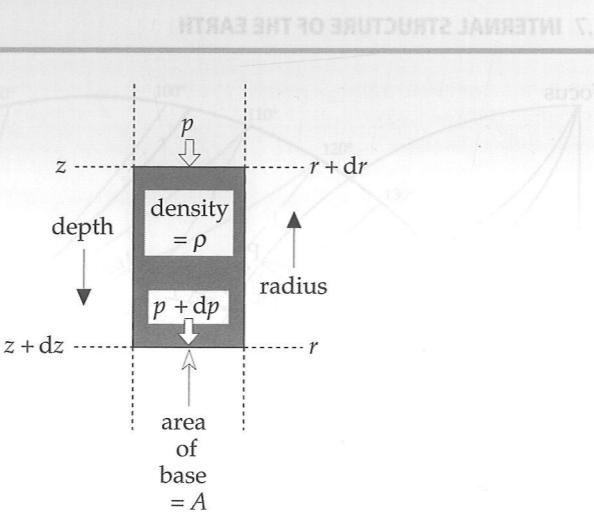
$$dp = \frac{w}{A} = \frac{(\text{volume} \times \rho)g}{A} = \frac{(A dz \rho)g}{A} = \rho g dz = -\rho g dr \quad (3.152)$$

From the definition of bulk modulus,  $K$  (Eq. (3.17)) we can write

$$K = -V \frac{dp}{dV} = \rho \frac{dp}{d\rho} \quad (3.153)$$

Combining Eqs. (3.151), (3.152) and (3.153) gives

$$\frac{-dp}{\rho g dr} = \frac{\rho}{K} = \frac{1}{\Phi} \quad (3.154)$$



**Fig. 3.82** Computation of hydrostatic pressure in the Earth, assuming that a change in pressure  $dp$  with depth increase  $dz$  is due only to the increase in weight of overlying material.

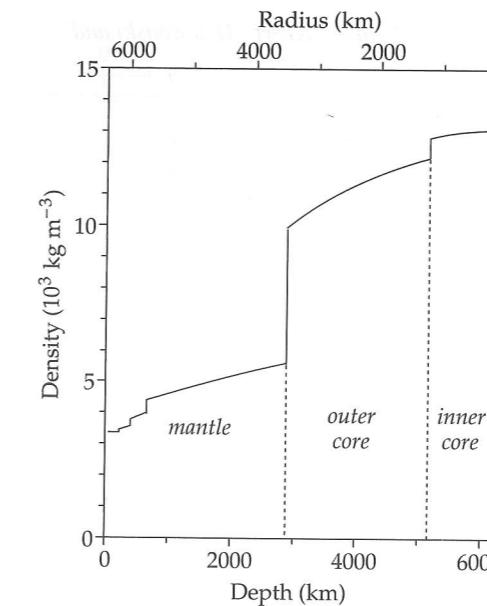
$$\frac{dp}{dr} = -\frac{\rho(r)g(r)}{\Phi(r)} \quad (3.155)$$

Equation (3.155) is known as the Adams–Williamson equation. It was first applied to the estimation of density in the Earth in 1923 by E. D. Williamson and L. H. Adams. It yields the density gradient at radius  $r$ , when the quantities on the right-hand side are known. The seismic parameter  $\Phi$  is known accurately, but the density  $\rho$  is unknown; it is in fact the goal of the calculation. The value of gravity  $g$  used in the equation must be computed separately for radius  $r$ . It is due only to the mass contained within the sphere of radius  $r$ , because external (homogeneous) shells of the Earth do not contribute to gravitation inside them. This mass is the total mass  $E$  of the Earth minus the cumulative mass of all spherical shells external to  $r$ .

The procedure requires that a starting value for  $\rho$  be assumed at a known depth. Analyses of crustal and upper mantle structure in isostatically balanced areas give estimates of upper mantle density around  $3300 \text{ kg m}^{-3}$ . Using this as a starting value at an initial radius  $r_1$  the density gradient in the uppermost mantle can be calculated from Eq. (3.155). Linear extrapolation of this gradient to a chosen greater depth gives the density  $\rho$  at radius  $r_2$ ; the corresponding new value of  $g$  can be calculated by subtracting the mass of the shell between the two depths; together with the value of  $\Phi(r_2)$  the density gradient can be computed at  $r_2$ . This iterative type of calculation gives the variation of density with depth (or radius). Fine steps in the extrapolations give a smooth density distribution (Fig. 3.83). There are two important boundary conditions on the computed density distribution. Integrated over the Earth's radius it must give the correct total mass ( $E = 5.974 \times 10^{24} \text{ kg}$ ). It must also fulfil the relationship  $C = 0.3308ER^2$  between Earth's moment of inertia ( $C$ ), mass ( $E$ ) and radius ( $R$ ), as explained in Section 2.4.3.

The density changes abruptly at the major seismic discontinuities (Fig. 3.83), showing that it is affected principally by changes in composition. If it could be mea-

## 3.7 INTERNAL STRUCTURE OF THE EARTH



**Fig. 3.83** Radial distribution of density within the Earth according to Earth model PREM (data source: Dziewonski, 1989).

sured at normal sea-level pressure and temperature, the density would be found to be around  $4200 \text{ kg m}^{-3}$  in the mantle,  $7600 \text{ kg m}^{-3}$  in the outer core and  $8000 \text{ kg m}^{-3}$  in the inner core. The smooth increase in density between the major compositional discontinuities is the result of the increases in pressure and temperature with depth.

#### 3.7.4.2 Gravity and pressure inside the Earth

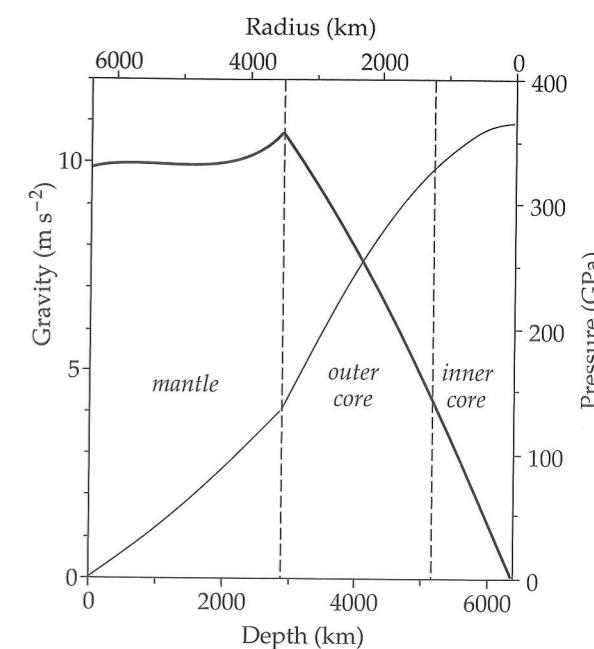
The radial variation of gravity can be computed from the density distribution. As stated above, the value of  $g(r)$  is due only to the mass  $m(r)$  contained within the sphere of radius  $r$ . Let the density at radius  $x$  ( $\leq r$ ) be  $\rho(x)$ ; the gravity at radius  $r$  is then given by

$$g(r) = -G \frac{m(r)}{r^2} = -\frac{G}{r^2} \int_0^r 4\pi x^2 \rho(x) dx \quad (3.156)$$

A remarkable feature of the internal gravity (Fig. 3.84) is that it maintains a value close to  $10 \text{ m s}^{-2}$  throughout the mantle, rising from  $9.8 \text{ m s}^{-2}$  at the surface to  $10.8 \text{ m s}^{-2}$  at the core–mantle boundary. It then decreases almost linearly to zero at the Earth's center.

Hydrostatic pressure is due to the force (N) per unit area ( $\text{m}^2$ ) exerted by overlying material. The SI unit of pressure is the pascal ( $1 \text{ Pa} = 1 \text{ N m}^{-2}$ ). In practice this is a small unit. The high pressures in the Earth are commonly quoted in units of gigapascal ( $1 \text{ GPa} = 10^9 \text{ Pa}$ ), or alternatively in kilobars or megabars ( $1 \text{ bar} = 10^5 \text{ Pa}$ ;  $1 \text{ kbar} = 10^8 \text{ Pa}$ ;  $1 \text{ Mbar} = 10^{11} \text{ Pa} = 100 \text{ GPa}$ ).

Within the Earth the hydrostatic pressure  $p(r)$  at radius  $r$  is due to the weight of the overlying Earth layers between  $r$  and the Earth's surface. It can be computed by integrating Eq. (3.152) using the derived distributions of density and gravity. This gives



**Fig. 3.84** Radial variations of internal gravity (thick curve) and pressure (thin curve) according to Earth model PREM (data source: Dziewonski, 1989).

The pressure increases continuously with increasing depth in the Earth (Fig. 3.84). The rate of increase (pressure gradient) changes at the depths of the major seismic discontinuities. The pressure reaches a value close to  $380 \text{ GPa}$  ( $3.8 \text{ Mbar}$ ) at the center of the Earth, which is about 4 million times atmospheric pressure at sea-level.

#### 3.7.5 Models of the Earth's internal structure

Once the velocity distributions of P- and S-waves inside the Earth were known the broad internal structure of the Earth – crust, mantle, inner and outer core – could be further refined. In 1940–1942 K. E. Bullen developed a model of the internal structure consisting of seven concentric shells. The boundaries between adjacent shells were located at sharp changes in the body-wave velocities or the velocity gradients. For ease of identification the layers were labelled A–G (Table 3.4); this nomenclature has been carried over into more modern models. The seismic layering of the Earth is better known than the composition of the layers, which must be inferred from laboratory experiments and petrological modelling.

In the original Model A the density distribution was not well constrained. Two different density distributions ( $A$  and  $A'$ ) that fitted the known mass and moment of inertia gave disparate central densities of  $17,300 \text{ kg m}^{-3}$  and  $12,300 \text{ kg m}^{-3}$ , respectively. In 1950 Bullen presented Earth Model B, in which the bulk modulus ( $K$ ) and seismic parameter ( $\Phi$ ) were assumed to vary

Table 3.4 Comparison of Earth's internal divisions according to Model A (Bullen, 1942) and PREM (Dziewonski and Anderson, 1981)

Model A Region [km]	PREM		
	Layer	Depth range [km]	Comments
A (0–33)	crust: upper lower	0–15 15–24	lithosphere 0–80 km
B (33–410)	upper mantle: uppermost mantle	24–80	
	low-velocity layer	80–220	asthenosphere
	transition zone	220–400	
C (410–1000)	upper mantle: transition zones	400–670 670–770	
	lower mantle: layer D'	770–2740	
D (1000–2900)	layer D"	2740–2890	
E (2900–4980)	outer core	2890–5150	
F (4980–5120)	transition layer		
G (5120–6370)	inner core	5150–6370	

smoothly with pressure below a depth of 1000 km. The model suggested a central density around 18,000 kg m<sup>-3</sup>.

In the 1950s the development of long-period seismographs made possible the observation of the natural oscillations of the Earth. After very large earthquakes numerous modes of free oscillation are excited with periods up to about one hour (Section 3.3.4). These were first observed unambiguously after the huge Chilean earthquake of 1960. The free oscillations form an independent constraint on Earth models. The lowest-frequency spheroidal modes involve radial displacements that take place against the restoring force of gravitation and are therefore affected by the density distribution. Starting from a spherically symmetric Earth model with known distributions of density and elastic properties, the *forward problem* consists of calculating how such a model will reverberate. The calculated and observed normal modes of oscillation are compared and the model is adjusted until the required fit is obtained. The *inverse problem* consists of computing the model of density and elastic properties by inverting the frequency spectrum of the free oscillations. The parametrized model PREM, based upon the inversion of body-wave, surface-wave and free-oscillation data, is the current standard model of the Earth's internal structure. It predicts a central density of 13,090 kg m<sup>-3</sup>.

### 3.7.5.1 The crust

The Earth's crust corresponds to Bullen's region A. The structures of the crust and upper mantle are complex and show strong lateral variations. This prohibits using the inversion of body-wave travel-times to get a vertical distribution of seismic velocities. The most reliable information on crustal seismic structure comes from seismic refraction profiles and deep crustal reflection sounding. The variation of seismic velocity with depth in the crust differs according to where the profiles are carried out. Ancient continental shield domains have different vertical velocity profiles than younger continental or oceanic domains. In view of this variability any generalized petrological model of crustal structure is bound to be an oversimplification. However, with this reservation, it is still possible to summarize some general features of crustal structure, and the corresponding petrological layering.

A generalized model of the structure of oceanic crust is shown in Fig. 3.85. Oceanic crust is only 5–10 km thick. Under a mean water depth of about 4.5 km the top part of the oceanic crust consists of a layer of sediments that increases in thickness away from the oceanic ridges. The igneous oceanic basement consists of a thin (~0.5 km) upper layer of superposed basaltic lava flows underlain by a complex of basaltic intrusions, the sheeted dike complex. Below this the oceanic crust consists of gabbroic rocks.

## 3.7 INTERNAL STRUCTURE OF THE EARTH

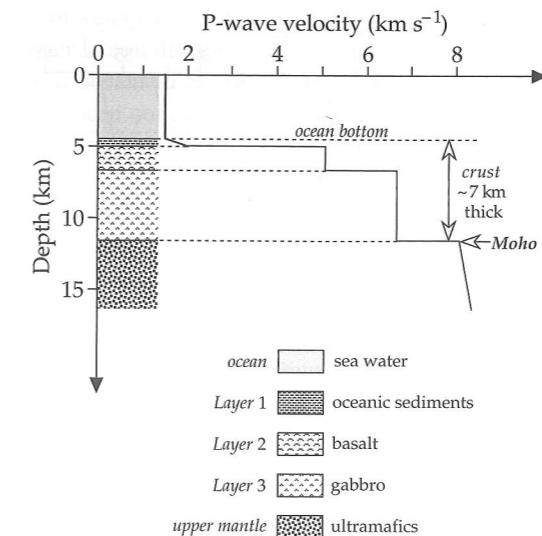


Fig. 3.85 Generalized petrological model and P-wave velocity–depth profile for oceanic crust.

The vertical structure of continental crust is more complicated than that of oceanic crust, and the structure under ancient shield areas differs from that under younger basins. It is more difficult to generalize a representative model (Fig. 3.86). The most striking difference is that the continental crust is much thicker than oceanic crust. Under stable continental areas the crust is 35–40 km thick and under young mountain ranges it is often 50–60 km thick. The continental Moho is not always a sharp boundary. In some places the transition from crust to mantle may be gradual, with a layered structure. Originally, the Conrad seismic discontinuity was believed to separate an upper crustal (granitic) layer from a lower crustal (basaltic) layer. However, the Conrad discontinuity is not found in all places and there is some doubt as to its real nature; it may represent a more complicated compositional or phase boundary. Crustal velocity studies have defined two anomalous zones that often disrupt the otherwise progressive increase of velocity with depth. A low-velocity layer within the middle crust is thought to be due to intruded granitic laccoliths; it is called the sialic low-velocity layer. It is underlain by a middle crustal layer composed of migmatites. Below this layer the velocity rises sharply, forming a “tooth” in the velocity profile. This tooth and the layer beneath it often make up a thinly layered lower crust. Refractions and reflections at the top of the tooth are thought to explain the Conrad discontinuity.

### 3.7.5.2 The upper mantle

In his 1942 model of the Earth's interior (see Table 3.4) Bullen made a distinction between the upper mantle (layers B and C) and the lower mantle (layer D). The upper mantle is characterized by several discontinuities of body-wave velocities and steep velocity gradients (Fig. 3.87). The top of the mantle is defined by the Mohorovičić

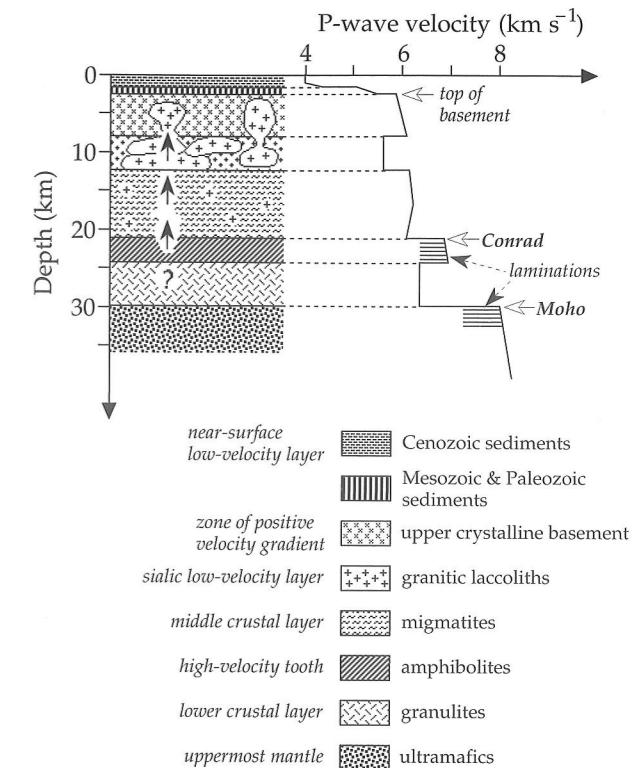
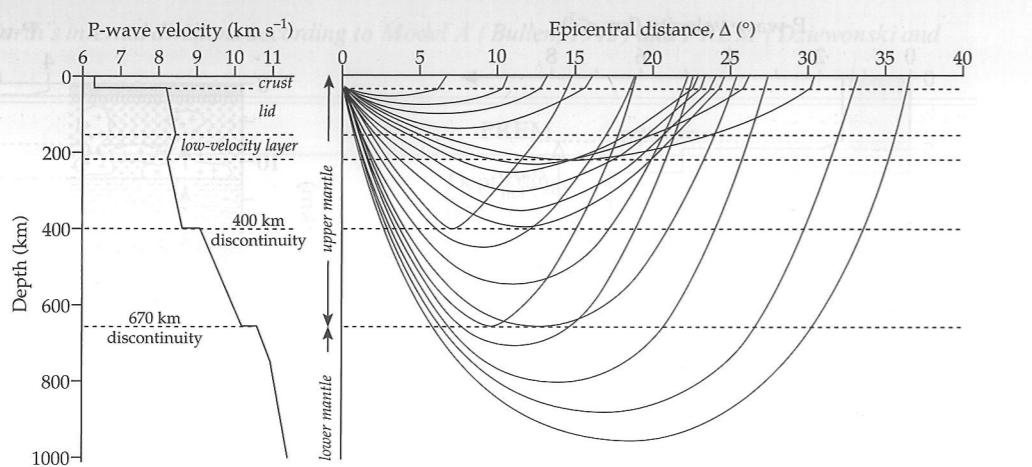


Fig. 3.86 Generalized petrological model and P-wave velocity–depth profile for continental crust (after Mueller, 1977).

discontinuity (Moho), below which the P-wave velocity exceeds 7.6 km s<sup>-1</sup>. The Moho depth is very variable, with a global mean value around 30–40 km. A weighted average of oceanic and continental structures equal to 24.4 km is used in model PREM. The assumption of a spherically symmetric Earth does not hold well for the crust and upper mantle. Lateral differences in structure are important down to depths of at least 400 km. The uppermost mantle between the Moho and a depth of 80–120 km is rigid, with increasing P- and S-wave velocities. This layer is sometimes called the *lid* of the underlying low-velocity layer. Together with the crust, the lid forms the *lithosphere*, the rigid outer shell of the Earth that takes part in plate tectonic processes (see Section 1.2). The lithosphere is subdivided laterally into tectonic plates that may be as large as 10,000 km across (e.g., the Pacific plate) or as small as a few thousand kilometers (e.g., the Philippines plate). The plates are very thin in comparison to their horizontal extent.

An abrupt increase of P- and S-wave velocities by 3–4% has been observed at around 220 ± 30 km depth; it is called the *Lehmann discontinuity*. Like the Conrad discontinuity in the crust it is not found everywhere and its true meaning is in question. Between the lid and the Lehmann discontinuity, in the depth range 100–200 km, body-wave velocity gradients are weakly negative, i.e., the velocities decrease with increasing depth. This layer is called the low-velocity layer (LVL). Its nature cannot be evaluated from body waves, because they do not bottom

**Fig. 3.87** (a) P-wave velocity-depth profile in the upper mantle beneath the Canadian shield, and (b) ray paths through the model; note the sharp bending of rays at the velocity discontinuities at depths of 400 km and 670 km (after LeFevre and Helmberger, 1989).



in the layer and its lower boundary is not sharp. The evidence for understanding the LVL comes from the inversion of surface-wave data. Only long-period surface waves with periods longer than about 200 s can penetrate to the depths of the base of the LVL. The data from surface waves are not precise. Their depth resolution is poor and only the S-wave velocity can be determined. Thus, the top and bottom of the LVL are not sharply defined.

The LVL is usually associated with the *asthenosphere*, which also plays an important role in plate tectonic theory. The decreases in seismic velocities are attributed to reduced rigidity in this layer. Over geological time intervals the mantle reacts like a viscous medium, with a viscosity that depends on temperature and composition.

From the point of view of plate tectonics the asthenosphere is a viscous layer that decouples the lithosphere from the deeper mantle; by allowing slow convection, it permits or promotes the relative motions of the global plates. The LVL and asthenosphere reflect changes of rheological properties of the upper mantle material. Brittle behavior in the crust and lithosphere gives way with increasing depth to ductile behavior in the low-rigidity asthenosphere. The brittle–ductile transition is gradual and depends on properties such as the rock composition, geothermal gradient, initial crustal thickness and strain rate. It probably occurs differently in oceanic and continental lithosphere, as suggested in Fig. 2.69.

The composition of the upper mantle is generally taken to be peridotitic, with olivine  $[(\text{Mg}, \text{Fe})_2\text{SiO}_4]$  as the dominant mineral. With increasing depth the hydrostatic pressure increases and eventually causes high-pressure transformation of the silicate minerals. This is reflected in the seismic properties. Travel-time– $\Delta$  curves of body waves show a distinct change in slope at epicentral distances of about 20°. This is attributed to a discontinuity in mantle velocities at a depth of around 400 km (Fig. 3.87). The 400 km discontinuity (or 20° discontinuity) is interpreted as due to a petrological change from an olivine-type lattice to a more closely packed spinel-type lattice.

The most interesting aspect of D'' is the presence – revealed by seismic tomographic imaging (Section 3.7.6.2)

– in the layer and its lower boundary is not sharp. The evidence for understanding the LVL comes from the inversion of surface-wave data. Only long-period surface waves with periods longer than about 200 s can penetrate to the depths of the base of the LVL. The data from surface waves are not precise. Their depth resolution is poor and only the S-wave velocity can be determined. Thus, the top and bottom of the LVL are not sharply defined.

The LVL is usually associated with the *asthenosphere*, which also plays an important role in plate tectonic theory. The decreases in seismic velocities are attributed to reduced rigidity in this layer. Over geological time intervals the mantle reacts like a viscous medium, with a viscosity that depends on temperature and composition.

From the point of view of plate tectonics the asthenosphere is a viscous layer that decouples the lithosphere from the deeper mantle; by allowing slow convection, it permits or promotes the relative motions of the global plates. The LVL and asthenosphere reflect changes of rheological properties of the upper mantle material. Brittle behavior in the crust and lithosphere gives way with increasing depth to ductile behavior in the low-rigidity asthenosphere. The brittle–ductile transition is gradual and depends on properties such as the rock composition, geothermal gradient, initial crustal thickness and strain rate. It probably occurs differently in oceanic and continental lithosphere, as suggested in Fig. 2.69.

The composition of the upper mantle is generally taken to be peridotitic, with olivine  $[(\text{Mg}, \text{Fe})_2\text{SiO}_4]$  as the dominant mineral. With increasing depth the hydrostatic pressure increases and eventually causes high-pressure transformation of the silicate minerals. This is reflected in the seismic properties. Travel-time– $\Delta$  curves of body waves show a distinct change in slope at epicentral distances of about 20°. This is attributed to a discontinuity in mantle velocities at a depth of around 400 km (Fig. 3.87). The 400 km discontinuity (or 20° discontinuity) is interpreted as due to a petrological change from an olivine-type lattice to a more closely packed spinel-type lattice.

The most interesting aspect of D'' is the presence –

### 3.7 INTERNAL STRUCTURE OF THE EARTH

of velocity variations of several percent that take place over lateral distances comparable in size to the continents and oceans in Earth's crust. The term "grand structure" has been coined for these regions; the thicker parts have also been termed "crypto-continent" and the thinner parts "crypto-oceans" (see Fig. 4.38). Moreover, the seismically fast regions (in which temperatures are presumed to be cooler than normal) lie beneath present subduction zones. This suggests that cold subducted lithosphere may eventually sink to the bottom of the mantle where it is colder and more rigid than the surrounding mantle and hence has higher body-wave velocities. A large low-velocity (hot) region underlies the Pacific basin, in which many centers of volcanism and locally high heat flow ("hotspots," see Section 1.2.8) are located. The D'' layer is suspected of being the source of the mantle plumes that cause these anomalies. Exceptionally hot material from D'' rises in thin pipe-like mantle plumes to the 670 km discontinuity, which opposes further upward motion in the same way that it resists the deeper subduction of cold lithospheric slabs. Occasionally a hot plume is able to break through the 670 km barrier, producing a surface hotspot.

If our current understanding of the D'' layer is correct, it plays an important role in geodynamic and geothermal behavior. On the one hand, D'' serves as the source of material for the mantle plumes that give rise to hotspots, which are important in plate tectonics. On the other hand, the thermal properties of D'' could influence the outward transport of heat from the Earth's core; in turn, this could affect the intricate processes that generate the Earth's magnetic field.

#### 3.7.5.3 The lower mantle

The lower mantle is now classified as the part below the important seismic discontinuity at 670 km. Its composition is rather poorly known, but it is thought to consist of oxides of iron and magnesium as well as iron–magnesium silicates with a perovskite structure. The uppermost part of the lower mantle between 670 and 770 km depth has a high positive velocity gradient and corresponds to the lower part of Bullen's layer C. Beneath it lies Bullen's layer D', which represents a great thickness of normal mantle, characterized by smooth velocity gradients and the absence of seismic discontinuities.

Just above the core–mantle boundary an anomalous layer, approximately 150–200 km thick, has been identified in which body-wave velocity gradients are very small and may even be negative. Although part of the lower mantle, it evidently serves as a boundary layer between the mantle and core. It is labelled D'' to distinguish it from the normal mantle above it. The structure and role of the D'' layer are not yet known with confidence, but it is the focus of intensive current research. Models of the internal structure of D'' have been proposed with positive velocity gradients, others with negative velocity gradients, and some with small velocity discontinuities. The latter possibility is important because it would imply some stratification within D''.

The most interesting aspect of D'' is the presence –

now been discarded. Improved seismographic resolution has yielded a large quantity of high-quality data for reflections from the boundaries of the outer core (PcP, ScS) and the inner core (PKiKP) which have helped clear up the nature of these boundaries. The PKiKP phase contains high frequencies; this implies that the inner core boundary is sharp, probably no more than 5 km thick. The seismic events earlier interpreted as due to a layer F are now regarded as rays that have been scattered by small-scale features at the bottom of the mantle.

The inner core transmits P-waves (PKiKP phase) but S-waves in the inner core (PKJKP phase), although in principle possible, have not yet been observed unequivocally. Body-wave travel-times do not constrain the rigidity of the inner core. The amplitude spectrum of the frequencies of higher modes of the Earth's free oscillations show that the inner core is likely solid. However, it is possible that it is not completely solid. Rather, it may be a mixture of solid and liquid phases at a temperature close to the solidification temperature. An analogy can be made with the mushy, semi-frozen state that water passes through on freezing to ice.

The outer core is fluid, with a viscosity similar to that of water. It is assumed to be homogeneous and its thermal state is supposed to be adiabatic. These are the conditions to be expected in a fluid that has been well mixed, in the Earth's case by convection and differential rotation. One theory of core dynamics holds that the iron-rich inner core is solidifying from the fluid outer core, leaving behind its lighter elements. These constitute a less-dense, therefore gravitationally buoyant, fluid, which rises through the denser overlying liquid. This compositional type of buoyancy could be an important contributor to convection in the outer core, and therefore to the dynamics of the core and generation of the Earth's magnetic field.

The core–mantle boundary (CMB) is also called the *Gutenberg discontinuity*. It is characterized by very large changes in body-wave velocities and is the most sharply defined seismic discontinuity. Seismic data show that the boundary is not smooth but has a topography of hills and valleys. Anomalies in the travel-times of PKKP phases – which are reflected once internally at the CMB – have been attributed to scattering by topographic features with a relief of a few hundred meters. However, depending on conditions in the hot D'' layer of the lower mantle immediately above the CMB some topographic features may be up to 10 km high. Interference between the CMB topography and fluid motions in the outermost core may couple the core and mantle to each other dynamically.

#### 3.7.6 Seismic tomography

A free translation of the term tomography is "representation in cross-section." Neighboring two-dimensional

cross-sections can be combined to give a three-dimensional model. The use of computer-aided tomography (CAT) in medical diagnosis is well known as a non-invasive method of examining internal organs for abnormal regions. X-rays or ultrasonic rays are absorbed unequally by different materials. CAT consists of studying the attenuation of x-rays or ultrasonic waves that pass through the body in distinctly controlled planar sections. Seismic tomography uses the same principles, with the difference that the travel-times of the signals, as well as their attenuation, are observed. Hence the technique may be described as the three-dimensional modelling of the velocity or attenuation distribution of seismic waves in the Earth. The technique requires powerful computational facilities and sophisticated programming.

The travel-time of a seismic wave from an earthquake focus to a seismograph is determined by the velocity distribution along its path. For an idealized, spherically symmetric Earth model the radial distributions of velocity are known. The velocities in the model are mean values, which average out lateral fluctuations. If such a velocity model were used to compute travel-times of different phases to any epicentral distance, a set of curves indistinguishable from Fig. 3.77 would result. In reality, the observed travel-times usually show small deviations from the calculated times. These discrepancies are called *travel-time residuals* or *anomalies*, and they can have several causes. An obvious cause is that the focal depth of an earthquake is not zero, as assumed in Fig. 3.77, but may be up to several hundred kilometers. The parametrized Earth model *iasp91* takes this into account and tabulates travel-times for several focal depths. Clearly, precise determination of earthquake focal parameters (epicentral location, depth and time of occurrence) are essential prerequisites for seismic tomography. An important factor is the assumption of spherical symmetry, which is not perfectly valid, and so the ellipticity of the Earth's figure must be taken into account.

A local source of travel-time residuals is the particular velocity-depth distribution under an observational network. Ideally, the local vertical velocity profile should be known, so as to allow compensation of an observed travel-time for local anomalous structure. In practice, the signals from a selected earthquake are averaged for several stations in a given area of the surface (e.g., about  $3^\circ$  square) to reduce local perturbations.

The lateral variations in P- and S-wave velocity at any given depth may amount to a few percent of the "average" velocity for that depth assumed in the reference model. If, at a certain depth, a seismic ray passes through a region in which the velocity is slightly faster than average, the wave will arrive slightly sooner than expected at the receiver; if the anomalous velocity is slower than average, the wave will arrive late. This permits classification of travel-time as "early" or "late" depending on

whether a ray has traversed a region that is "fast" or "slow" with respect to the assumed model.

The velocity of a seismic wave is determined by elastic parameters and density, which are affected by temperature. Thus, the velocity anomalies obtained from seismic tomography on a global scale are generally interpreted in terms of abnormal temperature and rigidity. A "slow" region is commonly associated with above-average temperature and lower rigidity, while a "fast" region is attributed to lower temperature and higher rigidity.

### 3.7.6.1 Travel-time residuals and velocity anomalies

The velocity distribution in the Earth is more heterogeneous than in a standard velocity-model. The time ( $t$ ) taken for a seismic ray to travel along the path from an earthquake to a recording station thus differs from the value ( $t_0$ ) predicted by the model. The difference ( $\Delta t = t_0 - t$ ) is a travel-time anomaly. A *travel-time residual* ( $\Delta t/t_0$ ) for the path is obtained by expressing the anomaly as a percentage of the expected travel-time (Box 3.4). If a given path is subdivided into segments with different velocities, the travel-time is equivalent to the sum of the travel-times through the individual segments. This is like an equation with one known value ( $\Delta t/t_0$ ) and several unknown terms (the velocity anomalies in each segment). For a given earthquake several seismic stations may record rays that traverse the region of interest. Each station may also have recorded rays that crossed the region from different earthquakes. A large set of travel-time residuals results, equivalent to a large number of equations. A mathematical procedure called matrix inversion is used to solve the set of equations and obtain the *velocity anomalies*. The sophisticated analysis involves intensive data processing, and is beyond the scope of this text. However, Box 3.5 illustrates for a simple pattern of travel-time residuals (Box 3.4, Fig. B3.4) how a distribution of velocity anomalies may be deduced by *back projection*. In this method successive adjustments are made to an initial velocity distribution to account for the observed travel-time residuals.

Tomographic imaging of the Earth's three-dimensional velocity distribution is able to resolve small differences in seismic velocities. In some studies the signals are of local origin, generated by earthquakes or explosions within or near the volume of interest. Other studies are based on *teleseismic* signals, which originated in earthquakes at more than  $20^\circ$  epicentral distance. The analysis requires precise location of the earthquake source, corrections for local effects (such as crustal structure close to each measurement station), and reconstruction of the seismic rays between source and seismometer.

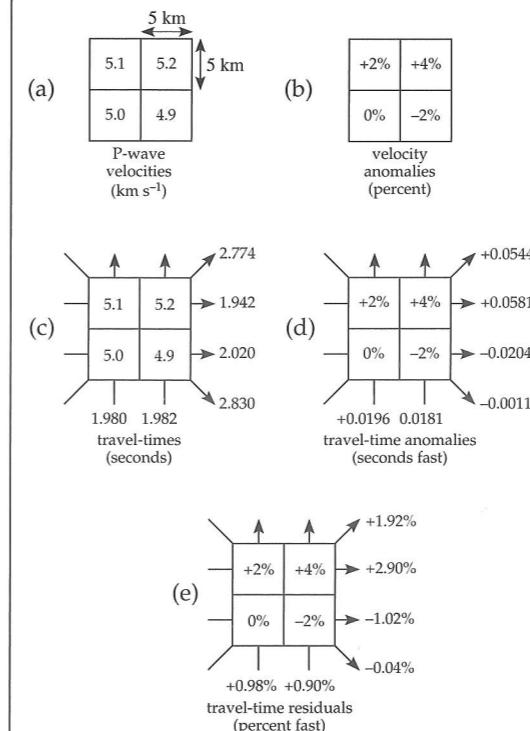
Seismic tomography can be based on either body waves or surface waves. The database for P-wave studies consists of hundreds of thousands of first arrival times from earthquakes that occurred in the past 30 years, which

## 3.7 INTERNAL STRUCTURE OF THE EARTH

### Box 3.4: Calculation of travel-time residuals

Consider the passage of seismic rays in six directions through a square region containing four equal areas, each a square with edge 5 km, in which the P-wave velocities ( $V$ ) are  $4.9 \text{ km s}^{-1}$ ,  $5.0 \text{ km s}^{-1}$ ,  $5.1 \text{ km s}^{-1}$  and  $5.2 \text{ km s}^{-1}$  (Fig. B3.4). Let the expected velocity ( $V_0$ ) throughout the region be  $5.0 \text{ km s}^{-1}$ . The velocity difference ( $\Delta V$ ) in each area is found by subtracting the reference value, and the velocity anomaly ( $\Delta V/V_0$ ) is obtained by expressing the difference as a percentage of the expected value. This gives zones that are 2% and 4% fast, a zone that is 2% slow and a zone with no anomaly.

Suppose that six seismic rays traverse the square region as in Fig. B3.4. If the velocity were  $5.0 \text{ km s}^{-1}$  in each area, the expected travel-time ( $t_0$ ) would be  $2.0 \text{ s}$  for each of the horizontal and vertical rays and  $2\sqrt{2} = 2.828 \text{ s}$  for the longer diagonal rays. However, the real velocity is different in each area. As a result, some of the observed travel-times are shorter and some are longer than the expected value. A travel-time anomaly ( $\Delta t$ ) is computed by subtracting the observed travel-times ( $t$ ) from the expected value (Fig. B3.4d). The *travel-time residual* ( $\Delta t/t_0$ ) is obtained by expressing the anomaly as a percentage of the expected travel-time (Fig. B3.4e). Note that the residuals are not simple averages of the velocity anomalies along each path.



**Fig. B3.4** Computation of relative travel-time residuals (in percent) for a simple four-block structure (modified after Kissling, 1993).

have been catalogued by the International Seismological Center. In addition, S-wave models have been derived for data sets from digital broadband seismic stations. The velocity structures of both P-waves and S-waves have been obtained for global models of the entire mantle as well as for regional studies (e.g., individual subduction zones).

### 3.7.6.2 Mantle tomography

The inversion of body-wave data provides evidence for the lateral variations in velocity in the deeper interior. The lateral variations at a given depth are equivalent to the variations on the surface of a sphere at that depth, and so can be depicted with the aid of spherical harmonic functions. The contoured velocity anomalies obtained by spherical harmonic analysis for a depth of 2500 km in the lower mantle are dominated by a ring of fast P-wave velocities around the Pacific and a slow-velocity region in the center (Fig. 3.88). Slow velocities are also present at depth under Africa. The pattern is present at nearly all depths in the lower mantle (i.e., below 1000 km). This is shown clearly by a vertical cross-section along a profile around the equator (Fig. 3.89, *bottom frame*). The slow velocities are interpreted as the expression of warm mantle material that may be rising from the core–mantle boundary as so-called "super-plumes." The role of these super-plumes in the circulation pattern of mantle convection is not yet understood. Deep fast-velocity ("cold") regions under America and Indonesia and slow-velocity ("warm") regions under the Pacific and Atlantic oceans extend from about 1000 km depth to the core–mantle boundary. The lower mantle velocity anomalies show little correlation with plate tectonic elements, which are features of the lithosphere.

Velocity anomalies in the upper mantle, on the other hand, are clearly related to plate tectonic features. The upper mantle velocity structures can be modelled by inverting teleseismic body waves, but these depths can also be probed by long-period surface waves, which are more sensitive to variations in rigidity (and thus temperature). The inversion of long-period surface-wave data is an effective technique for modelling upper mantle S-wave velocities, especially under oceanic areas, where the resolution of body-wave anomalies is poor. The pattern of S-wave velocity anomalies in the upper mantle along the equatorial cross-section (Fig. 3.89, *middle frame*) shows generally elevated velocities under the "cold" continents and reduced velocities under the "warm" oceanic ridge systems. In these analyses the large local variations in S-wave velocities in the heterogeneous crust constitute an impediment to greater resolution.

Important advances in understanding the geodynamical state of the Earth's mantle have been made as a result of numerous regional seismic tomography studies, which have focused on specific tectonic environments such as zones of tectonic plate convergence. In particular,

**Box 3.5: Calculation of velocity anomalies**

To illustrate how a velocity structure can be deduced from observed travel-time residuals we take as starting point the set of travel-time residuals ( $\Delta t/t_0$ ) computed for the velocity distribution in Box 3.4. The expected travel-time along a seismic path of length  $L$  is  $t_0 = L/V_0$ , where  $V_0$  is the reference velocity. If the true velocity is  $V = V_0 + \Delta V$ , the observed travel-time is

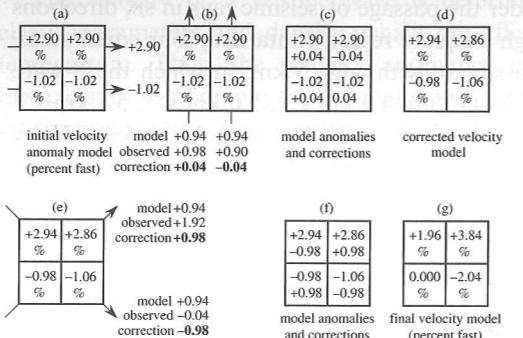
$$t = \frac{L}{V_0 + \Delta V} = \frac{L}{V_0} \left( 1 + \frac{\Delta V}{V_0} \right)^{-1}$$

$$= t_0 \left( 1 - \frac{\Delta V}{V_0} + \left( \frac{\Delta V}{V_0} \right)^2 + \dots \right) \quad (1)$$

$$\left( \frac{\Delta t}{t_0} \right) = \left( \frac{\Delta V}{V_0} \right) \left( 1 - \frac{\Delta V}{V_0} + \dots \right) \quad (2)$$

Next, we assume that the percentage velocity anomalies ( $\Delta V/V_0$ ) are the same as the percentage travel-time residuals. As shown by Eq. (2), this is not strictly true; the discrepancy is of the order of  $(\Delta V/V_0)$ . However, the velocity anomalies are usually very small (see Figs. 3.88–3.90), so assuming that the velocity anomaly is equal to the travel-time residual is a reasonable approximation.

Consider a horizontal ray that traverses the upper two “fast” blocks in Fig. B3.5. To account for the (early) travel-time anomaly of + 2.9% let each upper block be allocated a (fast) velocity anomaly of + 2.9%. Similarly, let each of the bottom two blocks be allocated velocity-anomalies of -1.02%. This simple velocity distribution (Fig. B3.5a) satisfies the travel-time anomalies for the horizontal rays. However, in the vertical direction it gives travel-time anomalies of + 0.94% (the mean of



**Fig. B3.5** Backward projection of relative travel-time residuals to obtain velocity anomalies (modified after Kissling, 1993).

+ 2.90% and -1.02%) for each ray (Fig. B3.5b). This does not agree with the observed anomalies for the two vertical rays (+ 0.98% and + 0.90%, respectively); one vertical anomaly is 0.04% too large, the other is 0.04% too small. The velocities in the blocks are adjusted accordingly by making a correction of + 0.04% to the left-hand blocks and -0.04% to the right-hand blocks (Fig. B3.5c). This gives a new distribution of velocity anomalies which satisfies the horizontal and vertical rays (Fig. B3.5d).

The model now gives travel-time anomalies along each of the diagonal rays of + 0.94%, compared to observed anomalies of + 1.92% and -0.04%, respectively (Fig. B3.5e). Further corrections of + 0.98% are now made to the upper right and lower left blocks, and -0.98% to the lower right and upper left blocks (Fig. B3.5f). The resulting distribution of velocity anomalies (Fig. B3.5g) satisfies all six rays through the anomalous region, and is close to the original distribution of velocity anomalies (Fig. B3.4b).

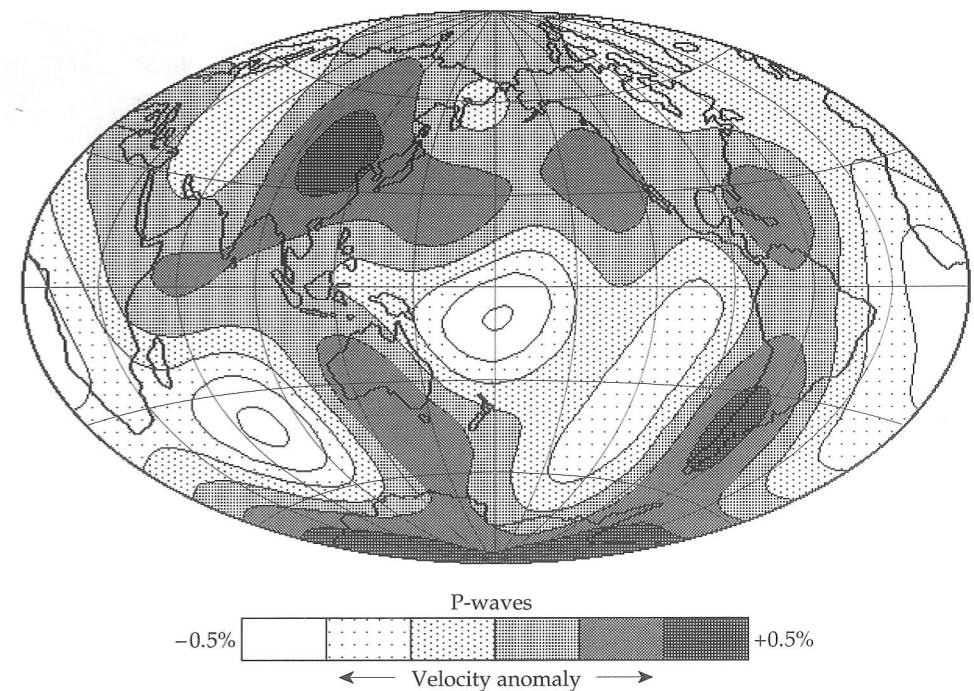
investigations of subduction zones have provided new insights into the effects on mantle convection of the seismic discontinuities at 400 km and 670 km depth. As discussed in more depth in Section 4.2.9.3, two models of mantle convection are prevalent. They differ mainly in the roles played by these seismic discontinuities. In the first – whole mantle convection – the entire mantle participates in the convection process. In the second – layered convection – the seismic discontinuities bound a transition zone that separates a system of convection cells in the upper mantle from a system of convection cells in the lower mantle.

Seismic tomography is a potent technique for describing conditions in the mantle. Many questions, such as the nature of the “super-plume” structures beneath the Pacific and Africa, remain unanswered. Despite convective mixing there may be more heterogeneity in mantle composition than is often supposed. Indeed, the assumption that fast and slow seismic velocities imply cold and hot temperatures, respectively, may be overly simplistic. Yet, seismic tomography is the best tool currently

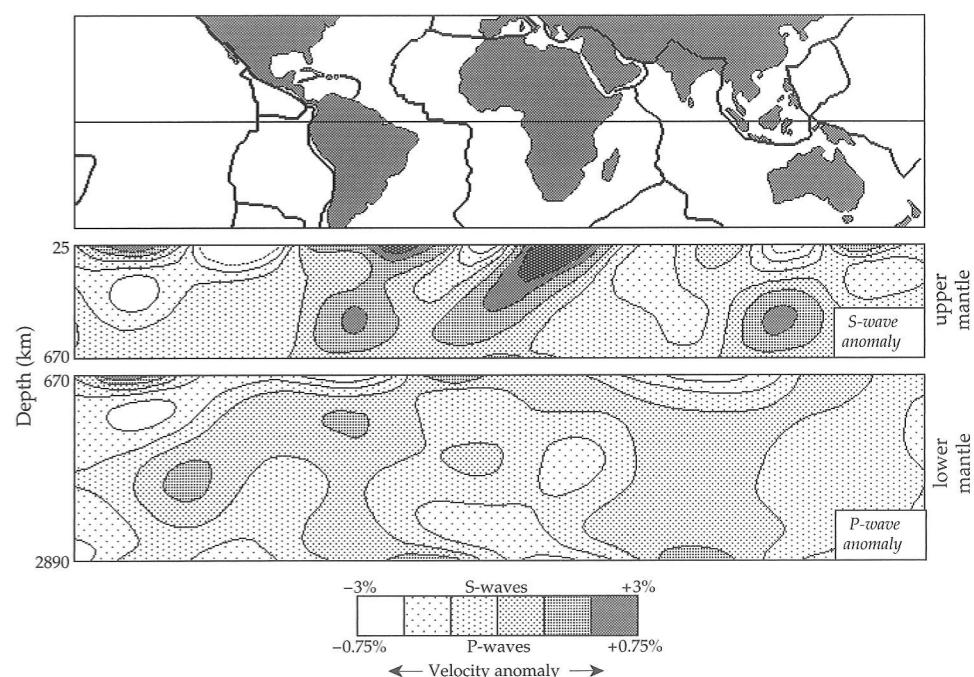
available for investigating geodynamic processes in the Earth’s interior.

**3.8 SUGGESTIONS FOR FURTHER READING**

**Fig. 3.88** Map of P-wave velocity anomalies in the lower mantle (2500 km depth). The deviations are plotted as percent faster or slower than the reference velocity at this depth (after Dziewonski, 1984, 1989).



**Fig. 3.89** Seismic tomographic section through the mantle along an equatorial profile. Middle frame: S-wave anomalies in the upper mantle to a depth of 670 km. Bottom frame: P-wave anomalies in the lower mantle in depths between 670 and 2890 km (after Woodhouse and Dziewonski, 1984).



Kearey, P., Brooks, M. and Hill, I. 2002. *An Introduction to Geophysical Exploration*, 3rd edn, Oxford: Blackwell Publishing.

Mussett, A. E. and Khan, M. A. 2000. *Looking into the Earth: An Introduction to Geological Geophysics*, Cambridge: Cambridge University Press.

Parasnis, D. S. 1997. *Principles of Applied Geophysics*, 5th edn, London: Chapman and Hall.

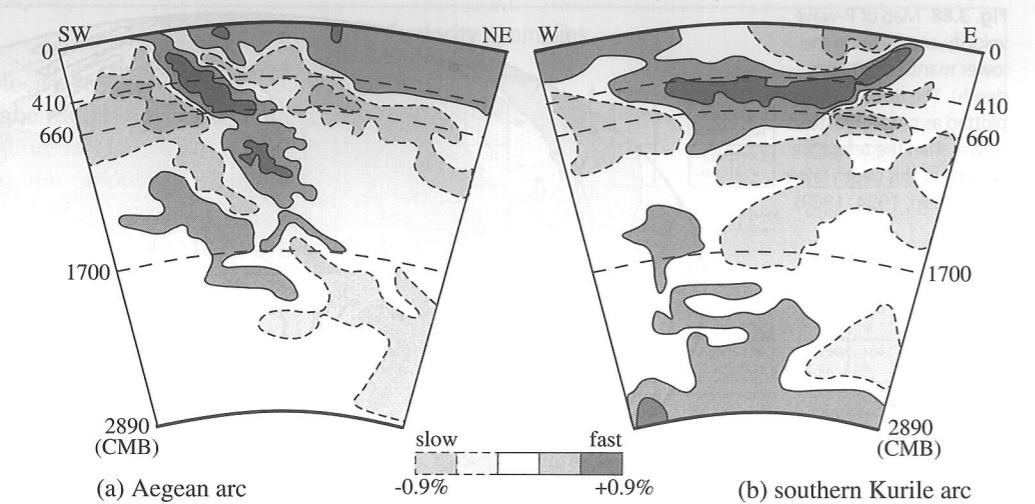
Sharma, P. V. 1997. *Environmental and Engineering Geophysics*, Cambridge: Cambridge University Press.

Walker, B. S. 1982. *Earthquake*, Alexandria, VA: Time-Life Books.

**3.8 SUGGESTIONS FOR FURTHER READING***Introductory level*

- Bolt, B. A. 1993. *Earthquakes*, New York: W. H. Freeman.  
Bryant, E. 2001. *Tsunami: The Underrated Hazard*, Cambridge: Cambridge University Press.

**Fig. 3.90** Seismic tomographic sections showing different styles of subduction. (a) At the Aegean arc the northward subducting African plate sinks through the 410 km and 660 m seismic discontinuities, while (b) at the southern Kurile arc the westward subducting Pacific plate appears to be unable to penetrate into the lower mantle and deflects horizontally (after Káráson and Van der Hilst, 2000).



#### Intermediate level

Dobrin, M. B. and Savit, C. H. 1988. *Introduction to Geophysical Prospecting*, 4th edn, New York: McGraw-Hill.

Fowler, C. M. R. 2004. *The Solid Earth: An Introduction to Global Geophysics*, 2nd edn, Cambridge: Cambridge University Press.

Gubbins, D. 1990. *Seismology and Plate Tectonics*, Cambridge: Cambridge University Press.

Lay, T. and Wallace, T. C. 1995. *Modern Global Seismology*, San Diego, CA: Academic Press.

Lillie, R. J. 1999. *Whole Earth Geophysics: An Introductory Textbook for Geologists and Geophysicists*, Englewood Cliffs, NJ: Prentice Hall.

Science: Special Section. 2005. The Sumatra–Andaman Earthquake. *Science*, **308**, 1125–1146.

Shearer, P. 1999. *Introduction to Seismology*, Cambridge: Cambridge University Press.

Sleep, N. H. and Fujita, K. 1997. *Principles of Geophysics*, Oxford: Blackwell Science.

Stein, S. and Wysession, M. 2003. *An Introduction to Seismology, Earthquakes and Earth Structure*, Oxford: Blackwell Publishing.

Telford, W. M., Geldart, L. P. and Sheriff, R. E. 1990. *Applied Geophysics*, Cambridge: Cambridge University Press.

#### Advanced level

Aki, K. and Richards, P. G. 1980. *Quantitative Seismology: Theory and Methods*, San Francisco, CA: W. H. Freeman.

Iyer, H. M. and Hirahara, K. (eds) 1993. *Seismic Tomography: Theory and Practice*, London: Chapman and Hall.

Nolet, G. (ed) 1987. *Seismic Tomography: With Applications in Global Seismology and Exploration Geophysics*, New York: Springer.

Officer, C. B. 1974. *Introduction to Theoretical Geophysics*, New York: Springer.

Sheriff, R. E. and Geldart, L. P. 1995. *Exploration Seismology*, 2nd edn, Cambridge: Cambridge University Press.

Stacey, F. D. 1992. *Physics of the Earth*, 3rd edn, Brisbane: Brookfield Press.

Udias, A. 2000. *Principles of Seismology*, Cambridge: Cambridge University Press.

#### 3.9 REVIEW QUESTIONS

- Describe the principle of the seismometer.
- Describe the particle motions relative to the direction of propagation of the two seismic body waves and the two seismic surface waves.
- What are the Lamé constants in seismic theory?
- What is meant by dispersion with regard to the propagation of surface waves?
- Sketch how the train of Rayleigh waves from a distant earthquake would appear on a seismogram.
- What are the free oscillations of the Earth? What kinds are possible? What is meant by the normal modes? What are higher modes? How do they relate to surface waves?
- How does the elastic rebound model explain the origin of a tectonic earthquake?
- What is the epicenter of an earthquake? Explain how the epicenter of an earthquake can be located? What is the minimum number of seismic records needed?
- Explain the difference between earthquake intensity and magnitude.
- If the magnitude of an earthquake is 0.5 greater than that of another, how much greater is the amount of energy it releases?
- How does a tsunami originate? Why are tsunamis barely noticeable over the open ocean but very dangerous near shore?
- Describe the geographical distribution of the Earth's seismically active zones.
- Describe with the aid of sketches the distribution of earthquakes at the three major plate boundaries:

#### 3.10 EXERCISES

- (a) spreading ridge, (b) transform fault, and (c) subduction zone.
- Sketch the fault-plane solutions that characterize earthquakes at each type of plate boundary. Describe their significance with regard to plate tectonic motions.
  - How do the distribution and focal solutions of earthquakes at a transform fault differ from those at a strike-slip fault?
  - What is the Mohorovičić discontinuity? What is the seismic evidence for this feature? What is its average depth under continents, under oceans, and for the entire world?
  - How do seismic wave velocities change at the major discontinuities in the Earth's internal structure? How are these discontinuities characterized?
  - What is the brittle–ductile transition in the Earth? What physical properties determine the depth and nature of this transition?
  - On a cross-section of the Earth, sketch the paths of the following seismic rays: (i) PKP, (ii) SKS, (iii) Pcp, (iv) PPP.
  - What is a PKIKP wave? Describe the refraction of this wave at each discontinuity it crosses.
  - What is the critical distance in seismic refraction surveying? What is the crossover distance? What is a head wave? What are supercritical reflections?
  - A seismic survey is conducted over level ground consisting of two horizontal layers. Sketch a travel-time diagram that shows the arrivals of the direct wave, the reflected wave, and the doubly refracted wave. How can the seismic velocity of each layer be determined from the diagram?
  - What is normal move-out in seismic reflection profiling?
  - What is meant by the migration of reflection seismic records? Why is it necessary?
  - Describe the split-spread and common-mid-point methods in reflection profiling? What is achieved by each of these methods?
- (a) From these quantities calculate the rigidity modulus,  $\mu$ , bulk modulus,  $K$ , and Poisson's ratio,  $\nu$ , at each depth.
- (b) Discuss in your own words the information that these data give about the deep interior of the Earth.
- A strong earthquake off the coast of Japan sets off a tsunami that propagates across the Pacific Ocean (average depth  $d = 5$  km).
    - Calculate the velocity of the wave in  $\text{km hr}^{-1}$  and the corresponding wavelength, when the wave has a dominant period of 30 min.
    - How long does the wave take to reach Hawaii, which is at an angular distance of  $54^\circ$  from the epicenter?
  - The dispersion relation between frequency  $\omega$  and wave number  $k$  of seismic water waves for water depth  $d$  is (Box 3.3)
 
$$\omega^2 = g \operatorname{tanh}(kd)$$
    - Modify this expression for wavelengths that are much shorter than the water depth.
    - Determine the phase velocity of these waves.
    - Show that the group velocity of the waves is half the phase velocity.
  - In a two-layer Earth the mantle and core are each homogeneous and the radius of the core is one-half the radius of the Earth. Derive a formula for the travel-time curve for the arrival time  $t$  of the phase Pcp at epicentral distance  $\Delta$ . Verify the formula for the maximum possible value of  $\Delta$  in this model.
  - Why might one expect an interface with a small critical angle to be a good reflector of seismic energy?
  - The P-wave from an earthquake arrives at a seismograph station at 10:20 a.m. and the S-wave arrives at

Region	Depth [km]	$\alpha$ [ $\text{km s}^{-1}$ ]	$\beta$ [ $\text{km s}^{-1}$ ]	$\rho$ [ $\text{kg m}^{-3}$ ]
Lower crust	33	7.4	4.8	3100
Upper mantle	400	8.5	12.2	3900
Lower mantle	2200	12.2	7.0	5300

10:25 a.m. Assuming that the P-wave velocity is 5 km s<sup>-1</sup> and that Poisson's ratio is 0.25, compute the time at which the earthquake occurred and its epicentral distance in degrees from the seismograph station.

8. The following table gives arrival times of P-waves ( $t_p$ ) and S-waves ( $t_s$ ) from a nearby earthquake:

Recording Station	Time of day [hr:min]	$t_p$ [s]	$t_s$ [s]
A	23:36	54.65	57.90
B	23:36	57.34	62.15
C	23:37	00.49	07.55
D	23:37	01.80	10.00
E	23:37	01.90	10.10
F	23:37	02.25	10.70
G	23:37	03.10	12.00
H	23:37	03.50	12.80
I	23:37	06.08	18.30
J	23:37	07.07	19.79
K	23:37	08.32	21.40
L	23:37	11.12	26.40
M	23:37	11.50	26.20
N	23:37	17.80	37.70

- (a) Plot the arrival-time differences ( $t_s - t_p$ ) against the arrival times of the P-wave to produce a Wadati diagram.  
(b) Determine the ratio  $\alpha/\beta$  of the seismic velocities.  
(c) Determine the time of occurrence ( $t_0$ ) of the earthquake.
9. A plane seismic wave, travelling vertically downwards in a rock of density 2200 kg m<sup>-3</sup> with seismic velocity 2000 m s<sup>-1</sup>, is incident on the horizontal top surface of a rock layer of density 2400 kg m<sup>-3</sup> and seismic velocity 3300 m s<sup>-1</sup>.  
(a) What are the amplitude ratios of the transmitted and reflected waves?  
(b) What fraction of the energy of the incident wave is transmitted into the lower medium?
10. A plane seismic wave travels vertically downwards at a velocity of 4800 m s<sup>-1</sup> through a salt layer with density 2100 kg m<sup>-3</sup>. The wave is incident upon the top surface of a sandstone layer with density 2400 kg m<sup>-3</sup>. The phase of the reflected wave is changed by 180° and the reflected amplitude is 2% of the incident amplitude. What is the seismic velocity of the sandstone?

11. (a) Calculate the minimum arrival times for seismic reflections from each of the reflecting interfaces in the following section. Consider the base of the lowermost bed to be a reflector as well.  
(b) What is the average velocity of the section for a reflection from the base of the dolomite?  
(c) Using the listed densities calculate the reflection coefficient for each interface (except the base of the dolomite). Which interface gives the strongest

Formation	Density [kg m <sup>-3</sup> ]	Thickness [m]	Formation velocity [ms <sup>-1</sup> ]
Alluvium	1500	150	600
Shale	2400	450	2700
Sandstone	2300	600	3000
Limestone	2500	900	5400
Salt	2200	300	4500
Dolomite	2700	600	6000

reflection and which the weakest? At which interfaces does a change in phase occur? What does this mean?

12. A reflection seismic record in an area of relatively flat dips gave the following data:

Distance shot-point to detector [m]	Travel-time ( $t_1$ ), 1st reflection [s]	Travel-time ( $t_2$ ), 2nd reflection [s]
30	1.000	1.200
90	1.002	1.201
150	1.003	1.201
210	1.007	1.202
270	1.011	1.203
330	1.017	1.205
390	1.023	1.207

- (a) Plot the  $t-x$  curves for these reflections to show the "moveout" effect.  
(b) On a different graph, plot the  $t^2-x^2$  curves (i.e., squared data) for the reflections.  
(c) Determine the average vertical velocity from the surface to each reflecting bed.  
(d) Use these velocities to compute the depths to the reflecting beds.

13. The following table gives two-way travel-times of seismic waves reflected from different reflecting interfaces in a horizontally layered medium.

Geophone to shot-point distance [m]	Travel-time [s] to		
	First reflector	Second reflector	Third reflector
500	0.299	0.364	0.592
1000	0.566	0.517	0.638
1500	0.841	0.701	0.708
2000	1.117	0.897	0.799
2500	1.393	1.099	0.896

- (a) Draw a plot of  $(\text{travel-time})^2$  against  $(\text{distance})^2$ .  
(b) Determine the vertical two-way travel-time ("echo-time") and average velocity to each reflecting interface.  
(c) Compute the depth of each reflector and the thickness of each layer.  
(d) Compute the true velocity (interval velocity) of each layer.

### 3.10 EXERCISES

- (e) Verify your results by computing the total vertical travel-time for a wave reflected from the deepest interface.

14. Assume the horizontally layered structure from the previous problem.

- (a) If a seismic ray leaves the surface at an angle of 15° to the vertical, how long does it take to return to the surface after reflecting from the basement?  
(b) At what horizontal distance from the shot-point does this ray reach the surface?

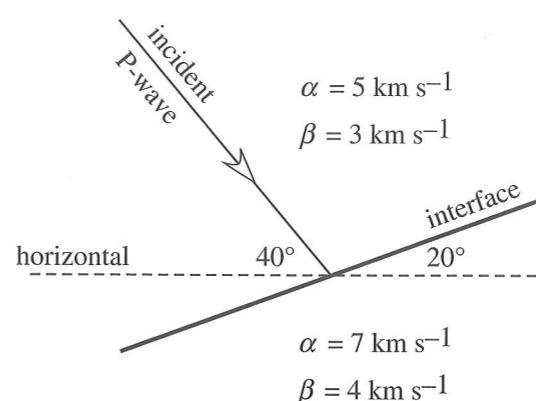
15. Assume that the three horizontal homogeneous rock layers in the previous problems have densities of 1800, 2200, and 2500 kg m<sup>-3</sup> respectively. The lowest layer overlies basement with velocity 5.8 km s<sup>-1</sup> and density 2700 kg m<sup>-3</sup>.

- (a) Compute the reflection and transmission coefficients at each interface for a plane P-wave travelling vertically downwards.  
(b) Calculate what fraction of the initial energy of the wave is transmitted into the basement.  
(c) Calculate the fraction of the initial energy carried in the reflection that returns to the surface from the basement.

16. An incident P-wave is converted into refracted and reflected P- and S-waves at an interface. Calculate all the critical angles in the following three cases, where  $\alpha$  and  $\beta$  are the P-wave and S-wave velocities, respectively:

Layer	Seismic wave	Case (a) [km s <sup>-1</sup> ]	Case (b) [km s <sup>-1</sup> ]	Case (c) [km s <sup>-1</sup> ]
Above interface	$\alpha$	3.5	4.0	5.5
	$\beta$	2.0	2.3	3.1
Below interface	$\alpha$	8.5	6.0	7.0
	$\beta$	5.0	3.5	4.0

17. An incident P-wave is converted into refracted and reflected P- and S-waves at an interface that is inclined at 20° to the horizontal, as in the figure below. The respective P- and S-wave velocities are 5 km s<sup>-1</sup> and 3 km s<sup>-1</sup> above the interface and



- 7 km s<sup>-1</sup> and 4 km s<sup>-1</sup> below the interface. If the incident P-wave strikes the interface at an angle of 40° to the horizontal, calculate the angles to the horizontal made by the reflected and refracted P- and S-waves.

18. A seismic refraction survey gave the following data for the first arrival times at various distances from the shot-point.

Distance [km]	Time [s]	Distance [km]	Time [s]
3.1	1.912	13.1	6.678
5.0	3.043	14.8	7.060
6.5	3.948	16.4	7.442
8.0	4.921	18.0	7.830
9.9	5.908	19.7	8.212
11.5	6.288		

- (a) Plot the travel-time curve for the first arrivals.  
(b) Calculate the seismic velocities of the layers.  
(c) Calculate the minimum depth to the refracting interface.  
(d) Calculate the critical angle of refraction for the interface.  
(e) Calculate the critical distance for the first arrival of refracted rays.  
(f) Calculate the crossover distance beyond which the first arrivals correspond to head waves.

19. A seismic refraction survey is carried out over a layered crust with flat-lying interfaces. In one case the crust is homogeneous and 30 km thick with a P-wave velocity 6 km s<sup>-1</sup> and overlies mantle with P-wave velocity of 8 km s<sup>-1</sup>. In the other case the crust consists of an upper layer 20 km thick with P-wave velocity 6 km s<sup>-1</sup> overlying a lower layer 10 km thick with P-wave velocity 5 km s<sup>-1</sup>. The upper mantle P-wave velocity is again 8 km s<sup>-1</sup>. On the same graph, plot the first arrival time curves for the two cases. What is the effect of the low-velocity layer on the estimation of depth to the top of the mantle?

20. The table below gives up-dip and down-dip travel-times of P-wave arrivals for refraction profiles over an inclined interface. The geophones are laid out in a straight line passing through the alternate shot-points A and B, which are 2700 m apart on the profile.  
(a) Plot the travel-time curves for each shot-point.  
(b) Calculate the true velocity of the upper layer.  
(c) Calculate the apparent velocities of the layer below the refractor.  
(d) In which direction does the refracting interface dip?  
(e) What is the angle of dip of the interface?  
(f) What is the true velocity of the layer below the refractor?

Distance from shot-point [m]	Travel-time [s]	
	from A	from B
300	0.139	0.139
600	0.278	0.278
900	0.417	0.417
1200	0.556	0.556
1500	0.695	0.695
1800	0.833	0.833
2100	0.972	0.972
2400	1.085	1.111
2700	1.170	1.170
3000	1.255	1.223
3300	1.339	1.276
3600	1.424	1.329

- (g) What are the closest distances to the refractor below A and B?
- (h) What are the vertical depths to the refractor below A and B?