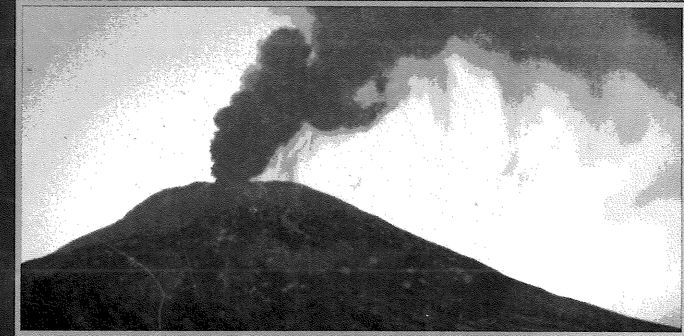
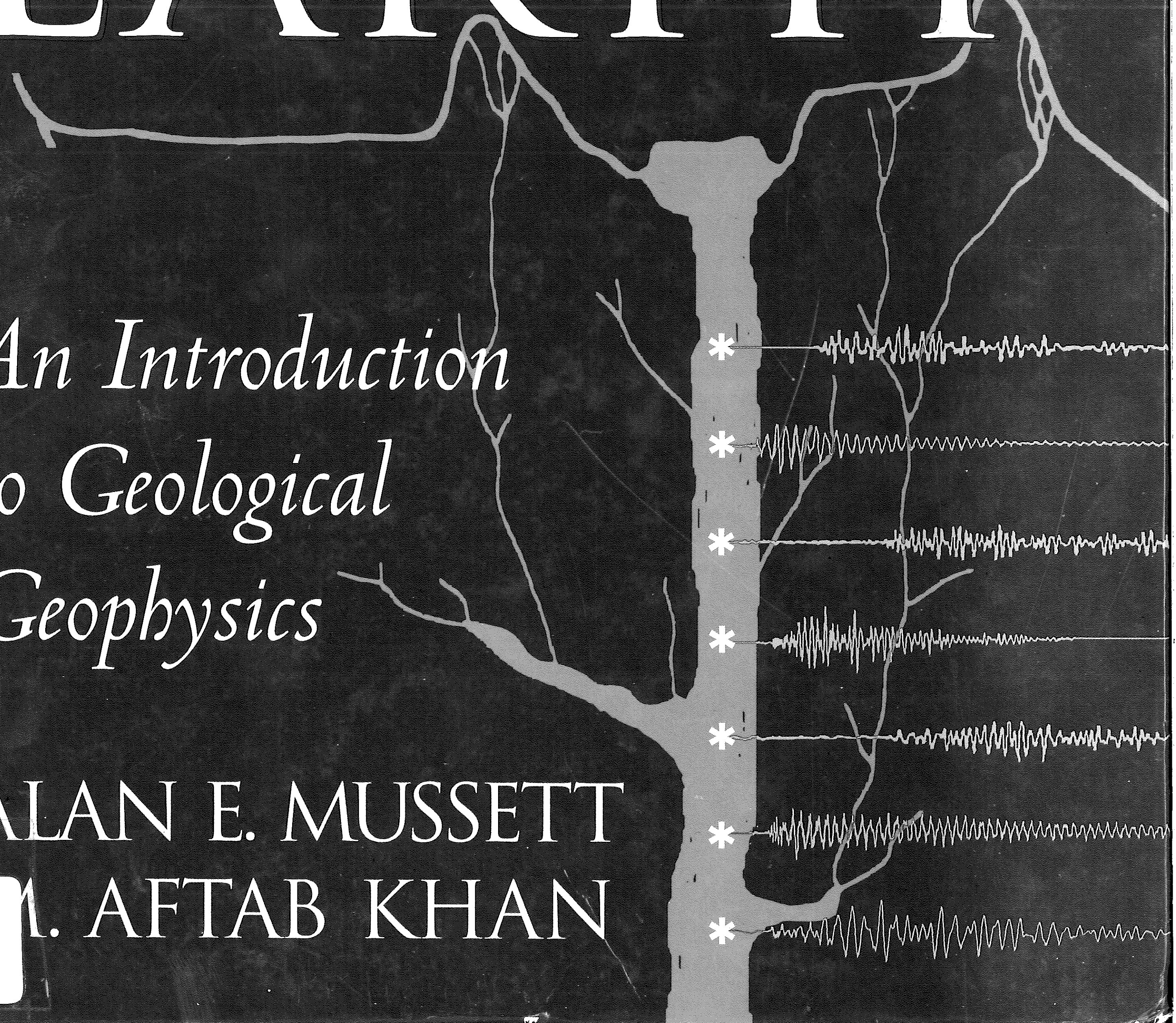
LOOKING INTO THE EARTH



An Introduction to Geological Geophysics



ALAN E. MUSSETT
M. AFTAB KHAN

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014, Madrid, Spain

First published 2000

Printed in the United States of America

Typeface 10/13 pt. Sabon *System* Quark XPress® [GH]

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

Library of Congress Cataloging-in-Publication Data
Mussett, A. E. (Alan E.)
Looking into the earth : an introduction to geological geophysics / Alan E.
Mussett, M. Aftab Khan ; [illustrations by Sue Button].
p. cm.
Includes bibliographical references and index.
ISBN 0-521-78085-3 (hc.) – ISBN 0-521-78574-X (pbk.)
1. Geophysics. I. Khan, M. Aftab. II. Title.
QE501.M87 2000
550 – dc21 00-020382

ISBN 0 521 78085 3 hardback
ISBN 0 521 78574 X paperback

Contents

Preface: Turning a Magician into an Expert

Geophysics is essential to understanding the solid Earth, particularly on a global scale. Modern ideas of the structure and evolution of continents and oceans, or of the formation of mountain chains on land and below the oceans, for instance, are based extensively on discoveries made using geophysics. But geophysics can contribute to geological knowledge on all scales, from the global, through the medium-scale such as regional mapping or the search for oil and minerals, down to the small-scale, such as civil engineering, archaeology, and groundwater pollution, as well as detailed geological mapping.

Geophysics differs from other methods for studying the Earth because it can 'look into the Earth', for its measurements are mostly made remotely from the target, usually at the surface. It is able to do this because it measures differences in the physical properties of the subsurface rocks or structures, which are revealed by their effects at the surface, such as the magnetic field of some rocks. But it describes the subsurface in physical terms – density, electrical resistivity, magnetism, and so on, not in terms of compositions, minerals, grain-sizes, and so on, which are familiar to the geologist. Because geologists are often unfamiliar with physics (and the associated mathematics), there is a tendency either to ignore geophysics or to accept what a geophysicist says without understanding the qualifications.

This last is simply to treat the geophysicist as some sort of a magician. All subjects have their assumptions, often unstated but known to practitioners. Anyone who has used a geological map knows that just because some area has the same colour is no guarantee that the same rock will be found everywhere within it, or that faults occur exactly where marked and nowhere else. Similarly, there are reservations and limitations on what geophysics tells you, which need to be understood.

The intention of this book is not to turn you into a geophysicist (you may be glad to know), but simply to provide a basic grasp of the subject, so that, for instance, if a seismic section is described as 'unmigrated' you know whether it matters (it well might!). Or you may need to know whether a geophysical survey could help solve some geological problem of yours, so when you call on the services of a geophysicist (and worse things may happen!) you need to be able to explain your problem and understand his or her advice. You would not buy a car without specifying the number of seats, engine size, whether a sedan or sports model, and so on that you would like, and the price you are prepared to pay; and you would listen to claims about its fuel consumption, insurance group, and the choice of radios. Similarly, you need an understanding of geophysics.

The purpose of this book is to explain how geophysics helps us understand the solid Earth beneath our feet, and how – combined with the traditional methods of geology – greatly extends what can be learned about the Earth.

SUBPART I.2

Seismology

Seismology is primarily concerned with determining the structure of the Earth – on all scales – and only subsidiarily with earthquakes. It uses the ability of seismic waves – vibrations of rocks – to propagate through the Earth. They do not generally travel in straight lines but are deflected, by refraction or reflection, by the layers they encounter, before they return to the surface of the Earth, and this allows the internal structure to be determined. Seismology is particularly useful for determining the positions of roughly horizontal interfaces between layers, which makes it the most useful single geophysical technique.

This subpart is divided into four chapters: **Global seismology and seismic waves** *first explains how seismology works and then uses it to explore the Earth's structure on the largest scale.* **Refraction seismology** *exploits the special case of seismic waves propagating along an interface after refraction, while* **Reflection seismology** *depends on waves reflected back up from interfaces.* **Earthquakes and seismotectonics** *– which comes second – is concerned with earthquakes and what they reveal about the tectonic processes that produce them.*

Chapter 4

Global Seismology and Seismic Waves

Global seismology reveals that, on the largest scale, the Earth is concentrically layered, with the primary divisions of crust, mantle, and core, plus other concentric features. It also shows that the core is divided into a liquid outer core and a solid inner core.

As seismology depends upon seismic waves, which travel in the Earth, the chapter begins by introducing these, together with their generation, propagation, and detection.

4.1 Waves, pulses, and rays

Waves. An example of a **wave** is a water wave, but so also are waves produced by shaking a rope attached at its further end, or pushing in and out a long spring (Fig. 4.1). If the end is moved rhythmically, a series of disturbances travels along the rope or spring; Figure 4.1 shows them before the disturbances have reached the further end.

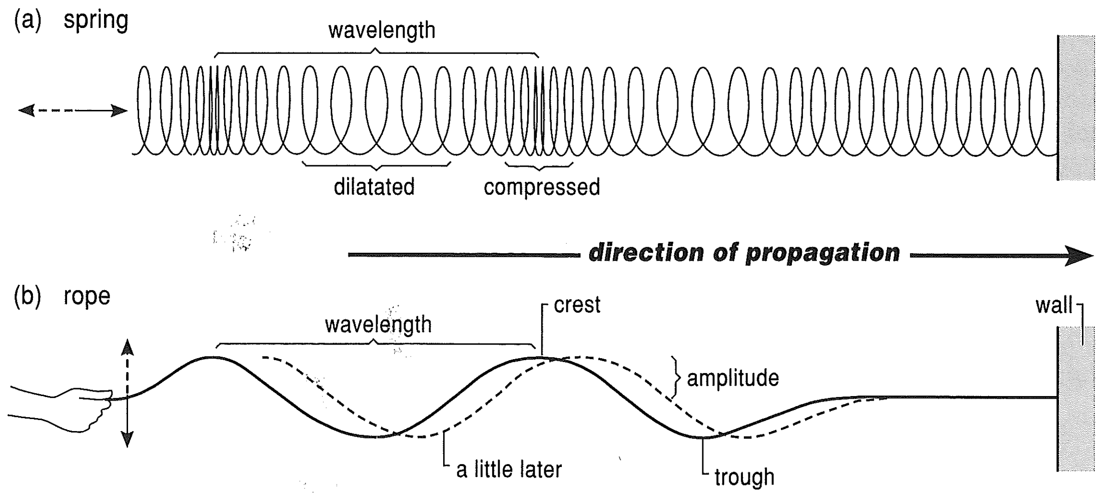


Figure 4.1 Waves along a spring and a rope.

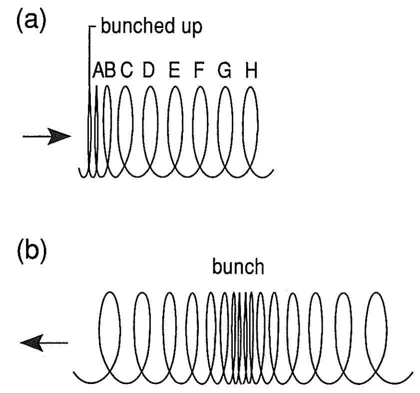


Figure 4.2 Propagation of waves.

Consider the spring in Figure 4.2. When it is held stretched but stationary, the turns will be equally spaced because the tension of the spring upon each turn is the same from either side. Pushing the end in quickly a short distance causes the first couple of turns, *A* and *B*, to bunch up (Fig. 4.2a), but then the next turn, *C*, being no longer pulled so hard from the left, moves to the right and bunches up against *D*, and so on, propagating a compression to the right. Then, if the end is pulled out, stretching the spring, a dilatation propagates along the spring, following the compression. So a rhythmic pushing in and pulling out of the end of the spring produces a regular series of compressions and dilatations that move along, forming waves.

Though the waves travel along, the spring does not, each turn only oscillating about its stationary position. Similarly, water is not moved along by water waves, which is why you cannot propel a ball across a pond by generating waves behind it by throwing in stones.

There are some terms you need to learn, most of them illustrated in Figure 4.1 (some were introduced in Section 3.1.1). **Wavelength, λ,** is the repeat length, conveniently measured between successive crests or compressions. The **amplitude, *a*,** is the maximum displacement from the stationary position. The waves travel along at some speed called, in seismology, the **seismic velocity, *v*** (a velocity should specify the direction of propagation as well as its speed, but this is often neglected in seismology). The number of crests or compressions that pass any fixed point on the rope, spring, and so on, in one second is the **frequency, *f*,** measured in Hz (Hertz, oscillations, or cycles, per second). In one second, f wave crests, each λ metres apart, will pass a point; and by the time the last one has passed, the first one will have travelled a distance of f times λ metres. As velocity is the distance travelled in one second,

$$v = f \times \lambda$$

velocity = frequency × wavelength Eq. 4.1

Pulses. You can see waves travelling along springs, ropes, and water, so it so easy to measure how fast they are moving. In seismology, we can learn a lot just by timing how long it takes seismic waves to travel different distances through the Earth, but, of course, we can't see them moving inside the Earth. We therefore need a way of 'marking' waves so that their progress can be observed. The simplest way is to generate just a few waves and time how long it is before the ground some distance away begins to move. A very short series of waves is called a **pulse.** Pulses can have various shapes, but a very simple one is just one compression (or crest) followed by one dilatation (or trough), as shown in Figure 4.3.

Seismic pulses are easily generated: Any quick, sharp disturbance of the ground does it, from a hammer blow to the onset of an earthquake. One common way is to fire a charge of explosive buried at the bottom of a hole (Fig. 4.4). The rapid expansion produces a compression that travels spherically outwards; then the ground tends to spring back into the cavity produced, causing a dilatation. The compression has a spherical surface called a **wave front,** and it expands away from the source. (Waves due to explosions are often described as shock waves, but shock waves exist only very close to the explosion, where material has been forced bodily outwards, moving faster than the natural speed of propagation of ground oscillations. They slow down within a metre or less and become normal seismic waves.)

Typical waves studied in seismology have velocities measured in km/sec and frequencies of some tens of Hertz. For example, a wave with velocity of 2 km/sec and frequency of 10 Hz has a wavelength, found using Eq. 4.1, of 0.2 km or 200 m. So seismic wavelengths can be quite long compared to thicknesses of strata or other types of layers. This has important consequences, explained in Section 7.8.2.

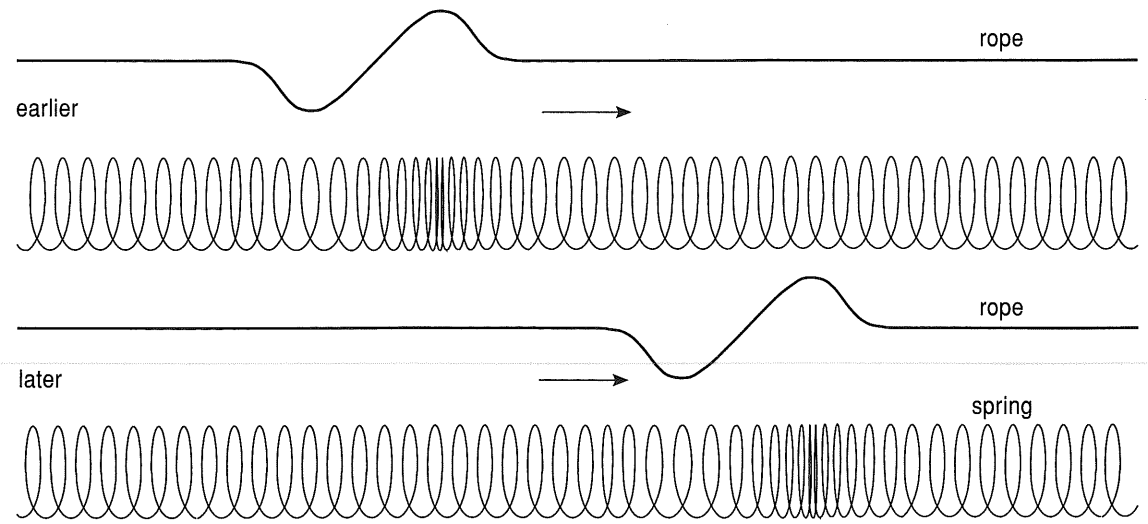


Figure 4.3 A pulse.

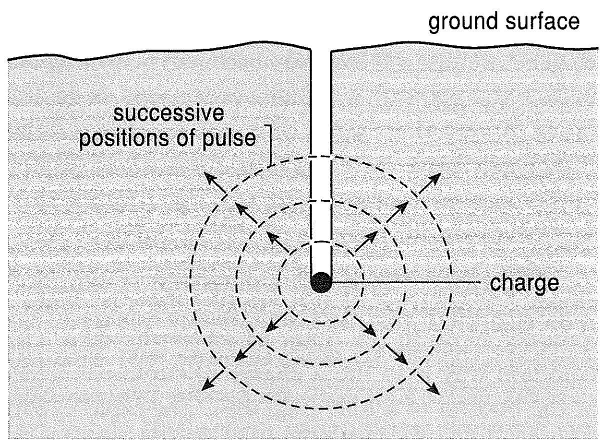


Figure 4.4 Pulse generation by an explosion.

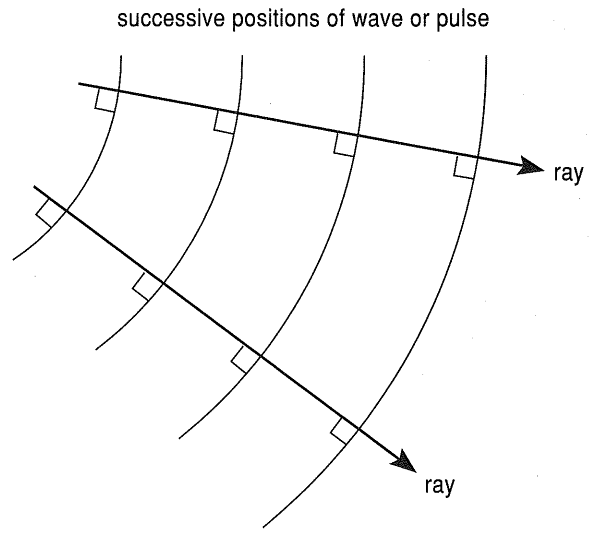


Figure 4.5 Rays and wave fronts.

Rays. In the solid Earth, waves travel outwards from their source in all directions (Fig. 4.4). If we are interested only in what happens in one direction we need consider only part of the wave front. The path of a tiny portion of the wave front, or pulse, forms a **ray** (Fig. 4.5). *Rays are always perpendicular to wave fronts, and vice versa.* As rays are simpler to consider than waves, most seismology theory will be explained using them.

4.2 Detecting seismic waves: Seismometers and geophones

When a seismic wave passes a point – perhaps after travelling deep in the Earth – it causes the ground to oscillate. It is easy to measure the movement of water waves if we are on the stationary shore, but how can we measure ground motion when we ourselves are moving up and down, or from side to side?

Consider a plumb bob hanging from a frame resting firmly on the ground (Fig. 4.6a). If the ground suddenly moves – to the left, say – the frame moves with it, but the bob tends to remain still. Thus the scale moves left past the bob, giving a positive deflection on the scale shown. This design of instrument detects horizontal ground motion; vertical ground motion could be measured using a bob on a spring (Fig. 4.6b).

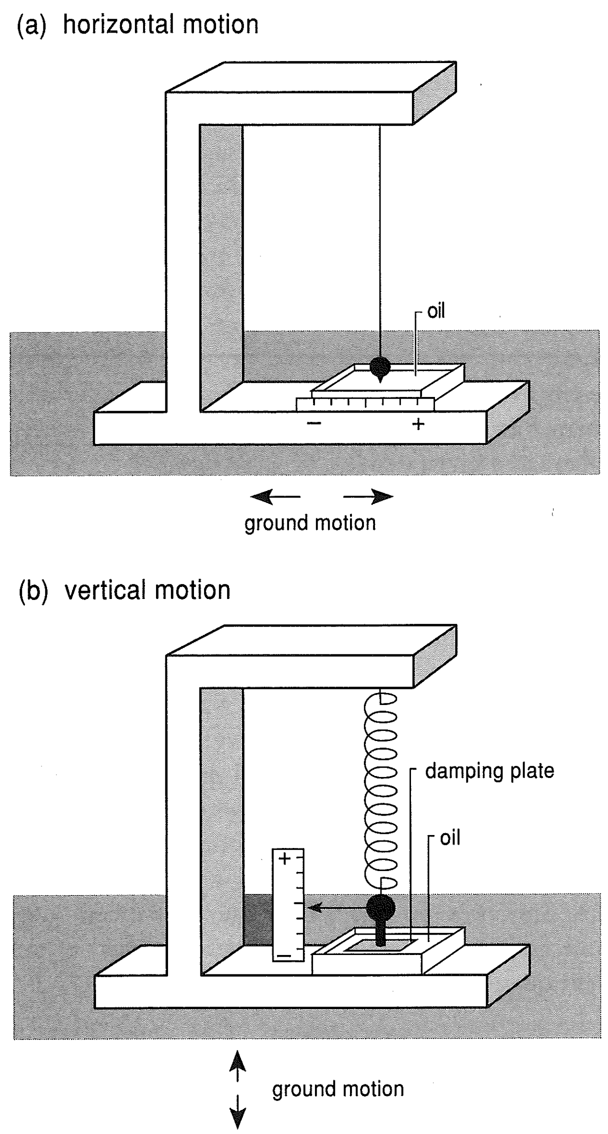


Figure 4.6 Principle of seismometers and geophones.

Of course, shortly after the frame has moved the bob will begin to move too, and thereafter it will tend to go on oscillating, obscuring the record of any subsequent ground motion. To reduce this effect the instrument is 'damped', which could be done by having part of the bob immersed in oil, causing the oscillations to die away once the ground stops moving.

Actual instruments have to be compact, damped in some way, and very sensitive, but depend on the above principles. Moving-coil and moving-magnet instruments are used. In Figure 4.7, the moving mass is a magnet suspended inside a coil of wire by a compact spring; relative movement of magnet and coil generates a small electric current (as described in Section 14.1.1), which is amplified electronically; this is the principle used in some record pickups. The amplified signal is sometimes recorded on paper on a moving drum for immediate inspection, but records are usually stored electronically, on magnetic tape, often employing digital recording (described in Section 7.7.1) because it is easier to process. Digital recording is also used, for instance, in compact discs.

Instruments are classed as **seismometers** or **geophones**, but function similarly. Compared to geophones, seismometers are much more sensitive – but are less robust and compact and have to be set up carefully – so they are used to measure weak signals, as is often the case with global seismology, because sources are distant. A modern seismometer can detect the ground motion caused by a person walking a kilometre away, if there is little other disturbance. A seismometer record – a **seismogram** – is shown later, in Figure 4.17. Geophones are used in small-scale seismic surveys where many instruments have to be set out quickly but the highest sensitivity is not needed. We call both seismometers and geophones seismic **receivers.**

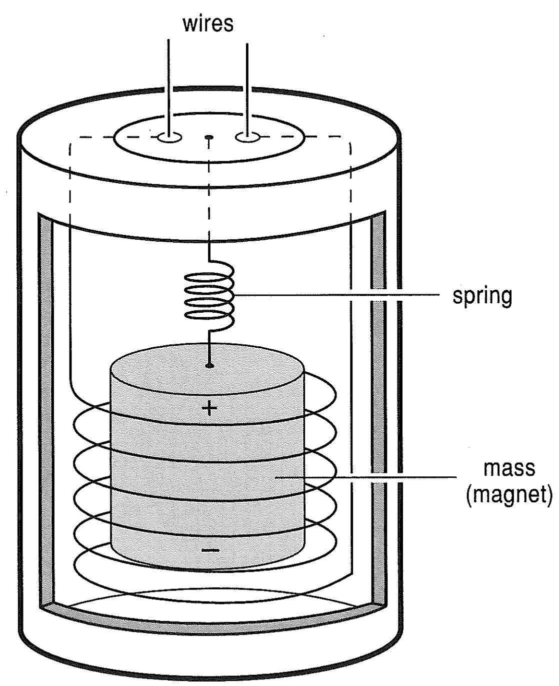


Figure 4.7 Moving-magnet seismometer.

To obtain full information about the ground motion, three seismometers are needed (sometimes combined into a single instrument): one to measure the vertical part or component of ground motion, the other two to measure horizontal motions at right angles, often N–S and E–W (Fig. 5.22 shows a three-component seismogram). Often, single-component instruments are used for economy.

4.3 The Earth is concentrically layered

4.3.1 Spherical symmetry of the Earth's interior

The Earth is nearly a sphere, as pictures taken from space show. Of course, it is not quite a sphere because of mountains and other topographic features, but the height of even Mount Everest is only a thousandth of the Earth's radius.

A much larger departure from sphericity is the equatorial bulge, caused by the centrifugal force of the Earth's rotation spinning out its material (see Fig. 9.17). The difference in radii is only about a third of one percent. The bulge has a smooth shape, unlike the jaggedness of mountains, and so can be allowed for in calculations; we shall assume this has been done and so shall seldom mention the bulge again in this chapter.

The first question to ask about the interior of the Earth is whether it also is spherically symmetrical: Is it layered like an onion, or are there large lateral differences as found for rocks near the Earth's surface? We can test this by timing how long it takes a pulse of seismic energy to travel between pairs of points *separated by the same distance*. In Figure 4.8 the ray paths are shown dashed because we don't yet know where they are; they serve just to connect the pairs of points, A_1B_1, A_2B_2, ... Be clear that we are not at this point testing whether the interior is seismologically uniform but only if it has the same velocity everywhere at a given depth.

In global seismology distances are usually given, not as kilometres around the surface, but as the angle subtended at the centre, **the epicentral angle, Δ.** Measurements show that the times to travel all paths with the same epicentral angle are nearly the same, and this is true for both large and small angles (Fig. 4.8a and b), so the Earth is indeed spherically symmetrical, like an onion. (There are very small differences, which will be considered in Section 4.6.) This is convenient because now we need not bother where the actual locations of sources and seismometers are, only how far apart they are, so we can combine results from all over the world to answer the next question.

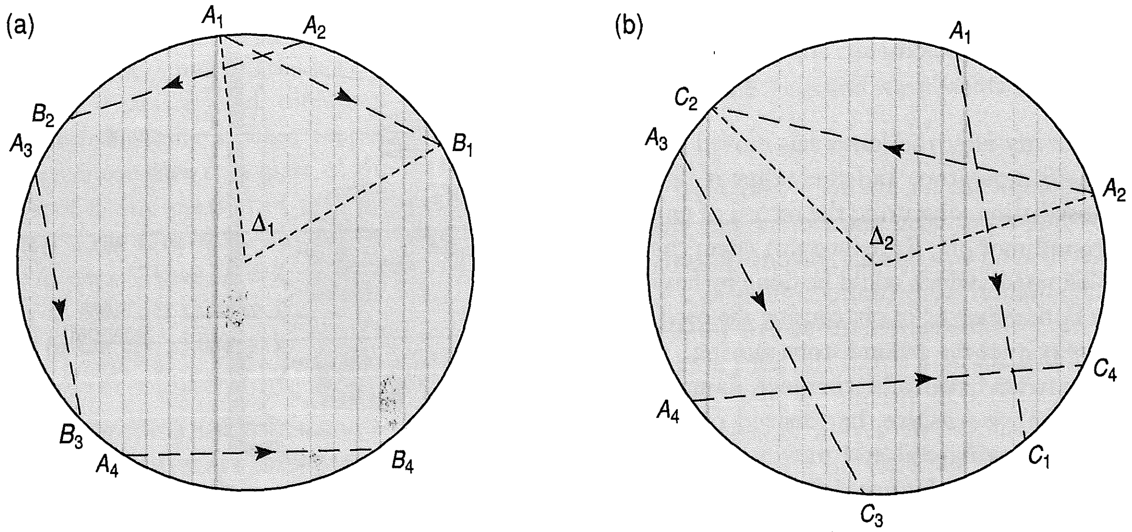


Figure 4.8 Same-length paths through the Earth.

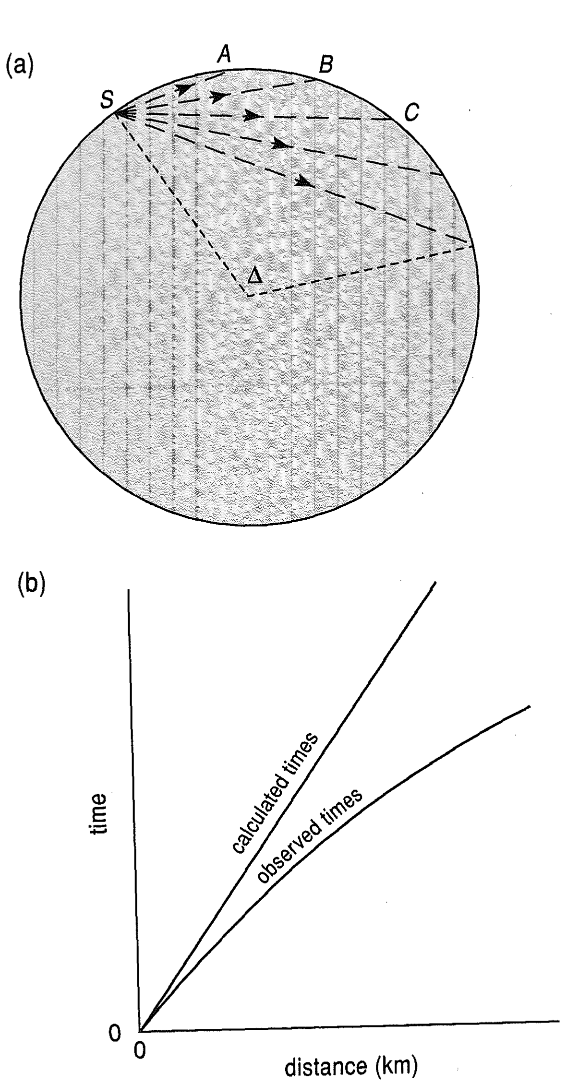


Figure 4.9 Travel-times for a uniform and the actual Earth.

4.3.2 Concentric layering

Having established that the Earth is symmetrical, we next ask, Is it perhaps uniform, like an onion with only one layer? If so, the seismic velocity would be the same at all depths, rays would be straight lines (Fig. 4.9a), and the time that rays take would be simply proportional to the distance they travel. This is easily tested by comparing times calculated for a uniform Earth with actual observed times.

The calculated times assume that all the Earth has a seismic velocity equal to the average of surface rocks, which we can measure. Figure 4.9b shows that as the source-to-receiver distance (straight through the Earth) is increased, the time observed is progressively *less* than for a uniform Earth. So the Earth is not seismically uniform, and as longer rays have travelled deeper into the Earth the seismic velocity is faster at depth.

How much faster? This is not so easy to answer because when the seismic velocity is not uniform, ray paths are not straight. So we need to know how wave propagation and rays are affected when they encounter a change of velocity, the topic of the next section.

4.4 Finding the path of a ray through the Earth

4.4.1 Refraction: Snell's law

When wave fronts cross obliquely into a rock with a higher seismic velocity, they speed up, which causes them to slew around and change direction (Fig. 4.10a), just as a line of people, arms linked, would slew if each person went faster once they had crossed a line. This bending is called **refraction.**

How much is the change of direction? First we note that a long way from the source a small part of a wave front will be nearly a flat plane, so two successive wave fronts – or successive positions of the same wave front – will be parallel, and rays will be straight, for they are perpendicular to wave fronts.

The *time* between successive wave fronts, AB and $A'B'$ in Figure 4.10b, remains unchanged, so the wavelength must increase in the second rock, in proportion to the increase in velocity. From geometry,

$$\frac{v_1}{v_2} = \frac{\lambda_1}{\lambda_2} = \frac{BB'}{AA'} \qquad \text{Eq. 4.2}$$

By trigonometry,

$$\sin i_1 = \frac{BB'}{AB'} \quad \sin i_2 = \frac{AA'}{AB'} \qquad \text{Eq. 4.3}$$

Therefore

$$AB' = \frac{BB'}{\sin i_1} = \frac{AA'}{\sin i_2} \qquad \text{Eq. 4.4}$$

As BB' and AA' are in proportion to the velocities v_1 and v_2 (Eq. 4.2), this can be rearranged to give

$$\frac{\sin i_1}{v_1} = \frac{\sin i_2}{v_2} \qquad \text{Snell's Law} \qquad \text{Eq. 4.5}$$

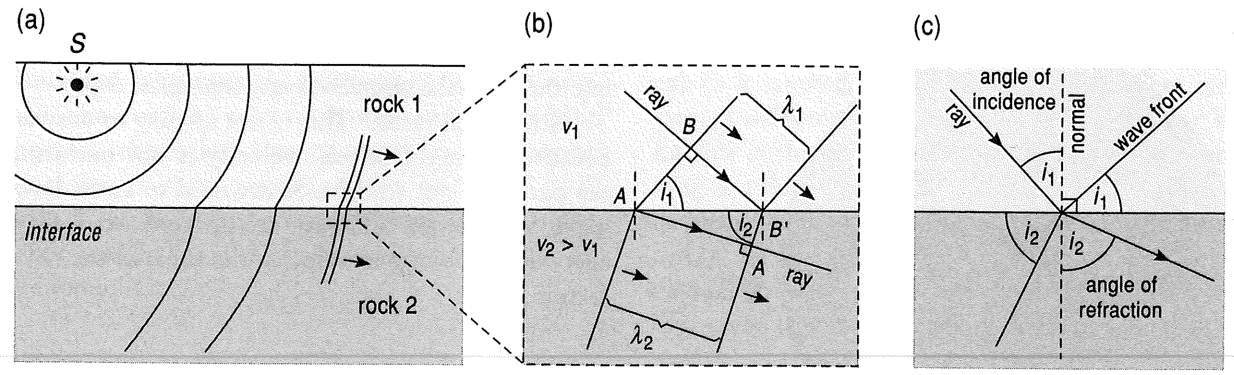


Figure 4.10 Refraction of a wave front.

This equation, **Snell's Law**, is essentially the same as the one used for the refraction of light, but this form is more useful for seismology (for light, the equation is rearranged with the velocities replaced by their ratio, which is called the refractive index).

Since it is more convenient to use rays, which show the direction of propagation, than wave fronts, i_1 and i_2 are measured as the angles between the rays and the perpendicular, or **normal,** to the **interface** between the two rock types (Fig. 4.10c); these are called the **angles of incidence** and **refraction.** They have the same values as the angles between the wave fronts and the interface.

As an example, and referring to Figure 4.11, at what angle would the ray leave the interface if the angle of incidence is 37°?

$$\frac{\sin 37°}{4} = \frac{\sin i_2}{5}$$

$$\sin i_2 = \frac{5}{4}\sin 37° \quad \text{so} \quad i_2 = 48.8°$$

After refraction the ray leaves the interface at 48.8°.

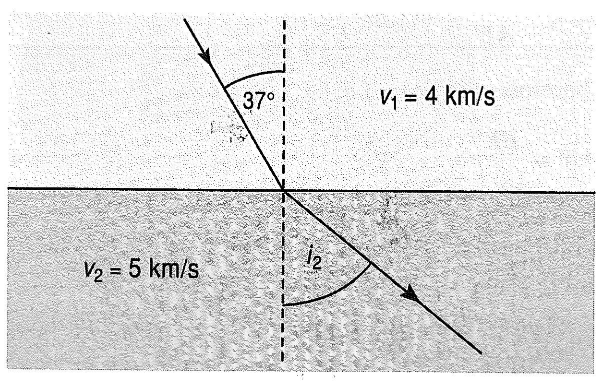


Figure 4.11 Example of refraction.

Now that Snell's Law has been introduced, most diagrams will be simplified to show only rays. Snell's Law also applies to **reflection,** as will be shown in Section 4.5.2.

4.4.2 Tracing rays through the Earth: The ray parameter, *p*

If there are several *parallel* and uniform layers (Fig. 4.12), a ray meets the next interface at the angle at which it left the last one, that is, $i_2 = i_1'$, and so on. Applying Snell's Law at each interface, we have

$$\frac{\sin i_1'}{v_1} = \frac{\sin i_2}{v_2} \quad \frac{\sin i_2'}{v_2} = \frac{\sin i_3}{v_3} \quad L \qquad \text{Eq. 4.6}$$

As $i_1' = i_1$, $i_2' = i_2$, and so on

$$\frac{\sin i_1}{v_1} = \frac{\sin i_2}{v_2} = \frac{\sin i_3}{v_3} = L = \text{constant} \qquad \text{Eq. 4.7}$$

The ratio (sin *i/v*) remains unchanged, or constant, along the ray path.

In the Earth, however, the layers are curved, so it is not true that $i_2 = i_2'$, and so on (Fig. 4.13). The differences between these angles depend not on the velocities of the layers but only on the geometry of triangle *ABO* (the relationship will not be derived). Snell's Law determines how the angle of a ray changes on crossing an interface, while geometry determines the change of angle between interfaces. These can be combined to give

$$\frac{r_1 \sin i_1}{v_1} = \frac{r_2 \sin i_2}{v_2} = L = \text{constant} = p \qquad \text{Eq. 4.8}$$

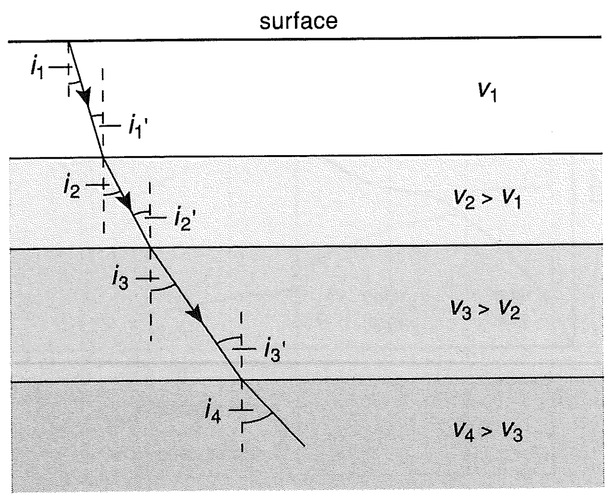


Figure 4.12 Refractions at parallel interfaces.

The constant *p* is known as the **ray parameter** and has the same value all along the path of any given ray, provided all three quantities, *v*, *i*, and *r*, are measured at the same place. For refraction at any single interface, *r* is the same on both sides and the *p* parameter simplifies to Snell's Law; thus the *p* parameter includes Snell's Law.

4.4.3 Ray tracing and the Earth's velocity–depth structure

If the variation of the seismic velocity with depth (or distance, *r*, from the centre) were known, then ray paths through the Earth could be deduced using the ray parameter, *p*. The Earth would be divided into a large number of thin shells (Fig. 4.14a), which would take account of any changes of seismic velocity, whether abruptly between real layers or gradationally within layers. Then a ray could be chosen that dives into the Earth at some angle *i*, the **take-off angle** (Fig. 4.14a), and its path through the layers could be followed using the ray parameter equation, Eq. 4.8 (Fig. 4.14b). The times to travel along each of the little sections within shells could be calculated and added together to give the total **travel-time.** By repeating this for take-off angles from horizontal to vertical, a travel-time versus distance, *t–Δ*, diagram could be worked out (Fig. 4.14c).

However, our problem is the inverse. We are confined to the surface and can only measure travel-times from a seismic source to detectors at different distances; from these, we have to determine how velocity changes with depth. This is an example of the inversion problem, introduced in Section 2.4. Generally, it is much harder to solve the inverse than the forward problem. The mathematics needed to invert travel-times into velocity versus depth were solved by the pioneers of global seismology but are beyond the level of this book. Nowadays, we can adopt a different approach using computers: Starting with some estimate of how the velocity depends on depth, such as earlier solutions, travel-times are calculated for the distances to actual seismic receivers and compared with the observed times. If there is a discrepancy, the velocity–depth curve is adjusted to minimise it. This is repeated for millions of seismic records of thousands of earthquakes obtained from hundreds of seismometers all over the world.

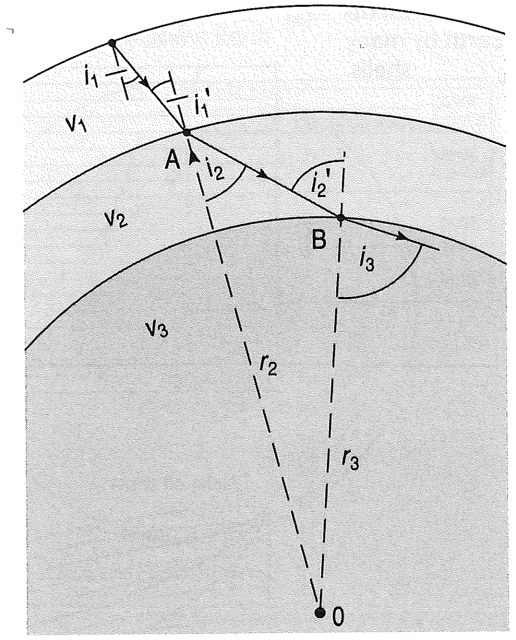


Figure 4.13 Ray path through layered Earth.

One result is shown in Figure 4.15a, a solution published by Kennett and Engdahl in 1991, and known as model iasp91. It is an average solution, treating the Earth as being perfectly spherically symmetrical (after allowance for the equatorial bulge), and ignoring small, lateral variations.

Why there are two curves and what they reveal about the Earth is discussed in the following section.

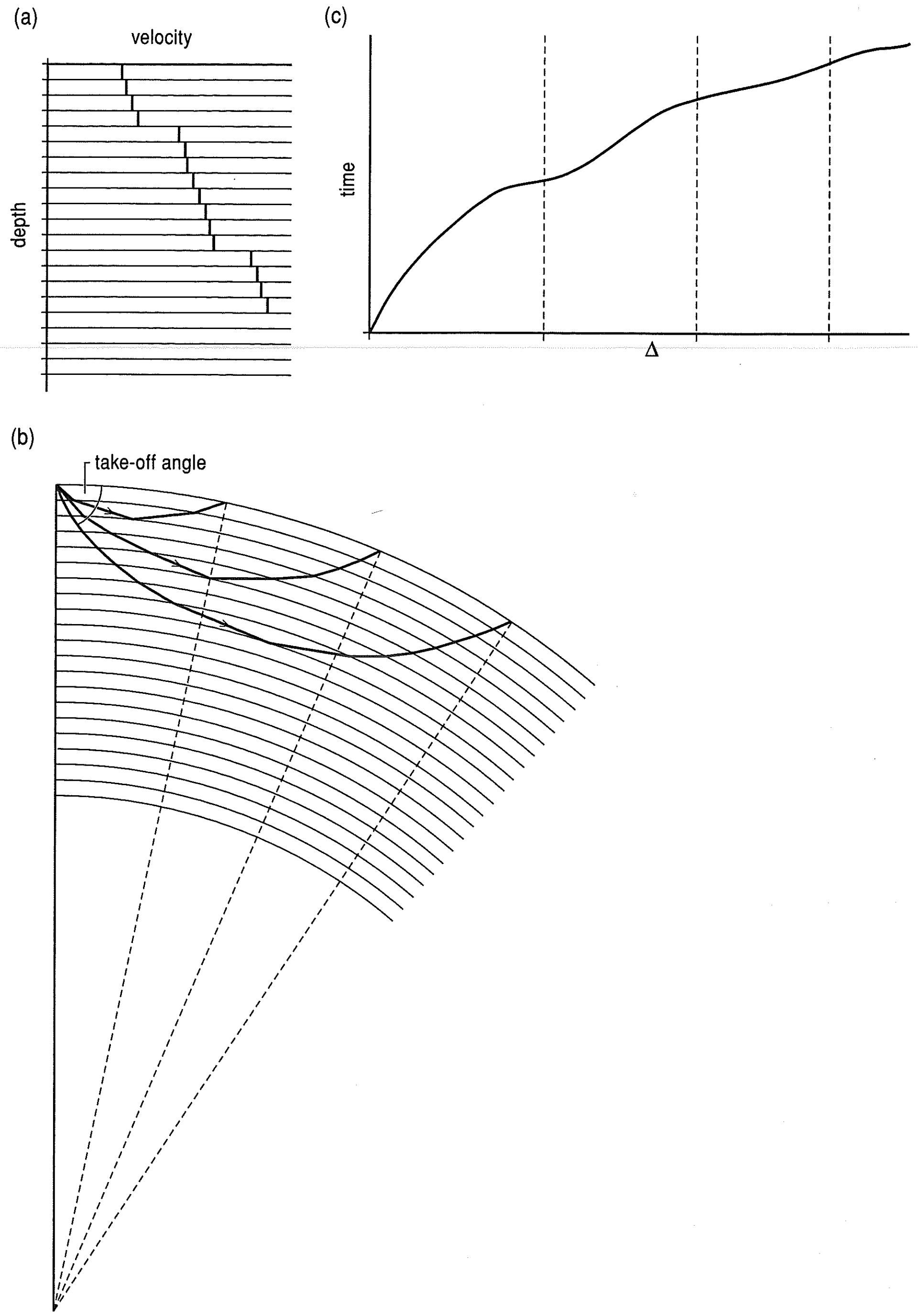


Figure 4.14. Representing the Earth by many shells.

4.5 Seismic features of the Earth

4.5.1 Core and mantle

Figure 4.15a shows that seismic velocity generally increases with depth, except that about halfway to the centre there is an abrupt decrease. An abrupt increase or decrease is called a **velocity discontinuity,** and because these mark the boundary between bodies with different properties they are also called **interfaces.** This interface is a major feature, and *by definition* divides the **mantle** from the **core** (Fig. 4.15b). It is therefore called the **core–mantle boundary** (often abbreviated to **CMB**). To find out what it is, we use the fact that there are different types of seismic waves, with different velocities.

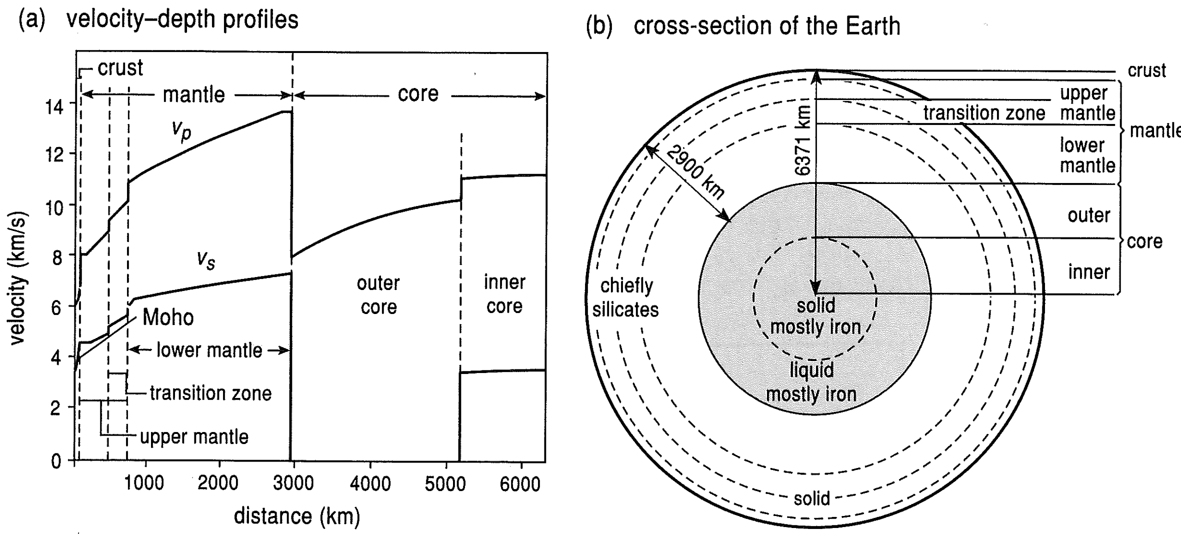


Figure 4.15 Velocity–depth curves for P- and S-waves, and the Earth's structure.

4.5.2 Longitudinal and transverse waves

Figure 4.1 showed waves along a spring and a rope. For the spring, the turns moved back and forth in the direction the wave travels; this is an example of a **longitudinal wave.** But for the rope, points oscillate at right angles to the direction of wave travel; this is a **transverse wave.** Seismic waves also exist in longitudinal and transverse forms; Figure 4.16 shows how an imagined column through rock is deformed as waves travel through it. If the column were marked into cubes, as shown to the right, the longitudinal wave would cause them to compress and then dilate as crests and troughs of the wave pass through, changing both their shape and size; but the transverse wave changes only their shape, making them lozenge shaped.

Longitudinal and transverse seismic waves are called **P-** and **S-waves** (for historical reasons); they may be remembered as pressure or push–pull waves, and shear or shake waves. P-waves are essentially the same as sound waves, except that many of the frequencies recorded are too low to be heard by the human ear. Sound waves travel well through most rocks, usually better and more quickly than through air, which explains, for instance, why the Indians of North America used to put their ear to the ground to hear the approach of distant cavalry.

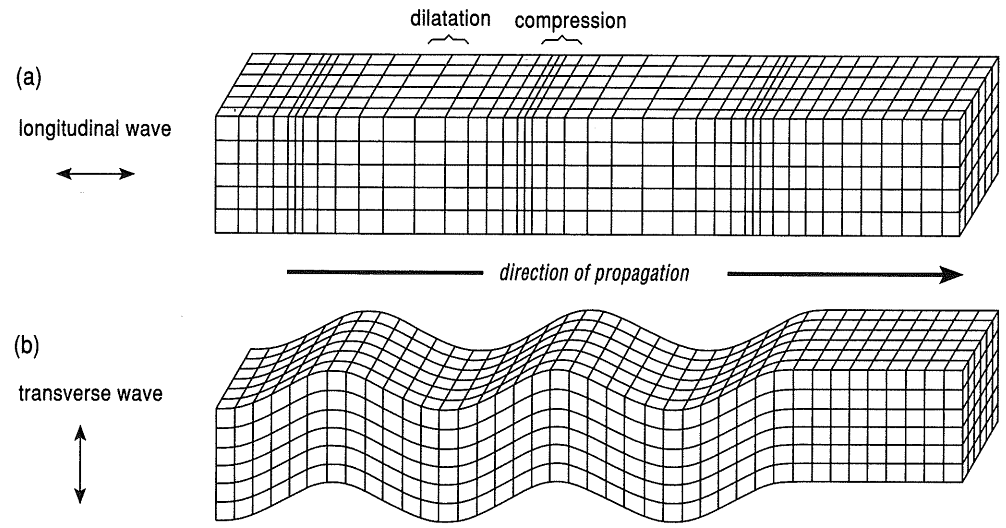


Figure 4.16 Longitudinal and transverse waves in rocks.

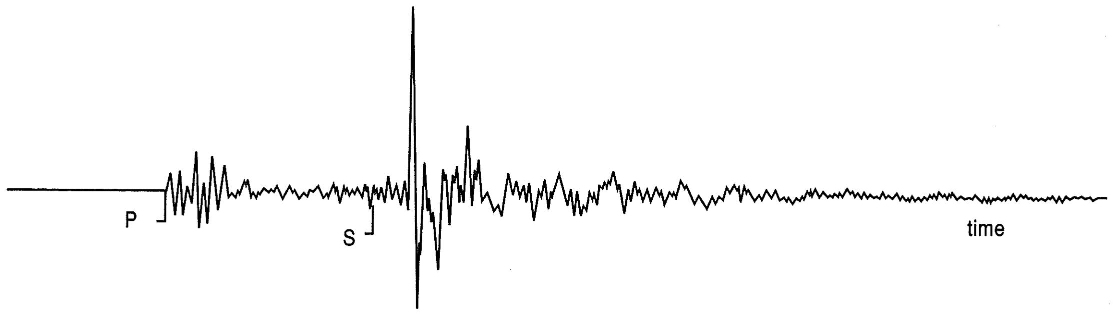


Figure 4.17 Seismogram.

Because P- and S-waves deform a rock in different ways the restoring forces are different, and as a result they travel at different speeds. P-waves are always faster, so P-waves arrive before S-waves, as shown in the seismogram of Figure 4.17. Because a liquid can take any shape (i.e., it has no shear strength), it has no tendency to straighten out if it is deformed as in Figure 4.16b, so S-waves cannot travel through a liquid; however, P-waves can because a liquid resists compression. What determines the velocities of P- and S-waves is explained quantitatively in Box 4.1.

Both P- and S-waves are generated by most seismic sources. They are also produced by **wave conversion** at an interface; when, for instance, a P-ray is refracted and reflected by an interface, S-rays are usually also produced (Fig. 4.18a), and similarly for S-rays. To find what directions the various rays take, Snell's Law is used, *for it applies to reflected as well as refracted rays,* using the velocities of P- or S-rays as appropriate. For the S-rays,

$$\text{reflection: } \frac{\sin i_{1P}}{v_{1P}} = \frac{\sin i_{1S}}{v_{1S}}$$

$$\text{refraction: } \frac{\sin i_{1P}}{v_{1P}} = \frac{\sin r_{2S}}{v_{2S}} \qquad \text{Eq. 4.9}$$

The refracted P-wave has been dealt with earlier, while, of course, the **angle of reflection** of a P-wave is the same as that of the incident P-ray but on the opposite side of the normal. If the velocities were as in Figure 4.18b, putting them into the equations would give the angles shown.

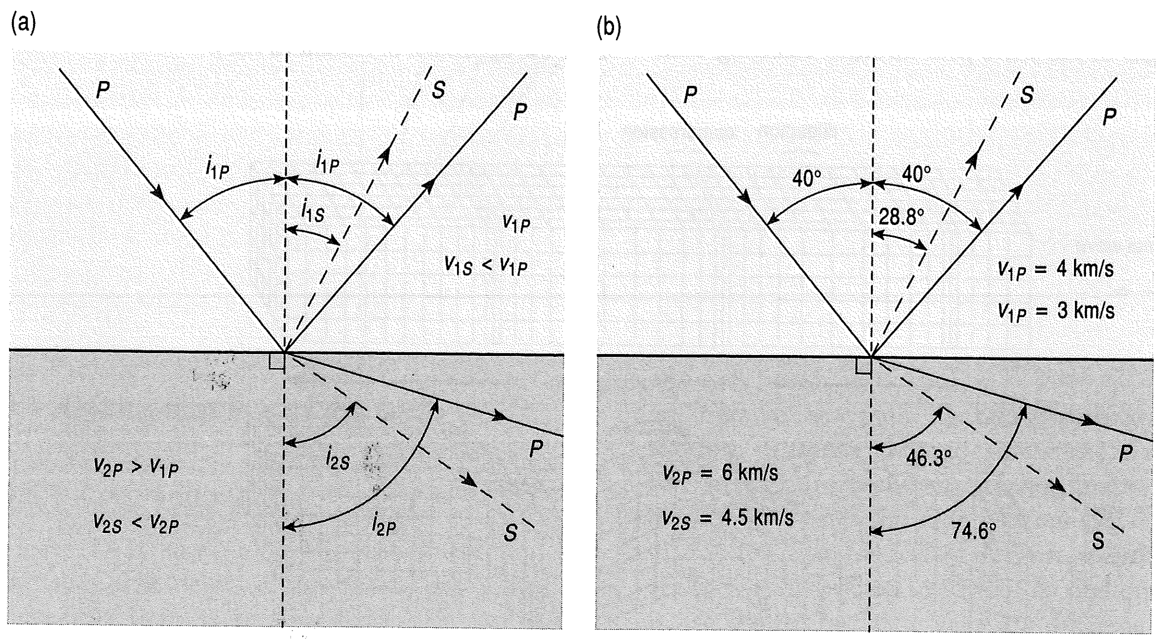


Figure 4.18 S-rays formed by conversion of P-rays.

Since the velocity of S-waves is always less than that of P-waves in the same medium (typically v_s is 0.55 v_p), the angles of S-rays are always smaller than those of the corresponding P-rays. The amplitudes of converted rays decrease towards zero as the angle of incidence, i_{ip}, decreases.

BOX 4.1 Elastic moduli and seismic velocities

Seismic velocities

In Section 4.5.2 it was explained that P- and S-waves have different velocities because they deform a rock in different ways, which are resisted by different elastic restoring forces. A wave propagates along a spring (e.g., Fig. 4.1), because when a turn is displaced the tensional forces from left and right become unequal and the imbalance tends to return it to the stationary position, as was explained earlier. The speed of propagation depends upon two quantities: (i) the stronger the spring, the greater the out-of-balance force produced by a given displacement and so the greater the propagation velocity; but as this force has to accelerate each turn, so (ii) the more massive the turns, the slower the acceleration, since force = mass × acceleration. For waves in general, the velocity increases as the restoring force from a given deformation increases, but decreases as the mass increases:

$$\text{velocity of waves} = \sqrt{\frac{\text{restoring stress}}{\text{appropriate mass}}} \qquad \text{Eq. 1}$$

For waves through an extended material such as rock, the appropriate mass is its density (the mass per unit volume), measured in megagrams per cubic metre (Mg/m³). This is the same for P- and S-waves, for they are travelling through the same rock, but the forces are different for the two types of waves.

Transverse waves involve just a change of shape, from square to lozenge shaped (Fig. 4.16b); to force a cube of rock to be lozenge shaped needs a shear force, and the size of the force depends on the shear, or rigidity, modulus μ (explained later in this box). But as a longitudinal wave causes changes of size as well as shape (Fig. 4.16a), the compresssibility modulus, κ, as well as the shear modulus, is involved. The expressions for their velocities are:

velocity of longitudinal, or P-, waves:

$$v_P = \sqrt{\frac{\kappa + \frac{4}{3}\mu}{\rho}}$$

velocity of transverse, or S-, waves:

$$v_S = \sqrt{\frac{\mu}{\rho}} \qquad \text{Eq. 2}$$

ρ is the density.

In a liquid μ is zero, so v_s is also zero. Because the velocity of longitudinal waves depends upon the value of the compressibility modulus as well as of the shear modulus, v_p is not zero in a liquid, and is always faster than v_s.

Compressibility and rigidity moduli

Elastic modulus is defined as the ratio of a stress to the strain or deformation it produces:

$$\text{modulus} = \frac{\text{stress}}{\text{strain}} \qquad \text{Eq. 3}$$

The higher the value of the modulus, the stronger the material, and the smaller is the strain produced by a given stress. For each different way of straining a material there is a different elastic modulus. Two are relevant to seismic waves.

The compressibility, or bulk, modulus, κ, is a measure of how much force is needed to change the *volume* of the material, without change of shape. This could be done by putting a sphere of the material in a liquid (Fig. 1a): Increasing the pressure of the liquid causes the sphere to shrink. The stress is just the pressure, P – the force on each unit area of the sphere's surface – while the strain is the resulting proportionate decrease of volume, $\delta V/V$.

$$\text{compressibility modulus, } \kappa = -\frac{P}{\left(\frac{dV}{V}\right)} \qquad \text{Eq. 4}$$

The minus sign is to keep κ positive, for an increase in pressure produces a decrease in volume. The actual size of the sphere does not matter. This measurement could be carried out for a liquid or even a gas contained in a balloon.

The shear, or stiffness or rigidity, modulus, μ, is a measure of the effort needed to change the *shape* of a material, without change of volume (Fig. 1b). Equal and opposite forces, F, are applied to top and basal faces of a block with area A, tending to change its rectangular section into a lozenge. The strain is $\delta x/w$, which equals the (small) angle $\delta\theta$. The pair of opposite and separated forces F form a couple, or torque, equal to $F \cdot w$ (F multiplied by w). Since doubling the thick-

BOX 4.1 Elastic moduli and seismic velocities *(continued)*

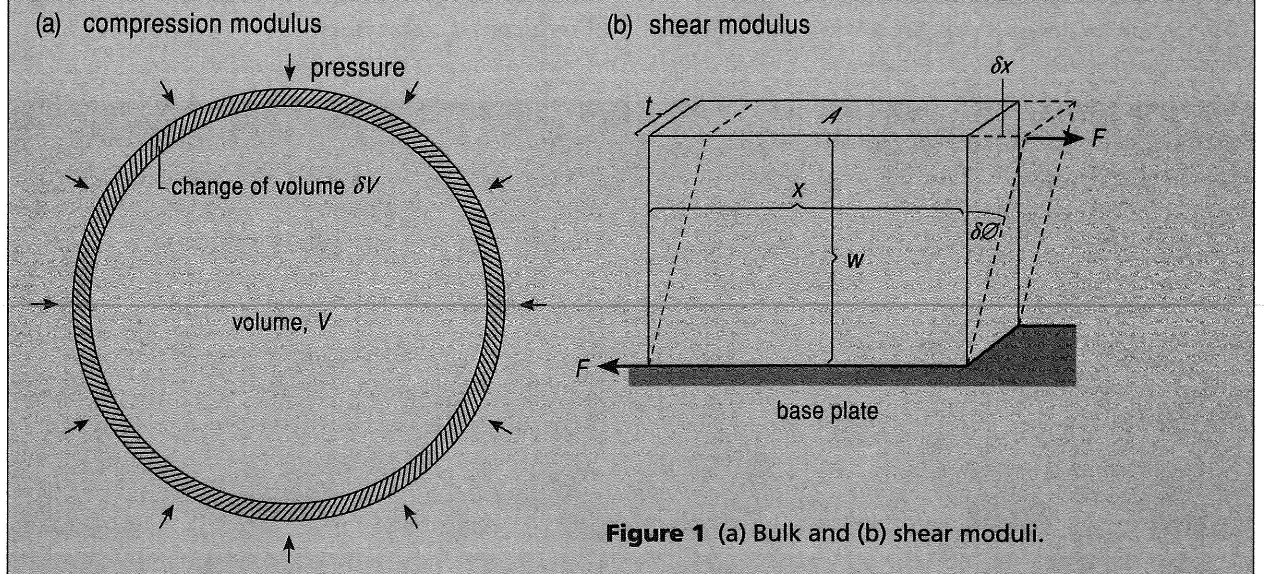


Figure 1 (a) Bulk and (b) shear moduli.

ness, t, or doubling the length, L, of the block would each require the force F to be doubled to produce the same strain, we need to measure the torque when the area A is unity, so the stress is F/A. Hence,

$$\text{rigidity modulus, } \mu = \frac{(F/A)}{d\theta} = \frac{F}{A\,d\theta} \qquad \text{Eq. 5}$$

As it takes no effort to change the shape of a liquid, its shear modulus is zero.

4.5.3 The mantle–core difference

The relevance of P- and S-waves is that they reveal that the core–mantle boundary is a change from solid mantle to liquid core, for S-waves do not travel through the core, as will be explained further in Section 4.5.6. A change from solid mantle to liquid core also helps account for the value of v_p being less at the top of the core than at the base of the mantle (Fig. 4.15a), for v_p is reduced if μ is changed to zero (Eq. 2 of Box 4.1). However, the core being liquid does not rule out it being made of a different material as well; in fact, the next question is why a liquid should occur below a solid, deep within the Earth. From arguments involving meteorites and theories of formation of the Earth (see, e.g., Brown and Mussett, 1993) we believe the mantle is composed of crystalline silicates while the core is predominantly of molten iron, which settled to the centre of the Earth because of its high density; the iron has a lower melting temperature than the silicates, and the temperature at the core–mantle boundary is between the melting points of the two materials.

4.5.4 Other seismological features of the Earth

The shallowest significant feature on Figure 4.15a, just discernible at this scale, is a small jump at a depth of a few kilometres. This is the **Moho** (the generally used abbreviation for the discontinuity named after its discoverer Mohorovičić), where the P-wave velocity increases abruptly from less than 7 km/sec to more than 7.6 km/sec. This *defines* the boundary between **crust** and mantle. The Moho exists over nearly the whole globe but varies in depth from 5 to 10 km (average 7) under the floors of the oceans to 70 km or more under the major mountain chains, with 40 km the average under continents. The term 'crust' suggests a solid layer over a liquid one, as in pie crust, but we know that crust and mantle are both solid, for S-waves travel through them.

About 100 km down is the **low-velocity zone** (LVZ), not a sharp change of velocity but a *decrease* over an interval before the general increase of velocity with depth resumes. Its lower boundary is chosen to be the depth where the velocity regains the value of the upper boundary (Fig. 4.19). The LVZ varies in thickness and in its velocity decrease, and it is not found at all under old continental cratons. Its cause is discussed in the next section.

The next important boundary is the 400-km discontinuity, where velocity increases abruptly. It is almost certainly due to a phase change, a reorganisation of the olivine and pyroxene – believed to be the dominant minerals at this depth – to more compact forms due to the increase of pressure with depth, just as graphite converts to diamond at sufficiently high pressures. Further down is the **660-km discontinuity,** whose nature is more controversial. Very probably it also is a phase change, to a yet more compact crystalline form, but there may be a small change of composition as well, also increasing the density. The region between 400 and 660 km is the **transition zone,** and it forms the lower part of the **upper mantle,** which is separated by the 660-km discontinuity from the **lower mantle.** There are probably other, smaller, discontinuities in the upper mantle.

The P-wave velocity increases smoothly with depth in the lower mantle, except for the bottommost 200 or so kilometres, where is it almost constant. This is called the D″ layer (after a former scheme for labelling the various layers of the Earth alphabetically). This may differ from the mantle above in having both higher temperatures and a somewhat different composition.

The core is divided into outer and inner parts by yet another discontinuity. The **inner core** is solid, probably mostly iron below its melting point at this depth. It exists within the liquid outer core, also predominantly of iron, mainly because the melting temperature of the **outer core** is lowered by the presence of other elements, such as sulphur.

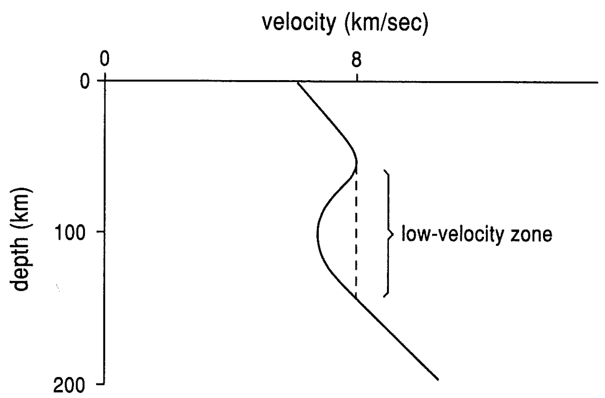


Figure 4.19 Low-velocity zone, LVZ.

4.5.5 Attenuation

The amplitude of seismic waves changes for two main reasons. One is that the wave front usually spreads out as it travels away from the source and, because the energy in it has to be shared over a greater area, the amplitude decreases. (Occasionally, energy is concentrated by reflection or refraction at interfaces and then the amplitude increases. If the waves come from a nearby earthquake, such concentration may increase the damage they do; see Section 5.10.1).

The second is when some of the wave energy is absorbed. This occurs if the rock is not fully elastic. A simple example is when waves enter unconsolidated sands: The sand grains move semi-independently, so the sand does not spring back to its original shape as a wave passes through. In loose sand the waves rapidly die away, but in partially consolidated sand their amplitude decreases progressively; this is called **attenuation.** We have seen that S-waves will not travel through liquids, though P-waves will; if a rock contains some liquid distributed through it S-waves will be noticeably attenuated, though the P-waves will be less affected. This is believed to be the reason for the low-velocity zone, for it is much more noticeable for S- than P-waves, and it occurs at a depth at which there is likely to be a few percent of partial melt; that is, between the crystalline grains of the rock is a few percent of liquid. Attenuation can therefore be used to map the presence of magma beneath volcanoes.

4.5.6 Ray paths in the Earth

The main ray paths through the Earth are summarised in Figure 4.20. A ray is named according to the parts of the Earth it travels through (the crust is ignored on this scale) and by whether it is a P- or an S-ray. A P-ray in the core is called K; thus a P-ray travelling successively through mantle, core, and again mantle is called PKP (sometimes abbreviated to P′). A reflection from the core is denoted by c, so a P-ray reflected back up from the core is PcP; if converted to an S-ray

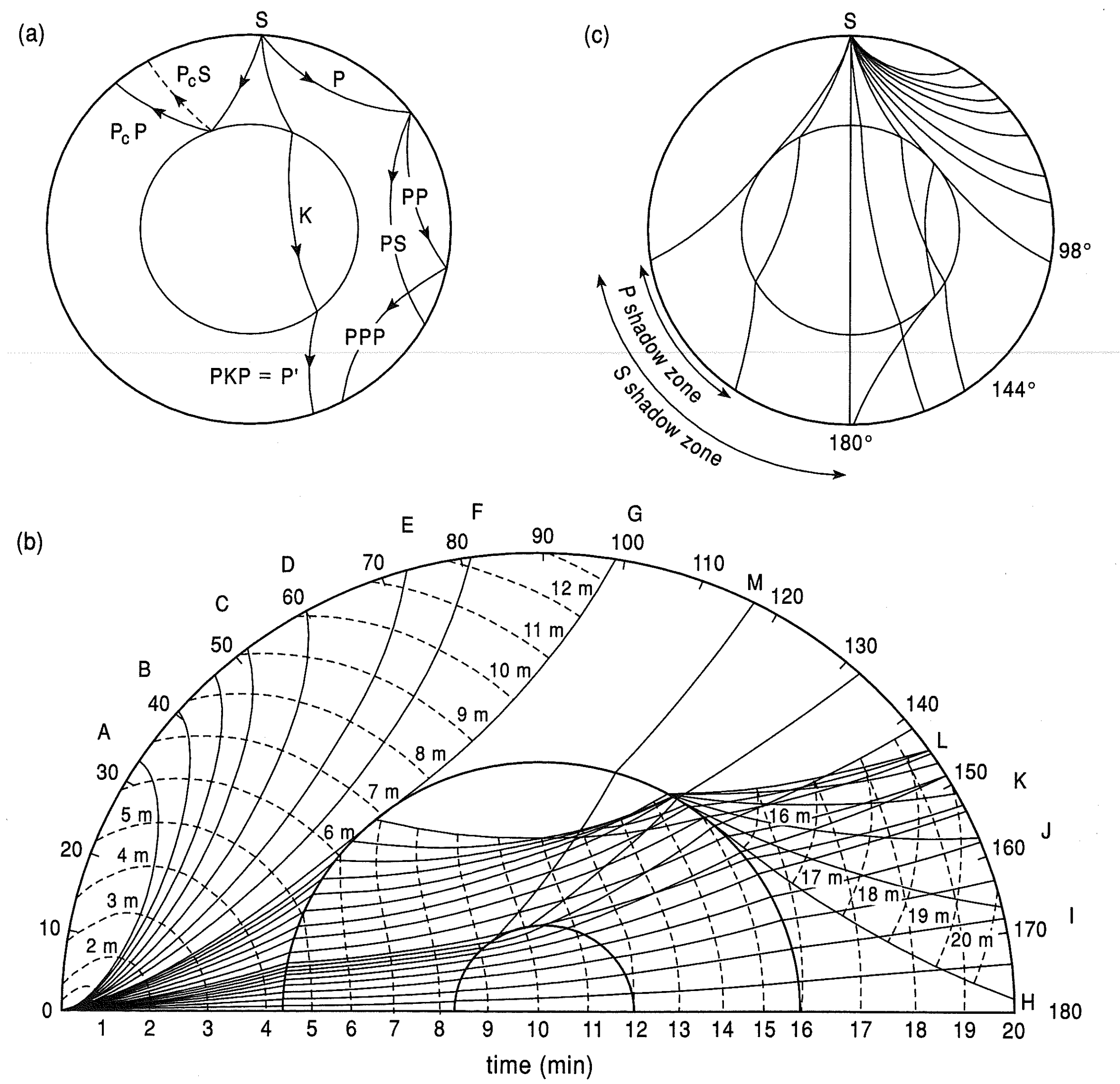


Figure 4.20 Ray paths in the Earth.

it is P_cS. A P-ray reflects from the underside of the surface to become a PP or PS ray. And so on. Arrivals from different paths can usually be recognised on seismograms and are referred to as phases.

Figure 4.20b, as well as showing ray paths, also shows wave fronts spaced 1 min apart, so that you can tell how long it takes for rays to travel various epicentral angles. For instance, it takes a little over 9 min for a seismometer 50° from an earthquake to receive its first arrival; going straight through the Earth takes a little over 20 min.

Let us follow the paths of rays leaving the surface at progressively steeper take-off angles. At first they travel in simple arcs (ignoring the small effects of the 400- and 660-km discontinuities), each going a little deeper and further than the previous one, *A, B, C*, . . . but beyond about 98°, *G*, they encounter the core and are refracted by it (reflected rays are not shown on this diagram). The ray that just intercepts the core is refracted into the core and emerges at 178°, *H*; actually, a *cone* of rays leaves the source with a given angle and so the rays arrive along a circle 2° around the antipodes to the source. With yet steeper rays the arrival point moves away from the antipodes, *I, J, K*, . . . as far as 144°, *L*, and then moves back towards the antipodes. There are no main P-ray arrivals in the interval 98° to 144°; this is called the P-ray (or wave) **shadow zone** (Fig. 4.20c).

There are no S-ray arrivals beyond 98°, so the S-ray shadow zone extends all the way to the antipodes.

But there are some, weaker, arrivals between 98° and 144°, because – as well as ones reflected into it, such as PP – the inner core reflects some rays into it (e.g., *M*); the inner core was discovered by detecting them.

Rays that arrive at a distance greater than 18° (2000 km) from their source are termed teleseismic rays. They are important for investigating the deeper parts of the Earth, not only because they travel deeply, but also because the rays travel up to the surface at a fairly steep angle and so spend little time in the variable crustal rocks that would affect their travel-times in an unknown way.

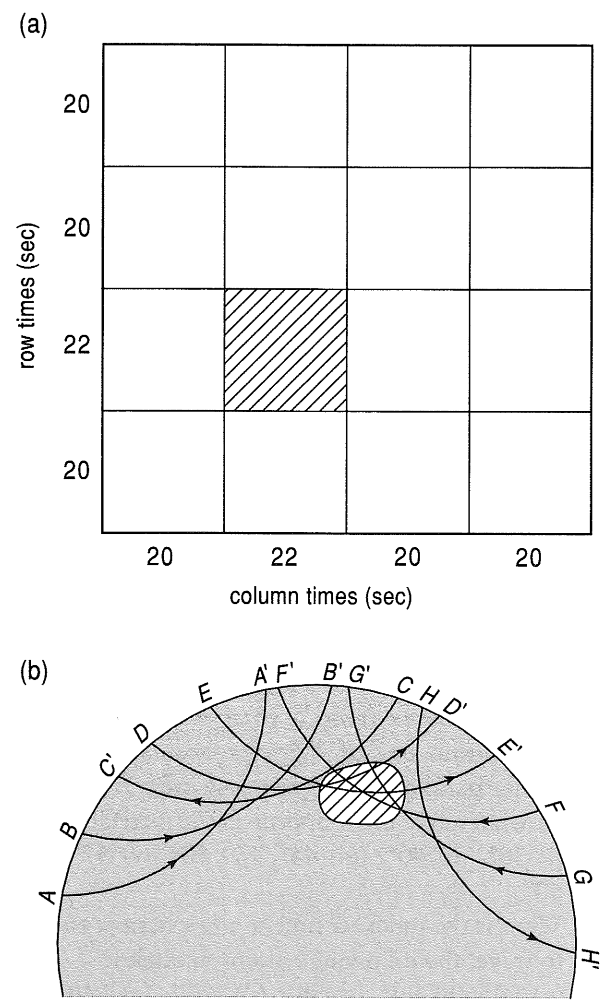


Figure 4.21 Simple tomographic example.

4.6 Seismic tomography

You may know the term 'tomography', perhaps in connection with the medical scanning technique of CAT (computer-aided tomography). **Tomography** is a general method for investigating interiors, such as human bodies and the Earth, whether or not they have internal interfaces. We introduce it with a simple example, shown in Figure 4.21a. Suppose seismic rays travel either along the rows or the columns. The times taken are shown and all are 20 sec, except for the two 22-sec times. Clearly, all the squares are the same and take 5 sec to cross except the shaded one, which takes 7 sec.

The corresponding situation in global seismology is illustrated in Figure 4.21b. If we knew that there were only one anomalous volume, these ray paths would be sufficient to find it. In the actual case, where there may be anomalous volumes at various depths, we need ray paths that penetrate to different depths, and computers are needed to disentangle the effects of more than one anomalous volume. Even with records of many thousands of earthquakes recorded at many seismic stations, it is still only possible to recognise anomalies that are hundreds of kilometres across or, in the deep mantle, thousands of kilometres. The differences of seismic velocity from the average for the depth are small, typically a few percent in the top few hundred kilometres, decreasing to less than 1% in the lower mantle, with differences in v_s greater than those for v_p.

The most successful results so far have been for the upper mantle, to investigate the deep structure of the Earth associated with plate tectonics (Chapter 20). An example is given in Plate 1.

Tomography is important because it can detect variations of seismic velocity, lateral or vertical, that are not separated by sharp interfaces.

Summary

Items within [ ] refer to material in a box.

1. Waves are cycles of displacement that propagate through solids, liquids, and so on. They have a wavelength λ, a frequency f, and a velocity v, related by the equation $v = \lambda f$. A ray is the path followed by a small part of a wave front. Rays and wave fronts always intersect at right angles. A pulse is a very short 'burst' or train of waves.

2. Waves can be longitudinal or transverse, depending whether particle motion is in the direction of wave propagation, or at right angles to it. Longitudinal (push–pull) seismic waves are called P-waves, transverse (shear or shake) ones S-waves.

   P-waves always travel faster than S-waves in a given rock. S-waves cannot travel in liquids. [Velocities depend on the compressibility and rigidity moduli and the density.]
3. Rays reflect and refract at interfaces according to Snell's Law (Eq. 4.5).

   Rays incident obliquely on an interface produce both P- and S- reflected and refracted rays, by conversion. Snell's Law applies to all, provided the appropriate velocities are used.
4. As a ray refracts through the Earth, the p parameter, $(r \sin i)/v$, remains unchanged. It can be used to trace rays through the Earth if the velocity–depth variation is known.
5. The Earth is close to spherical symmetry (after allowance for the equatorial bulge) but is concentrically layered.
6. The main seismic features of the Earth are, at increasing depths: the Moho, separating the crust from the mantle; the low-velocity zone (LVZ); discontinuities at about 400 and 660 km; the D″ layer at the base of the mantle; the core–mantle boundary (CMB), separating the solid mantle from the liquid outer core; and the outer core–inner core boundary separating outer from solid inner core. Between discontinuities velocity usually increases with depth.
7. Major ray paths in the Earth are P, S, PKP (P′), SKS (S′), PcP, PcS, PP, PS, and so on.
8. Seismic tomography is used to detect anomalous volumes by their effect on the times of different rays that pass through them. It does not require the anomalies to have sharp boundaries.
9. Important terms: wave, wave front, pulse, ray, wavelength, amplitude, frequency, seismic velocity; interface, discontinuity; refraction, reflection; normal, angles of incidence, refraction, and reflection; seismometer, geophone, receiver, seismogram; longitudinal and transverse waves, P- and S-waves, wave conversion; Snell's Law, p parameter; attenuation; take-off angle, epicentral angle, travel-time; crust, Moho, low-velocity zone, transition zone, shadow zone, 400-km discontinuity, 600-km discontinuity, upper and lower mantles, core–mantle boundary, outer and inner cores, tomography.

Further Reading

Doyle (1995) gives a comprehensive account of waves and global seismology, while Bolt (1982, 1993a) give elementary accounts of how seismology has revealed the internal structure of the Earth, and Gubbins (1990) gives a more advanced one. Fowler (1990) gives a brief account.

Problems

For some questions, you need to take data from figures given in the chapter. [ ] denotes questions that draw on material in a box.

1. A swing-door in a wall aligned N–S would swing in response to which components of ground motions?
   (i) N-S, (ii) E-W, (iii) vertical.
2. A ray, travelling in a rock with seismic velocity 3 km/sec, encounters an interface with a rock of 4 km/sec at an angle of 45°. At what angle from the normal does it leave the interface?
   (i) 19°, (ii) 45°, (iii) 60°, (iv), 63°, (v) 67°, (vi) 71°.
3. A ray approaches the interface of the previous question at an angle of only 10° to the interface. It leaves the interface at an angle of:
   (i) 7.5°, (ii) 10°, (iii) 13°, (iv) 48°, (v) 60°, (vi) 80°.
4. If a P-ray arrives at the core–mantle boundary at an angle of 25°, at what angle does it enter the core?
   (i) 14°, (ii) 32°, (iii) 34°, (iv) 46°, (v) 73°.
5. A ray crosses from a rock with velocity 2 km/sec into one of 4 km/sec such that it runs along (parallel to) the interface after refraction. At what angle did it approach the interface?
   (i) 30°, (ii) 40°, (iii) 45°, (iv) 50°, (v) 57°, (vi) 60°.
6. What is the quickest time it takes seismic energy to travel the following epicentral angles?
   (a) 50°, (b) 90°, (c) 98°, (d) 142°, (e) 180°, (f) 183°.
   Select from the following list (times are in minutes):
   (i) 9, (ii) 13, (iii) 14, (iv) 15, (v) 17, (vi) 19, (vii) 19.5, (viii) 20, (ix) 22, (x) 23.
7. What are the least and greatest epicentral angles at which a P′ ray can arrive from its source?
   (i) 0°, (ii) 65°, (iii) 104°, (iv) 143°, (v) 155°, (vi) 169°, (vii) 180°.
8. A rock has v_p = 2.5 km/sec and v_s = 1.5 km/sec. If a P-ray travelling through the rock meets the surface at an angle of 30° (to the normal) at what angle does the S-ray reflect?
   (i) 17°, (ii) 30°, (iii) 56°, (iv) 60°, (v) 73°.
9. A P-ray reflects from an interface as both P- and S-rays. Compared to the angle of reflection of the P-wave, that of the S-wave is:
   (i) The same.
   (ii) Always bigger.
   (iii) Always smaller.
   (iv) Sometimes bigger, sometimes smaller.
10. An underground explosion produces a spherical expansion, so you might expect a seismogram of it to lack an S-wave arrival. But S-wave arrivals are recorded. What might produce them?

[11. A granite has a density at least 2½ times that of water, but sound travels faster through it. (a) Why might one expect a higher density to be associated with a lower velocity? (b) In what other way must the properties of granite differ from those of water?]

12. If the Earth were seismically uniform rays travelling an epicentral angle of 90° would take longer than rays travelling 60° by a factor of approximately:
   (i) 1.2, (ii) 1.3, (iii) 1.4, (iv) 1.5, (v) 1.6.
13. Repeat the previous question but for the real Earth (P-waves):
   (i) 1.2, (ii) 1.3, (iii), 1.4, (iv) 1.5 (v) 1.6.
14. A ray, travelling down through the interior of a spherically layered planet, encounters a layer that extends from 3100 km to 3000 km radius. If the velocities above, within, and below the layer are respectively 10, 11, and 12 km/sec, and the ray was incident to the layer at 40°, then it leaves the layer at an angle of:
   (i) 31°, (ii) 34°, (iii) 48°, (iv) 53°, (v) 57°.

[15. A rock has its rigidity modulus equal to three fourths of its bulk modulus. If melting the rock does not change its bulk modulus or density, the ratio of v_p in the solid to that in the liquid would be:
(i) 0.50, (ii) 0.75, (iii) 1.33, (iv) 1.41, (v) 2.00.]

Chapter 5

Earthquakes and Seismotectonics

'Earthquake' is used for two different things: (i) the shaking of the ground – sometimes destructive – and (ii) the cause of that shaking, usually a sudden displacement, or rupture, of a fault. The rupture usually results when increasing tectonic stress strains rocks beyond the limit of their strength. These two aspects have led to two types of study: earthquakes are studied in their own right, particularly because of their destructive power, while in seismotectonics they are used to help reveal tectonic forces and processes.

This chapter describes the basics of earthquake and tectonic studies, which have much in common: how earthquakes are located, how the orientation of the ruptured fault, the direction of its displacement, and the underlying strain are determined, how the sizes, or strengths, of earthquakes are measured, and how damage is caused.

5.1 What is an earthquake?

If you experience an earthquake, there is shaking of the ground and its consequences: rattling of crockery, bells ringing, and – if the shaking is severe – cracking of walls, collapse of buildings, and so on, a terrifying experience. The shaking is caused by the passage of seismic waves. If these waves are traced back to their source, they are usually found to be due to a sudden movement, or displacement, along a fault. The term **earthquake** can refer to either the shaking of the ground at some location or to the source – perhaps distant – of that shaking.

Suppose there are steadily increasing shear forces on two blocks of rocks separated by an existing fault (Fig. 5.1a), tending, in the example, to move the western block northwards, and the eastern one southwards. Because of friction, there is no movement initially; instead, the blocks are distorted so that lines originally straight across the fault – such as fences and roads – become oblique (Fig. 5.1b). Finally, the strain becomes more than the fault can support, and at its weakest part the fault suddenly slips, and this rupture extends rapidly along part of the fault plane, allowing the blocks on either side of it to 'jerk' into a less strained state. The half-arrows beside the fault in Figure 5.1c show the extent of this sudden displacement, called the **elastic rebound.** The energy that had been accumulating in the strained volume of rock is abruptly released; some of it is converted by friction into heat along the fault, some may go into lifting the blocks (if the displacement has a vertical component), and some goes into seismic waves that travel outwards from the fault.

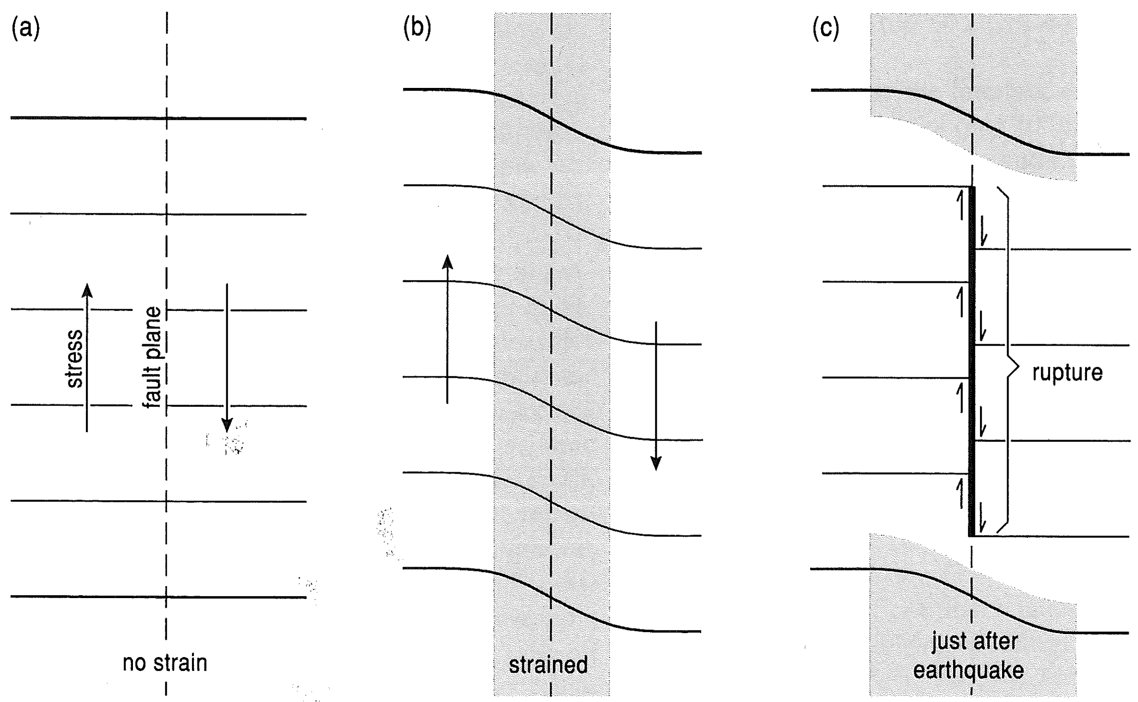


Figure 5.1 Elastic rebound.

A small percentage of earthquakes result from volcanic eruptions (e.g., Section 24.1), filling of reservoirs, mining subsidence, and other causes, but even these often involve displacement of a fault.

5.2 Locating an earthquake

Often we do not see the cause of an earthquake, because it is too deep or too remote or beneath the sea. When no fault displacement is observed, we know neither the location of the earthquake nor the time at which it started, the origin time. However, we can deduce them from seismograms, the traces recorded at a number of seismometer stations, as in Figure 5.2.

Suppose we use only the times of the **first arrivals** (the P-wave arrivals) at different places (Fig. 5.2a). According to the times of arrival, the earthquake was closest to station *A*, furthest from *B*, and so on, so it is somewhere within the shaded area (Fig. 5.2b). This is not very precise, because we don't know how long the signal took to reach the stations and time is not simply proportional to the distance, because the velocity within the Earth is not uniform (Section 4.3.2).

There is a more precise method, that uses the S-wave arrivals as well. As S-waves travel more slowly than P-waves, the more distant the earthquake from the receiver the greater the lag of the S after the P arrival. By matching this delay to standard P and S travel-time curves the *distance* of the earthquake from a given station can be read off (Fig. 5.3). Here, the P–S difference of 6½ min corresponds to an epicentral angle of about 46°. We can also deduce when the earthquake occurred (in the case above, about 8 min before the first P arrival).

As we do not know the *direction* of the earthquake, we can locate the source only on an arc centred on the seismogram station, but if the procedure is repeated for other stations then each distance can be drawn as an arc and their intersection locates the earthquake (Fig. 5.4). Three stations are sufficient to locate an earthquake (if it is on the surface), but usually more are used, to improve the precision of the location and to reveal the size of its error, through the scatter of the intersection positions.

So far, we have assumed that the earthquake is near the surface but it may not be. The point where the seismic waves originate is called the **hypocentre** (or focus); the point on the surface above the focus is the **epicentre.**

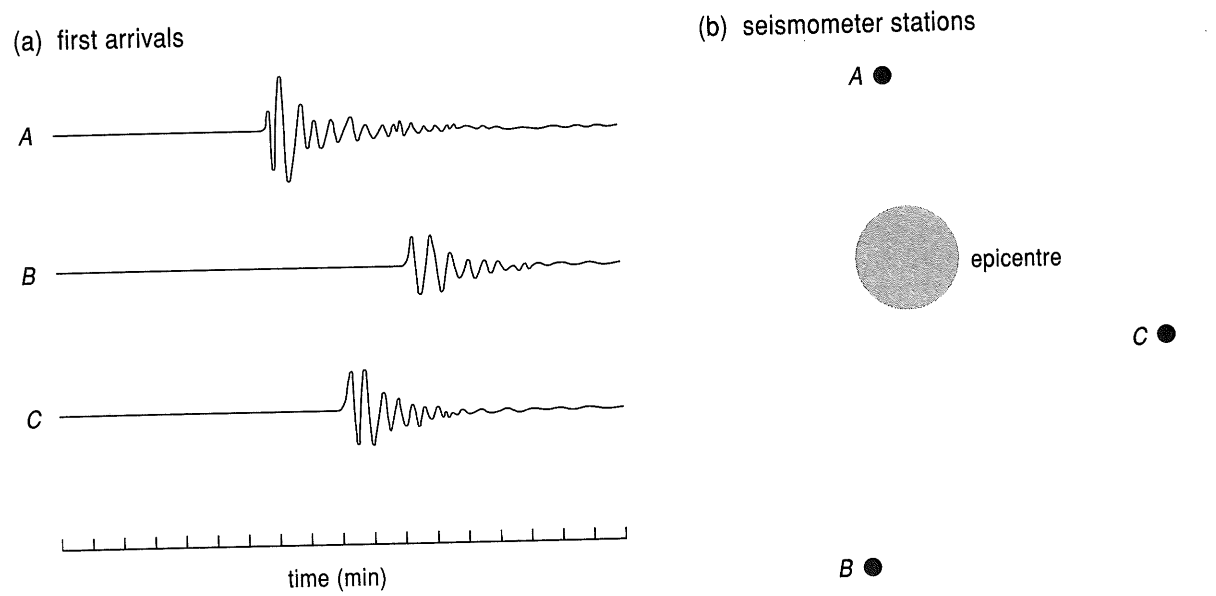


Figure 5.2 Locating an earthquake using first arrivals.

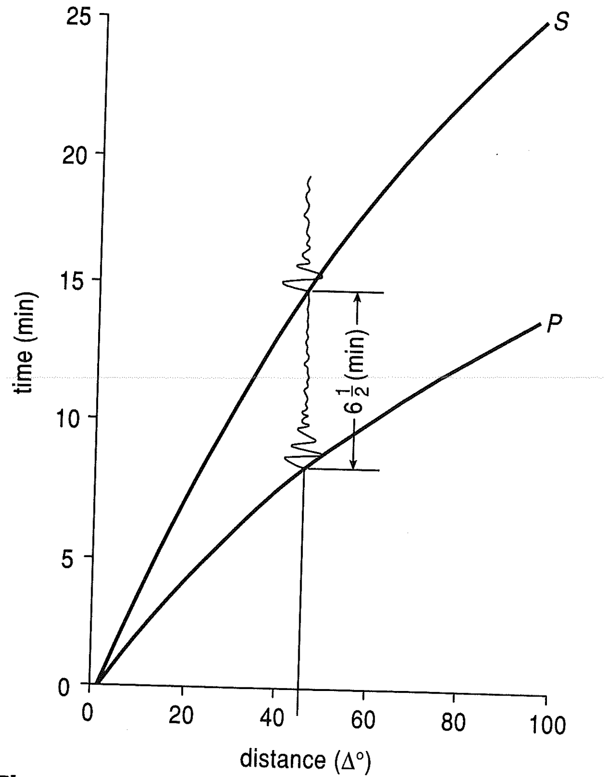


Figure 5.3 Distance of earthquake from difference of P and S arrival times.

The depth of the hypocentre can be found by measuring the difference in arrival of the P-ray – the direct ray – and the ray that reflects from the surface, pP (Fig. 5.5). If a series of earthquakes occurred at progressively greater depths below the same epicentre, the pP–P difference would increase. The differences have been tabled to give depths. (The P–S arrival-time differences are also affected, but there are tables that take this into account.)

5.3 Fault-plane solutions and stresses

5.3.1 Fault-plane solutions

As well as finding where and at what depth an earthquake occurred, it is also possible to make a **fault-plane solution**, which is deducing the orientation of the **fault plane** and the direction of the displacement in that plane. These are found from the directions of the first arrivals at a number of receivers encircling the epicentre. To understand how first motion at a seismometer depends on the orientation of the fault, first consider a peg driven firmly into the ground and struck by a hammer moving, say, northwards (Fig. 5.6a).

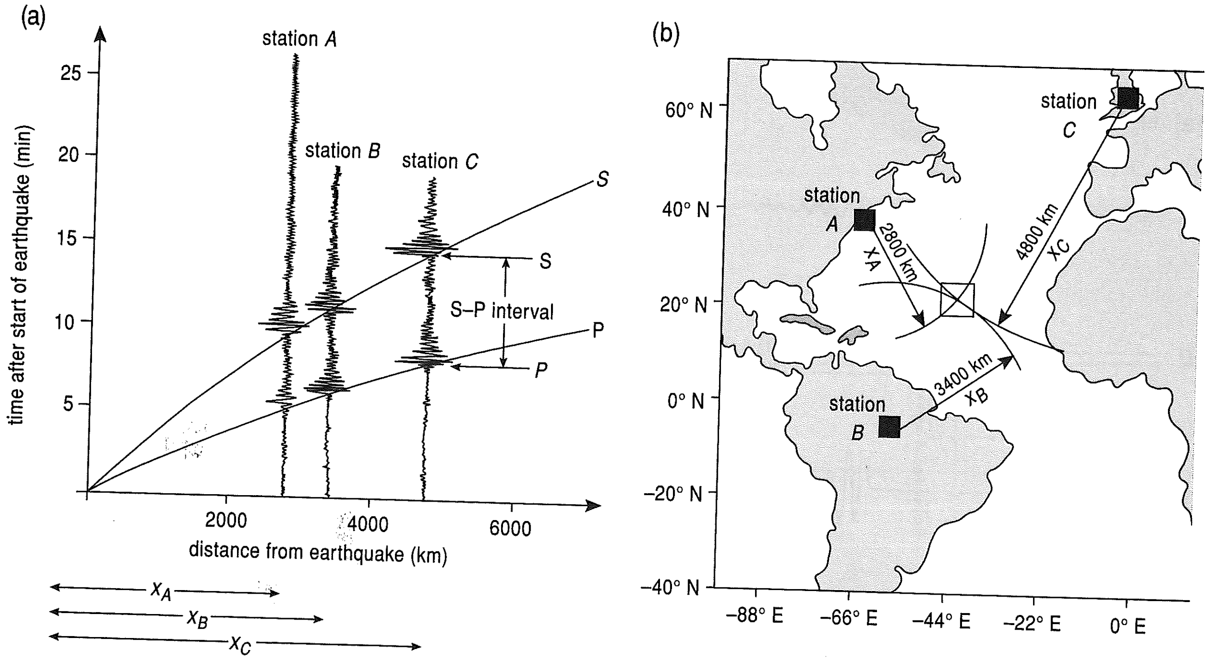


Figure 5.4 Location of an earthquake using P–S arrival-time differences.

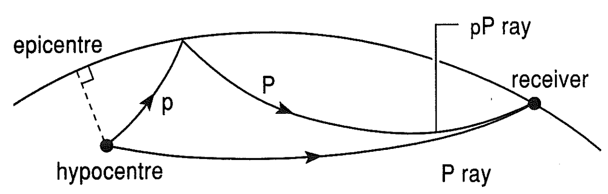


Figure 5.5 Ray paths from a deep earthquake.

Since P-wave oscillations are back and forth in the direction of propagation, there can be no P-waves to either the east or west, but they propagate with maximum amplitude to north and south. In other directions there will be P-waves with lesser amplitudes. Figure 5.6b shows, in plan view, relative amplitudes of the waves by the length of an arrow in the relevant direction; the tips of the arrows trace out two lobes. As the peg is struck towards the north, the *first arrival* or movement of the ground to the north of the peg is a **compression**, while to the south it is a rarefaction or **dilatation**, denoted respectively by + and – signs. S-waves, being transverse, are not propagated to north or south, but have progressively larger amplitudes as the direction approaches east or west, as summarised by Figure 5.6c; however, they are of little help in finding the fault plane.

We next extend these ideas to a fault displacement, which is more complex because the sides move in opposite directions. In the example of Figure 5.7a – a dextral strike-slip fault striking N–S – the amounts of *rebound* are shown by the half-arrows. On the eastern side of the fault the motion of the rebound is similar to that of the peg being struck towards the south, so that – in the eastern half – the P-wave first-motion pattern is the same as in the western half of Figure 5.6b. On the western side, with displacement in the opposite direction, the pattern is reversed.

What is recorded by receivers beyond the limits of the fault rupture? Exactly north and south, on the line of fault (Fig. 5.7a), the effects of the opposite displacements of the two sides cancel and there is no ground movement; nor is there any to east and west, for no waves are generated in these directions. In other directions, the effect of displacement of the further side of the fault is shielded by the fault, so the nearer side has the greater effect. Their combined effects are shown in Figure 5.7b, a *four*-lobed radiation pattern, with alternately compressive and dilatational first motions.

Because the radiation pattern is symmetrical about the N–S and E–W directions, it is impossible to tell from the radiation pattern alone which of these two directions is the fault plane and which a plane perpendicular to it, called the **auxiliary plane;** exactly the same patterns would result from an E–W sinistral fault (far side of fault moves to the left) as for the N–S dextral one (far side to the right) of the example. However, other considerations can usually be used to recognise which is the fault plane; these include (i) the fault plane may be revealed by a surface break; (ii) any aftershocks will be along the fault plane (Section 5.4); (iii) if several earthquakes in an area lie along a line, they are likely to be on a single fault plane rather than on several separate ones; (iv) the trend of faults in the area may be known from geological maps or photographs. Of course, receivers may not be so regularly distributed, and then it may only be possible to locate the fault and auxiliary planes approximately.

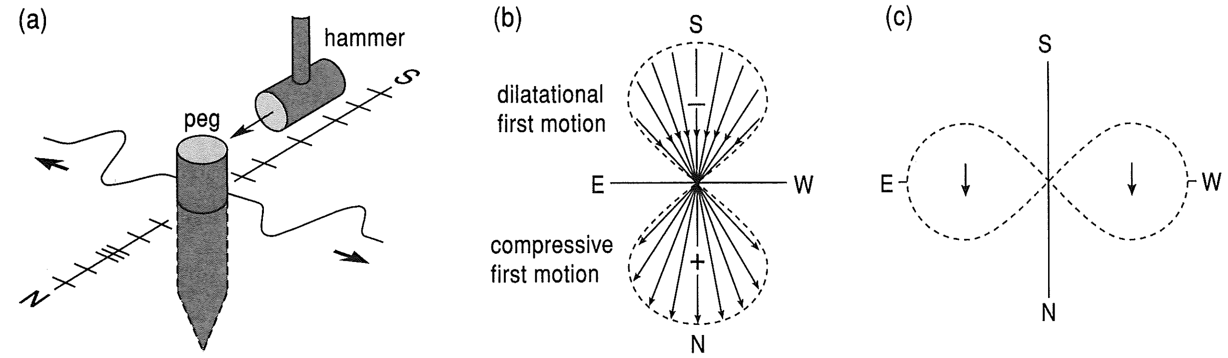


Figure 5.6 P- and S-waves around a struck peg.

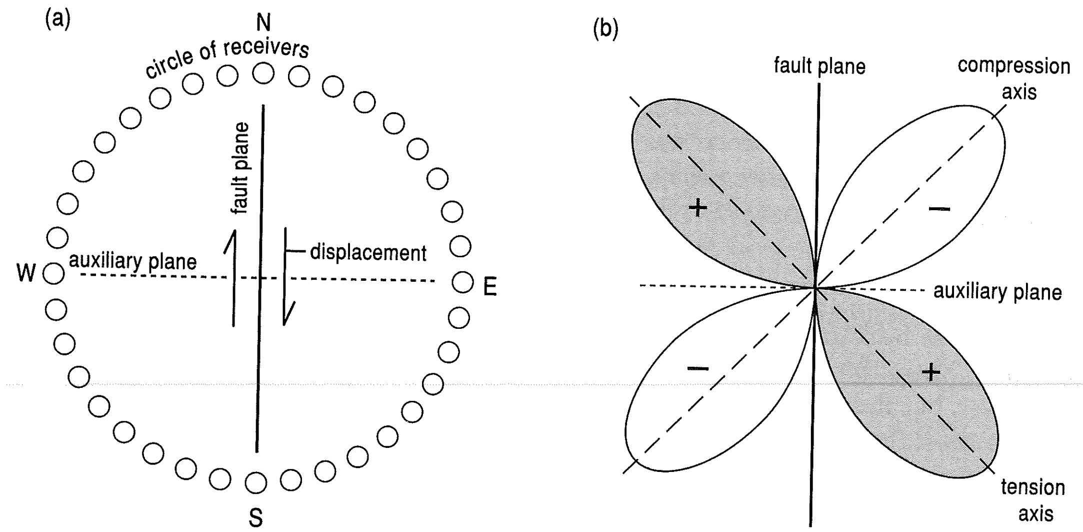


Figure 5.7 Radiation pattern of a fault displacement.

Not all faults are strike-slip, so we need to generalise to all orientations. Since faults are approximately planes and the displacement is in the fault plane, radiation patterns will look the same if viewed along a direction that lies in the fault plane and is perpendicular to the direction of the displacement. Therefore, if we turn the fault plane of Figure 5.7 into another orientation, the radiation pattern will be turned with it. To record the tipped-over radiation pattern we now need not just a circle of receivers but a spherical shell of them (how this is possible will be explained later). The fault and auxiliary planes will cut this **focal sphere,** dividing it into four parts; however, as we use only the lower half sphere, sometimes only three parts appear, as shown in Figure 5.8a, which shows a normal fault striking at 125° (clockwise of north), and dipping at 65°, with displacement entirely down dip. As it is inconvenient to draw spheres, a projection is used, as shown in the right-hand part of the figure; the lines where the two planes cut the focal sphere appear as arcs.

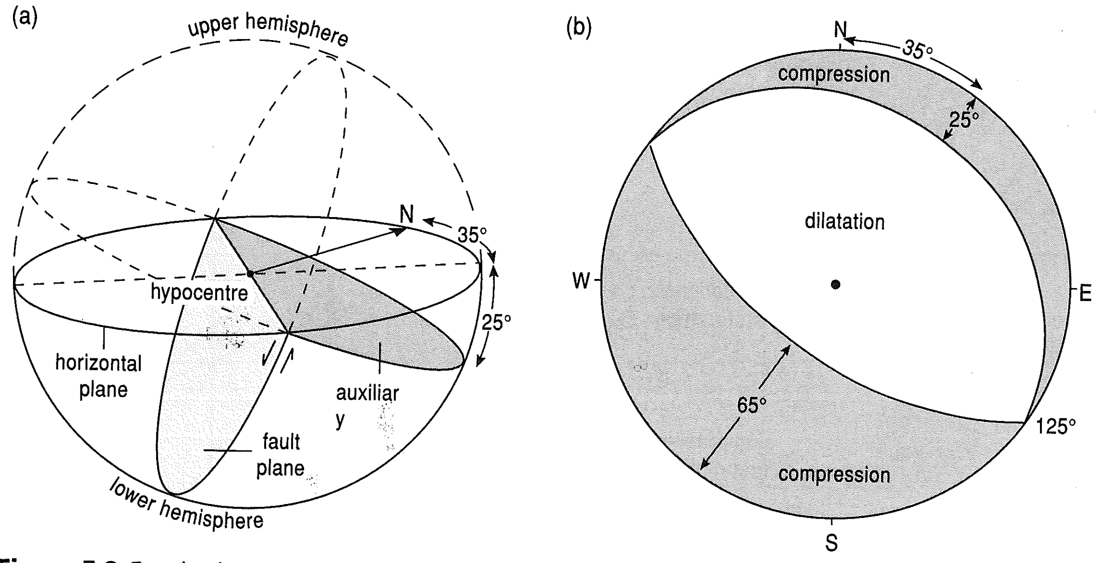


Figure 5.8 Focal sphere.

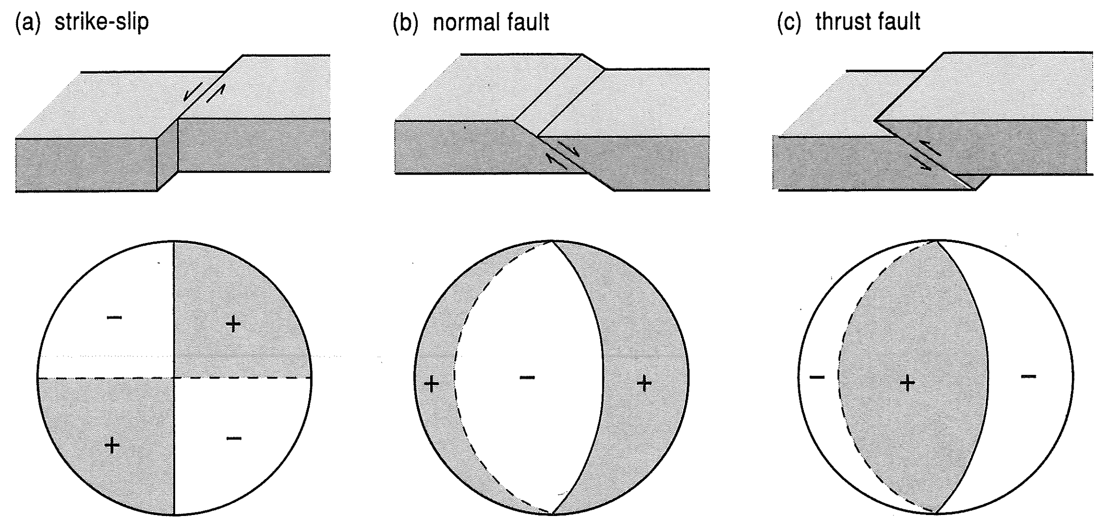


Figure 5.9 Beach balls for various faults.

Because of its appearance, the fault-plane solution is often called a beach ball. If the displacement had been oblique, the intersection of the auxiliary with the fault plane would not be horizontal and the beach ball would show four quadrants.

Different types of fault produce different beach-ball figures. For simplicity, assume that in the following examples the strike of the fault is N–S. The two planes of a strike-slip fault form a +; imagine a sphere centred on Figure 5.7b and the two planes extended downwards to cut it. A normal fault (Fig. 5.9b) has a steep fault plane, which gives an arc passing close to the centre, whereas the auxiliary is nearly horizontal; in the example they are 60° and 30° on opposite sides of the vertical, totalling 90°, of course. Figure 5.9c shows a low-angle thrust fault: It particularly differs from the normal fault in having a *compressional* middle segment, not a *dilatational* one. So a strike-slip fault gives an +, and normal and reverse or thrust faults have the centre point in dilatational and compressional segments, respectively.

When the strike of the fault is not N–S, the beach ball is rotated about its centre to the strike angle (see Box 5.1). When the slip vectors are not as shown above (horizontal for the strike-slip fault, down dip for the other two) the fault plane is unaffected but the auxiliary plane is rotated by the angle the slip vector has been rotated (see Box 5.1).

How is it possible to record the radiation pattern over the lower half of the focal sphere when it is impracticable to place receivers deep in the ground? The P-rays that would reach buried receivers arrive back at the surface somewhere, where they produce compressive or dilatational first motions. Their curved paths can be traced back to find their **take-off angle** (Fig. 5.10), which is used to construct the beach-ball diagram. In practice, it is looked up in a table of take-off angle versus arrival distance.

A step-by-step description of how to find the fault-plane solution is given in Box 5.1. Examples of fault-plane solutions are given in Figure 1 of Box 5.1, and in Figures 20.7 and 20.12.

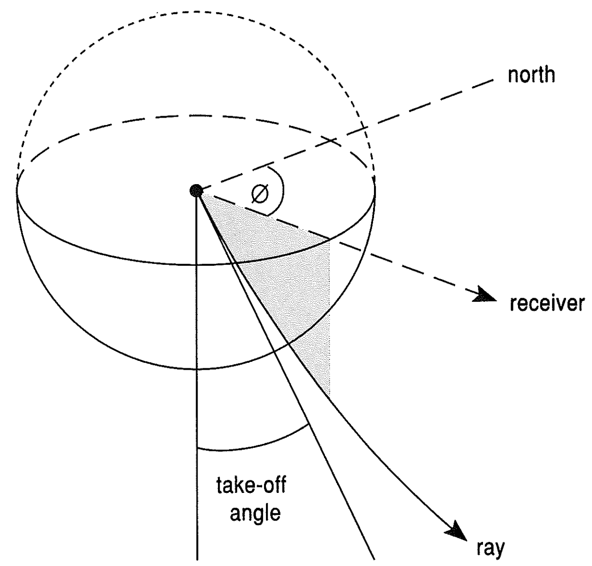


Figure 5.10 Ray path through focal sphere and back to surface.

BOX 5.1 Finding a fault-plane solution

Data needed

The epicentre and hypocentre of the earthquake (found as described in Section 5.2). Then for each receiver,

its bearing (azimuth) with respect to the epicentre,
its take-off angle (using a table that shows angles versus distance),
whether the sense of first motion of P arrivals is compressional or dilatational.

To plot the data

To illustrate how plotting is done, suppose the values of the above quantities for one receiver are azimuth 34°, take-off angle 46°, and dilatational.

(i) Place a piece of tracing paper over a pinpoint protruding through the centre of a Lambert (equal area) projection net (also known as a Schmidt net), and mark the position of north on it. Count clockwise round the net from the north an angle equal to the azimuth value (34°), and make a mark on the circumference. Rotate the tracing paper until this mark is at north; count from the centre towards north an angle equal to the take-off angle (46°) as in Figure 1a.

(ii) Put a small circle at the position, solid for compressional first arrival, open for dilatational.

(iii) Repeat for each receiver.

Suppose the result is as in Figure 1b (some of the data are tabled at the end of the box).

Locating the fault and auxiliary planes

(iv) Rotate the tracing paper (Fig. 1c) to find an arc running from north to south of the stereo net (corresponding to a plane through the centre of the focal sphere) that separates compressional and dilatational points; the compressional and dilatational points will probably 'change sides' along the arc (as shown in Fig. 5.1d). There may be more than one arc that fits, depending on the distribution of points.

(v) Without moving the tracing paper, count 90° along the 'equator' of the stereo net from where the arc intersects it, passing through the centre of the net. Mark this point as *P*, for pole.

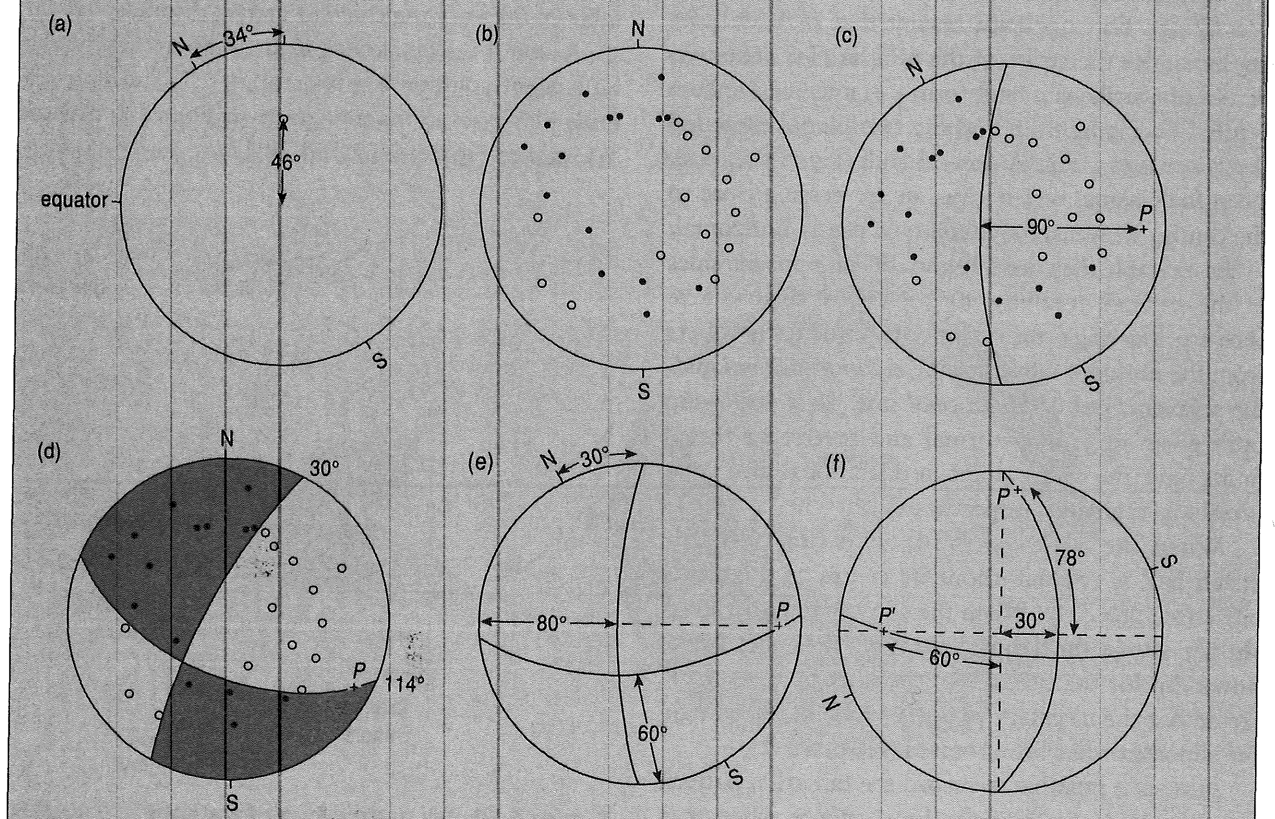


Figure 1 Finding fault-plane solutions

BOX 5.1 Finding a fault-plane solution *(continued)*

(vi) Rotate the tracing paper to find a second arc that separates compressional and dilatational points and *that also goes through P* (this ensures that its plane is perpendicular to the first plane). The two planes are shown in Figure 1d.

(vii) If you cannot find a second plane that separates the compressional and dilatational points, try with another arc that satisfies (iv).

If the points permit a wide choice of planes – because the points are few or poorly distributed – the planes are poorly defined.

Finding the values of azimuth and dip of the planes

(viii) When the two arcs have been found, position the tracing paper with north at the top. Read off the azimuth angles round the rim of the stereo net to the first end of each arc that you encounter (Fig. 1d) (the other end will be 180° further). These are the strike angles of the fault and auxiliary planes (you don't yet know which is which). In the example the azimuths are 30° and 114°.

(ix) Rotate the tracing paper until one of the arcs has its ends at north and south of the stereo net (Fig. 1e); read off the angle from its midpoint along the 'equator' to the edge of the net: this is the angle of dip of the plane (80°).

Repeat for the other arc – giving a dip of 60° in this example.

(x) Recognise whether the solution is for a predominantly strike-slip, normal, or reverse fault. This example is closest to a strike-slip fault but differs significantly from one.

(xii) If the solution is not approximately one of these three types of fault, you need to find the slip angle on the fault plane.

Guess that one arc is the fault plane; rotate the tracing paper until the ends of the *other arc* are at north and south (Fig. 1f); read off the angle from its midpoint to *P*, to give the slip angle (here 78°) on the guessed fault plane (it would be 90° for a pure strike-slip fault). Repeat for the other arc, after finding its pole P', as described above (it should lie on the first arc); it gives 60° for this example. In this example, the fault has either azimuth 30°, dip 80°, with a slip direction of 78° from the vertical, or 114°, 60°, and 60°.

(xiii) Determine which of the two planes is the fault plane, using other information as described in Section 5.3.1.

(*Alternatively*, the stress directions – which are unambiguous – may be deduced: mark on the two possible couples where the two planes intersect and combine them as explained in Section 5.3.2.)

Table 1 Selection of data from Figure 1b

Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*
14	41	C	134	74	C	218	62	D
24	49	D	144	51	C	234	65	D
49	45	D	177	54	C	266	54	D
76	23	D	178	37	C	278	49	C
134	54	D	212	39	C	348	50	C

*C: compressional; D: dilatational

5.3.2 The earthquake stress field and the double-couple mechanism

The directions of the stresses that cause earthquakes can also be deduced from the radiation pattern. The elastic rebound is not confined to just beside the fault rupture but extends out – in lessening amount – on either side. Squares drawn on the surface before the strain had begun to build up would have become lozenges (rhombuses) before the earthquake occurred (Fig. 5.11); these lozenges, of course, are just the tops of columns. To distort a square into a lozenge requires a shear couple (explained in Box 4.1); this could be produced, for example, by pulling on ropes attached to opposite faces (Fig. 5.12a), except that a single couple would rotate the rectangle (Fig. 5.12b).

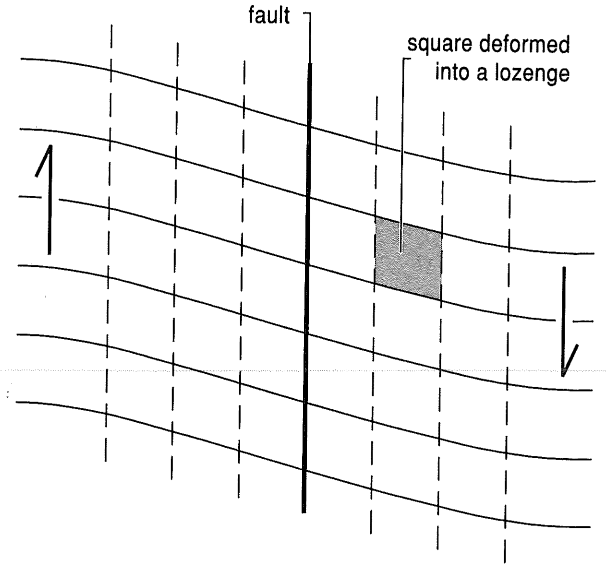


Figure 5.11 Strain just before an earthquake.

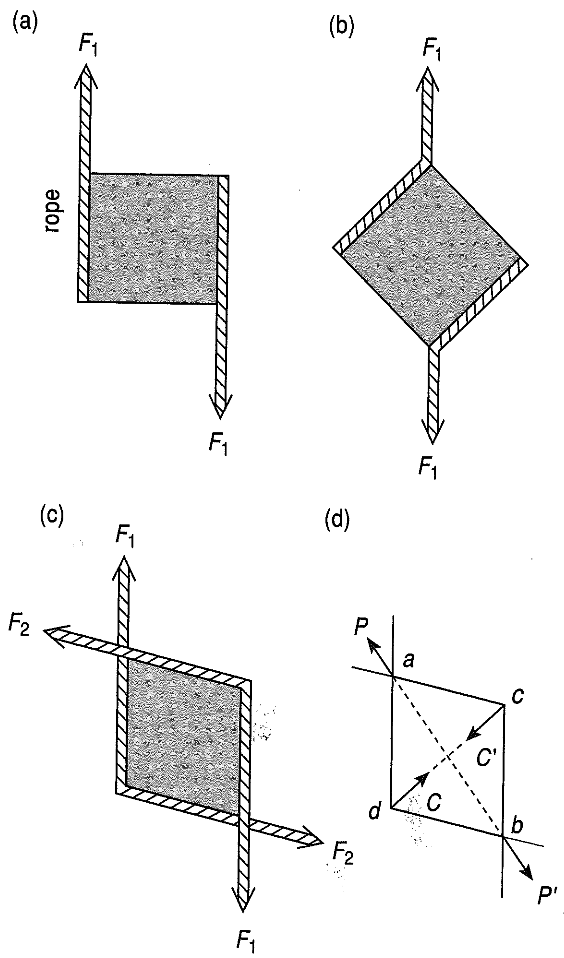


Figure 5.12 Shearing a block without rotating it.

To prevent the rotation requires an equal couple in the opposite sense (Fig. 5.12c). At corner a (Fig. 5.12d) the pulls of the two ropes could be replaced by a single pull, P, and at corner b by P'; similarly, at corners c and d the inward pulls are combined to give compressions C and C'; the two pairs of forces are equivalent to a double couple. For this reason, the system of forces that caused the earthquake is called a **double-couple mechanism.** As the forces of each pair, P and P', C and C', are aligned, there is no twisting of the square. (A few earthquakes do not fit the double-couple mechanism but will not be considered.)

Thus a fault-plane solution can be given either in terms of the fault plane and the sense of the displacement, or as the directions of tension and compression (Fig. 5.13). For some problems it is more useful to think in terms of stresses; an example is described in Section 20.4.1. There is no ambiguity which are the directions of compression and tension, unlike the situation with fault and auxiliary planes.

5.4 Rupture dimensions and displacements

What length of a fault ruptures in an earthquake, how large are the displacements, and how far away from the fault plane does the elastic rebound extend? The ruptured area of a fault due to a large earthquake has a length along strike that is usually much greater than its depth down dip, so we can approximate it by a rectangle, as shown in Figure 5.14.

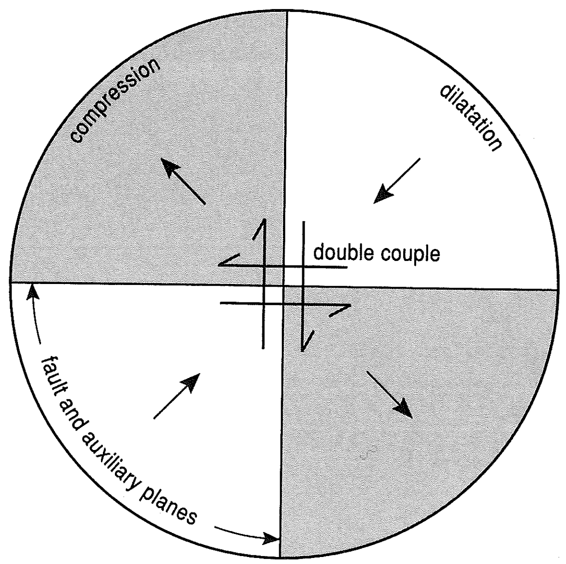


Figure 5.13 Stresses and possible fault planes.

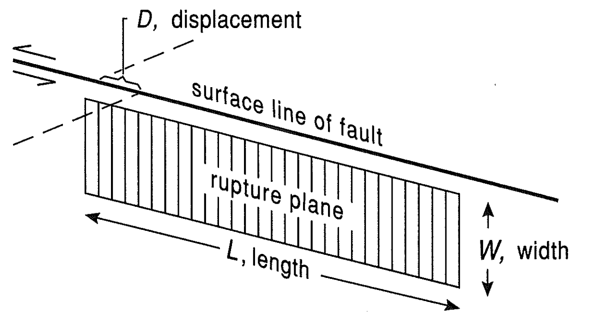


Figure 5.14 Length, width, and displacement of a rupture.

The simplest way to measure the length, L, and displacement, D, of a fault is to look at the newly faulted surface, or fault break, but often ruptures do not break the surface or are under water so indirect means are used, as they must be to estimate the down-dip extent, W, of the rupture.

After large earthquakes there follow many smaller earthquakes, called **aftershocks,** whose numbers dwindle with time. These are believed to reveal the rupture plane because most of them lie on a plane that also includes the main shock, and when there is a visible fault break the aftershock plane is found to match it in position and length. An example is shown in Figure 5.15; most of the epicentres are southwest of the rupture (Fig. 5.15a) because the rupture plane dips steeply to the southwest. There are few aftershocks near the surface (Fig. 5.15b) because the near-surface rocks are poorly consolidated sediments and so are unable to accumulate strain, while below about 12 km the rocks are too ductile to support the elastic strains needed to produce earthquakes. The aftershock plane provides values of L and W (Fig. 5.14).

The extent of strain release can be deduced by comparing accurate maps made just before and after an earthquake. Though maps are unlikely to be made just before an earthquake, earlier maps can reveal the extent of the strained area, and – coupled with measurements of the rebound in a few places – give a good idea of the area and amount of strain release. Figure 5.16 shows contoured horizontal displacements for a large thrust fault, based on measurements at locations shown by the dots. In addition to the above methods, there are other, more sophisticated ways of estimating the dimensions of rupture and its displacement (such as examining the form of the seismic waves radiated in different directions), but these will not be described.

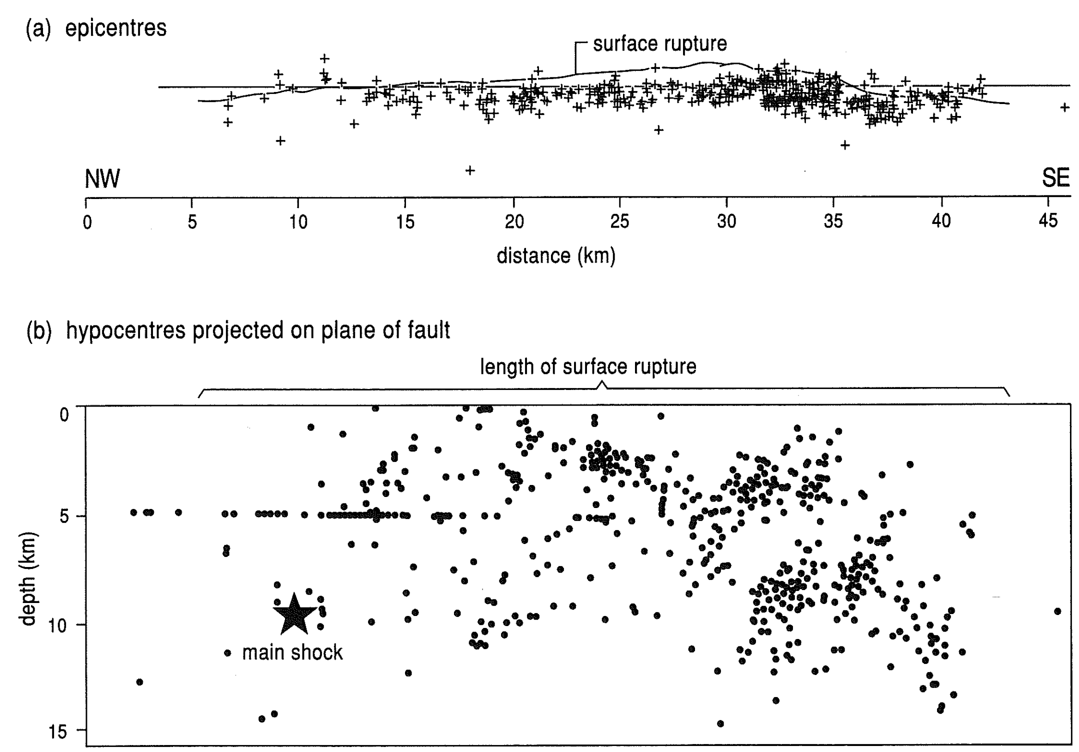


Figure 5.15 Distribution of the aftershocks of the Parkfield earthquake, 1966.

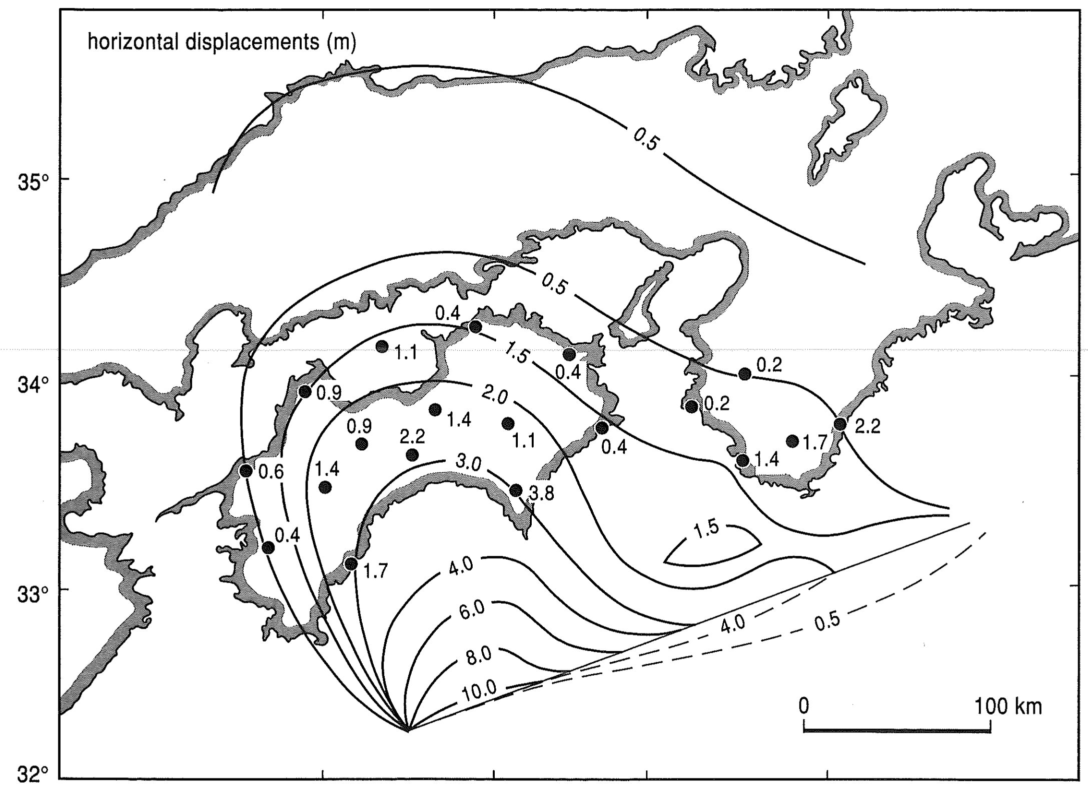


Figure 5.16 Horizontal displacements due to the Nankaido, Japan, earthquake of 1946.

Earthquakes have all sizes of rupture lengths up to hundreds of kilometres (see Section 5.6), widths up to tens of kilometres, and displacements up to tens of metres. The maximum strain varies far less, and has a value of roughly 10^{-4}, that is, the corner of a square in Figure 5.11 moves no more than about one ten-thousandth of the length of a side before there is failure. This is because rocks at depth do not vary greatly in their strength (surface rocks, such as alluvium, are often too weak to accumulate any significant strain, as mentioned above). Therefore, the energy of an earthquake depends mainly on the size of the strained volume – which can be millions of cubic kilometres – not on how much it is strained.

You may be wondering what is meant by the hypocentre of an earthquake, if a rupture can extend for hundreds of kilometres. When strain has built up to the point of rupture, there will be some point on the fault plane that is just a little weaker than the rest, and this is where rupture starts. This puts more strain on adjacent areas, which yield in turn, and failure rapidly spreads out until it reaches parts of the fault plane where the strain is much less. As the rupture spreads more slowly than P- and S-waves, the first arrival at a distant receiver is of waves that originated at the initial point of rupture, and so that is where the hypocentre is located. Waves from other parts of the fault plane arrive a little later, and this is one reason why seismograms of earthquakes are complex.

5.5 Measures of earthquake size

5.5.1 Intensity: Severity of an earthquake at a locality

We need a way of measuring the size, or strength, of earthquakes. There are several ways, but first there is a fundamental distinction to grasp. Just as the term 'earthquake' may refer to either the effects at a locality or to the sudden displacement at the source, so the size can be a measure of the effects at a locality or of the disturbance at the source. We shall be concerned chiefly with the latter but consider the former first.

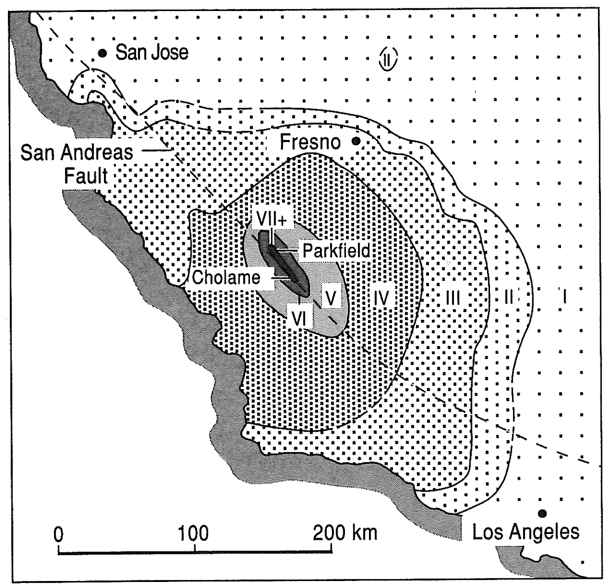


Figure 5.17 Isoseismals for Parkfield 1966 earthquake.

Intensity is a measure of the severity of an earthquake at a locality, judged by the effects produced, such as perceptible shaking, lamps swinging, destruction of masonry, and so on. The scale commonly used is the Modified Mercalli Scale of 1931, which has twelve categories, I to XII. For instance, intensity III includes, 'Felt indoors. Hanging objects swing. Vibration like passing of light lorries (trucks). Duration estimated. May not be recognised as an earthquake'. VI is 'Felt by all . . . many frightened and run outdoors . . .'. Most alarming of all is XII: 'Damage nearly total. Large rock masses displaced. Line of sight and level distorted. Objects thrown in air' (the dedicated geophysicist continues to take observations!).

This may not sound as 'scientific' as, say, recording the local ground motion, but to do this would require having anticipated the earthquake and set up instruments. Instead, it allows the intensity at each locality to be estimated *after* the earthquake, by observing the damage and asking people what they felt and noticed (though it should be done as soon as possible, while memories are fresh and damage unrepaired). If enough data is collected the intensities can be contoured as **isoseismals** (Fig. 5.17). The zone of highest intensity indicates the position of the fault rupture, but only roughly, partly because data may be poor, but also because intensity is partly determined by the local geology; for instance, damage is less to buildings built on solid rock than on alluvium.

Measurement of intensity may not tell us much about the strength of the earthquake at its source, for a small but nearby earthquake may cause as much damage at a locality as a large but remote one. Besides, estimating intensity is time-consuming, not very precise, and limited to places where there are buildings to be disturbed and people to report their observations. A way of measuring the source strength of an earthquake is explained next.

5.5.2 Seismic moment: Size of the earthquake at source

The most commonly quoted measure of earthquake size is the Richter magnitude (explained in Section 5.8) but a later and better measure is the **seismic moment**, M_0. Just before a fault ruptures, the shear forces on either side of the fault (Fig. 5.18) exert a couple, whose size, or moment, equals the product of the shear forces and the perpendicular distance between them, that is, $2Fb$. The force F depends on the strain, the area of the rupture, A, and the rigidity modulus, μ (Box 4.1). The strain depends on the fault offset and the width of the strained volume, and so equals $d/2b$, where d is the average displacement; values are estimated as described in Section 5.4. These are combined as follows:

$$\text{moment of couple} = F \cdot 2b \qquad \text{Eq. 5.1}$$

As $F = \mu A \times \text{strain}$, and $\text{strain} = \dfrac{d}{2b}$,

$$\text{moment of couple} = \mu A d = M_0 \qquad \text{Eq. 5.2}$$

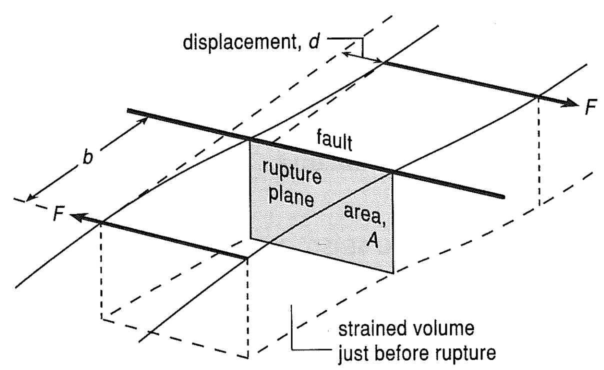


Figure 5.18 Seismic moment.

The rigidity modulus is measured on samples of rock or is estimated from a knowledge of the rocks in the area; as this quantity varies far less than do the dimensions of the rupture, its value need not be known exactly. The seismic moment can also be estimated from the forms of seismograms, useful when there are no aftershock data, but this will not be described.

Seismic moment and the sizes of fault ruptures. Broadly, the larger the rupture, the larger will be the earthquake. The ruptures of small earthquakes do not extend over the whole of the down-dip extent, *W*, of the fault plane (Fig. 5.19a), but above some size of earthquake they do, and then a rupture can enlarge only by increasing its length, *L*, along the fault. As a result, the dependence of seismic moment on rupture length is different for small and large earthquakes (Fig. 5.19b). The relationship also depends on the type of fault. Strike-slip faults do not have so large a depth *W* as thrust faults do, so, to produce a given seismic moment, the rupture is longer, as shown in Figure 5.19c; normal faults lie between ('interplate' and 'intraplate' are explained in Chapter 20). The largest earthquakes have very great rupture lengths; for instance, for the 1960 Chile earthquake, one of the largest in the past hundred years, the displacement, which was along a thrust fault, was 800 km long. This suggests that earthquakes cannot be much larger, for to be so they would have to extend across whole continents.

5.6 Seismotectonics: Deducing tectonic processes

Since almost all earthquakes are due to tectonic processes, they can be used to study tectonic movements and stresses, an application called **seismotectonics.** Earthquakes are particularly useful for revealing hidden tectonic processes, perhaps deep in the Earth, or beneath the oceans. Deductions may be qualitative or quantitative.

5.6.1 Qualitative seismotectonics

We saw in Section 5.3 that fault-plane solutions can be used to deduce fault planes and displacements or, alternatively, tension and compression axes. Qualitative solutions, without the amount of the displacement or the magnitude of the stress, are sometimes all that is needed, and some examples – important for understanding plate tectonics – are given in Sections 20.3 and 20.4.1.

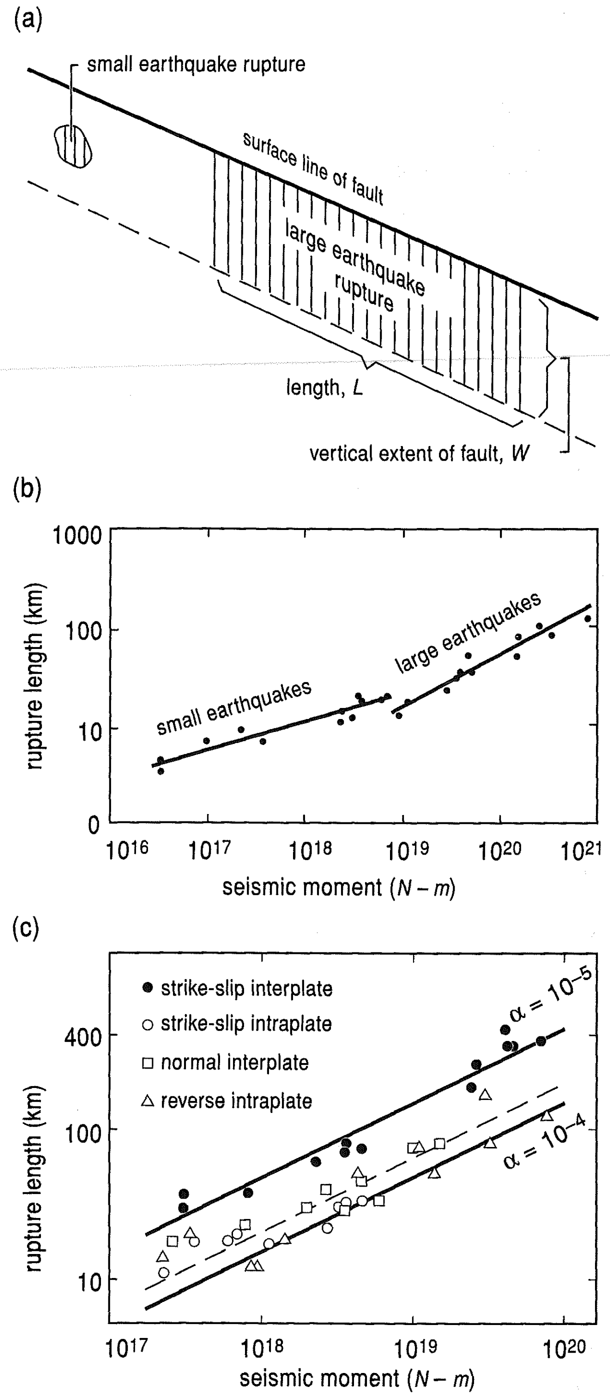


Figure 5.19 Ruptures of small and large earthquakes.

5.6.2 Quantitative seismotectonics: Seismic and aseismic faulting

A different type of problem is to deduce what the average rate of displacement is along a fault, and whether movement occurs simultaneously at all depths. Also important is whether displacements can occur aseismically, that is, without causing earthquakes.

A single earthquake rupture occurs on only part of a fault, so if strain is building up over a much longer extent of the fault – perhaps a fault between two tectonic plates (Chapter 20) – how is it relieved? Over a period of time ruptures occur on different parts of the fault. If the area, *A*, and displacement, *D*, of each rupture are known, they can be added together to give the average displacement, $\bar{D}$:

$$\bar{D} = \frac{1}{L_t W_t}\Sigma AD \qquad \text{Eq. 5.3}$$

where the fault has a total length of L_t and width W_t. If the value of *D* of any earthquake is not known directly, as is often the case, the seismic moment M_0 is used instead:

$$\bar{D} = \frac{1}{\mu L W}\Sigma M_0 \qquad \text{Eq. 5.4}$$

($M_0 = \mu AD = \mu LWD$; in this case, M_0 is found by analysis of the waves recorded on a seismogram, as mentioned above).

This average displacement, $\bar{D}$, can be compared with the displacement found from the long-term offsetting of rivers, roads, fences, and so on that cross the fault, or from surveying. To get a true average, records over a long period may be needed, because the largest earthquakes on a fault may recur only after several hundred years. Results show that on most faults most of the displacement is associated with earthquakes; the remaining displacement is due to aseismic faulting or slippage, which may be steady, have periods of faster and slower movement, or be interrupted by earthquakes. Only on a few faults, or parts of a fault, does aseismic slip account for most of the displacement.

The San Andreas Fault of California (whose origin is described in Section 20.5.2) will be given as an example (Fig. 5.20). It is a dextral strike-slip fault and is the most important of a number of faults that accommodate the northwards motion of about

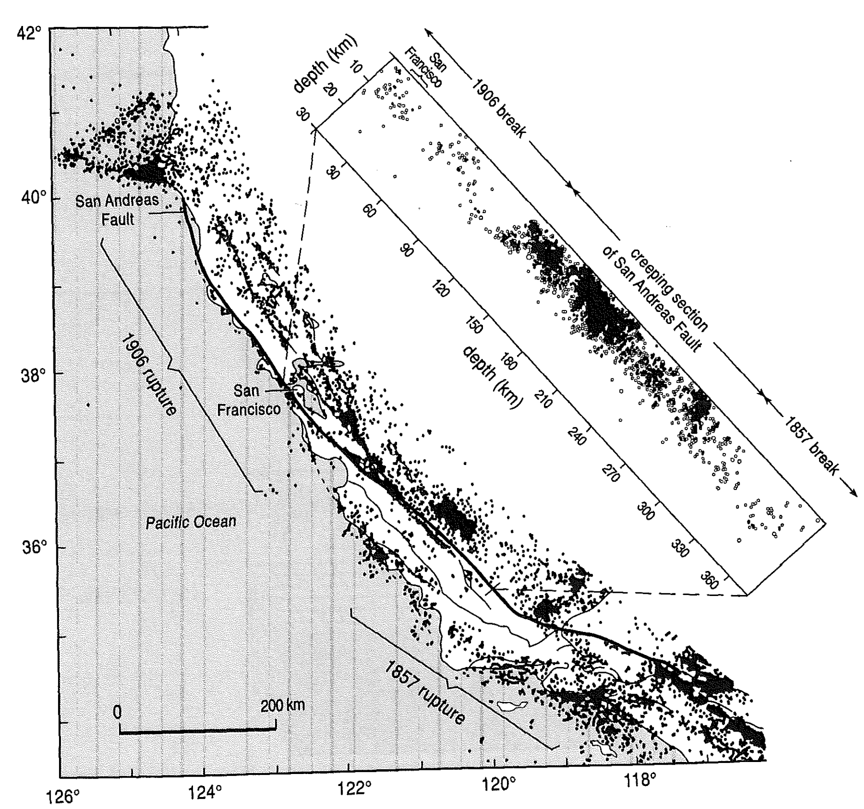


Figure 5.20 San Andreas Fault, California.

50 mm/year of the Pacific relative to the North American Plate (Section 20.5.2). On parts of it there occur large but infrequent earthquakes, notably those of 1857 (Fort Tejon) and 1906 (San Francisco), with ruptures extending for hundreds of kilometres along the fault. These parts of the fault experience negligible aseismic slippage, and also little seismicity between the large earthquakes, as illustrated by Figure 5.20; they are described as being 'locked'. Between these two ruptures is a length with mainly aseismic slip or creep, which has been constant over the 40 years of records, with a maximum movement of about 30 mm a year in the middle of the region; though there are frequent small earthquakes, they account for less than 10% of the slippage. Towards either end of this creeping section are transitional zones where significant earthquakes occur, such as the Parkfield earthquakes of moderate strength, which recur every few decades (one was the earthquake of Fig. 5.15).

The San Andreas Fault, though often described, is not typical. Most faults in continental crust and also those in oceanic crust (except subduction zones, Section 20.4.1) experience no more than a small proportion of aseismic slippage, so estimates using seismic moments will give good approximations of the total slippage.

5.7 Surface waves

In the previous chapter, P-waves and S-waves, which travel through the body of the Earth, were introduced. There are two other types of seismic waves (Fig. 5.21). Both are mainly confined to near the surface, and so are called **surface waves,** in contrast to P- and S-waves, which are termed **body waves.** Water waves are an example of a surface wave: A cork bobbing on the water reveals the amplitude of the waves, but a fish swimming down from the surface would find that the amplitude decreases rapidly (so one way to escape seasickness is to travel deep in a submarine). Similarly, the amplitude of seismic surface waves decreases rapidly with depth (Fig. 5.21) and this characterises surface waves. The amplitude of long waves decreases less rapidly than that of short ones, but in proportion to their wavelengths the decrease is the same; for instance, the amplitude of a 100-m wavelength wave would have decreased at a depth of 100 m by the same fraction as that of a 50-m wave at 50 m. P- and S-waves are called body waves because they can travel through the body of the Earth rather than only near the surface (their amplitudes do decrease, due to spreading out, reflections, and some absorption – as is also the case for surface waves – but not because of distance below the surface).

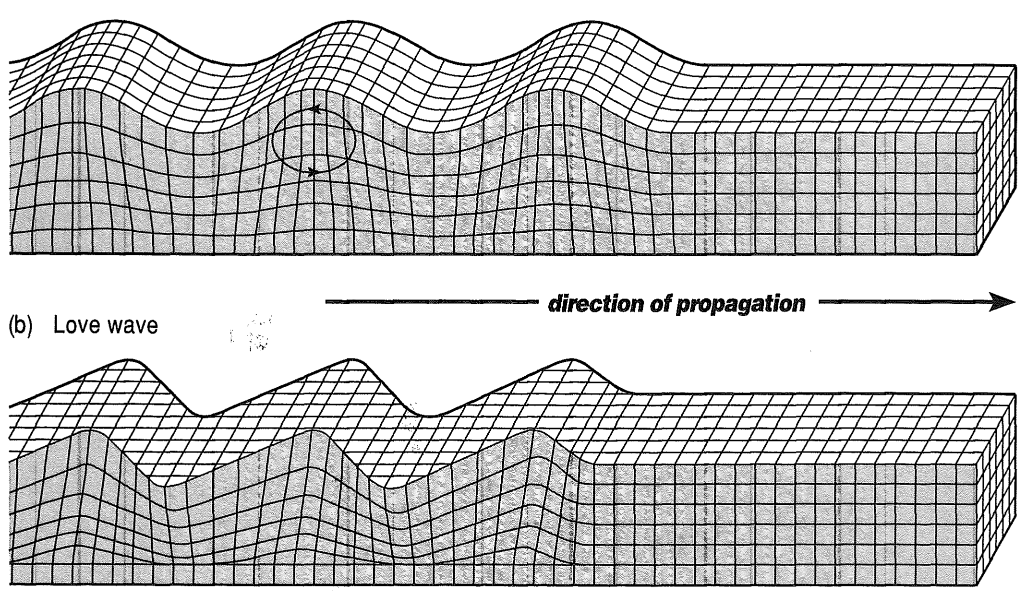


Figure 5.21 Particle motion in Rayleigh and Love waves.

The two types of seismic surface waves are **Rayleigh waves** and – also named after their discoverer – **Love waves.** The main difference between them is that in Rayleigh waves particle motion is a vertical ellipse, whereas in Love waves it is horizontal and transverse (similar to horizontal S-waves) (Fig. 5.21). Therefore Love waves are not detected by a vertical component seismometer, whereas Rayleigh waves are detected by horizontal as well as vertical ones. In both cases, a horizontal component seismometer will only respond to the waves if it has its sensitive direction in the direction of particle motion. In Figure 5.22, the N–S instrument records by the far the largest Love wave amplitudes but the smallest Rayleigh amplitudes; the waves are therefore travelling approximately east or west.

Both types of wave are slower than P- and S-waves – as Figure 5.22 shows – and so are little used for measuring travel-times, but they are important in other branches of seismology. They are generated by most seismic sources and often have the largest amplitude, as Figure 5.22 illustrates; they are responsible for much earthquake damage (Section 5.10.1). They are also used to measure the size of earthquakes, as described in the following section.

5.8 Magnitude: Another measure of earthquake size

Seismic moment (Section 5.5.2) is the preferred measure of the size or strength of an earthquake at its source, but an earlier measure, **magnitude,** is still widely quoted. Magnitude was the first measure of the source strength of an earthquake, and the first magnitude scale was devised by Richter, an eminent seismologist, in 1935, and – with modifications (see Box 5.2) – is the one usually quoted by the media, as the **Richter magnitude.** Richter measured the amplitude in microns (millionths of a metre) of the largest oscillations – the surface waves – recorded by a particular type of seismometer, 100 km from the earthquake source. As the values have a very large

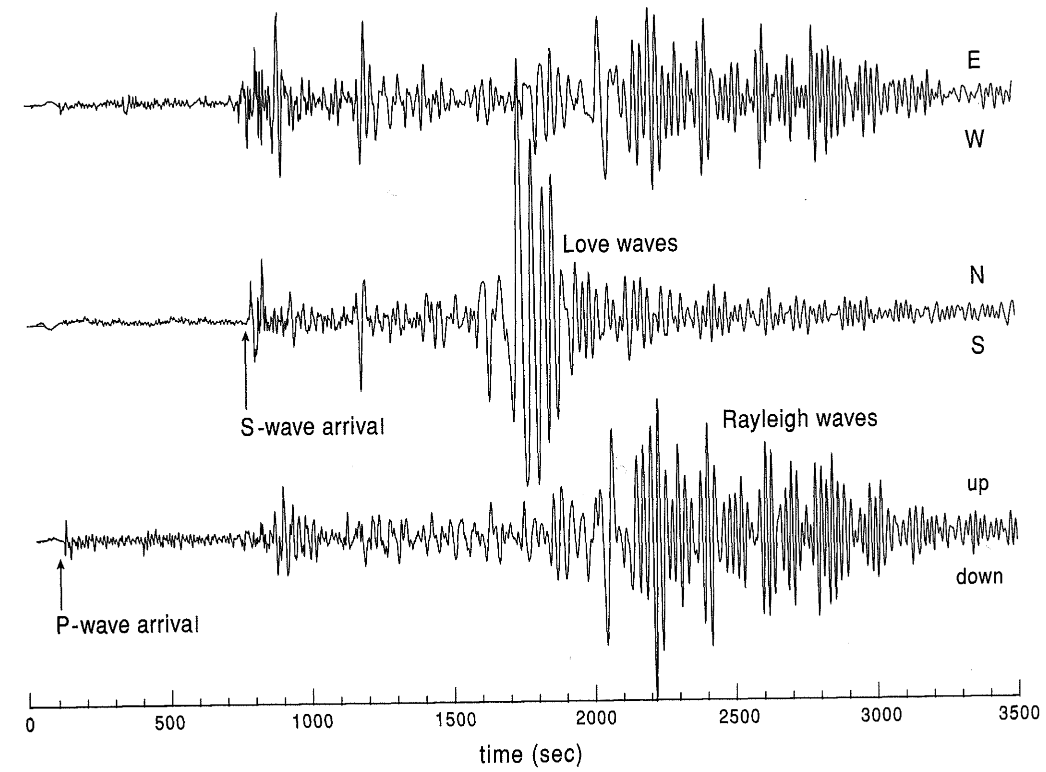


Figure 5.22 Seismograms showing surface waves.

range, he took the logarithm (to base 10) to make the numbers more manageable, so an increase of 1 in magnitude means the amplitude is 10 times as large (the energy is about 30 times as large):

$$\text{magnitude} = \log_{10}(\text{maximum amplitude of oscillation, in units of } 10^{-6}\text{m})$$

Of course, earthquakes are seldom exactly 100 km from the seismometer, so correction for distance is needed. Figure 5.23 shows one way for doing this: The amplitude is marked on the right-hand scale, the difference in P and S arrivals (which depends on the distance, as explained in Section 5.2) on the left-hand scale; the line joining them cuts the middle scale to give the magnitude, nearly 4 in this example.

Unlike the intensity scale, but like seismic moment, the magnitude scale is open ended; that is, no maximum nor minimum values are defined. Even negative magnitudes are possible: they have amplitudes less than a millionth of a metre. The largest Richter magnitude recorded is about 8.9 (Japan, 1923), but any over 7 are still very large and usually very destructive.

Richter's original definition was for shallow earthquakes comparatively near to the receiver and depended upon using a particular type of seismometer. The definition has been modified to deal with earthquakes at a wide range of distances and depths, in terms of the amplitude of ground motion rather than of a particular type of seismometer. Even so, it is still rather an arbitrary measure: for example, different expressions are needed for shallow and deep earthquakes, and these are not fully consistent.

Another drawback is that it underestimates the size of the very largest earthquakes (see Box 5.2). To surmount these limitations, the **moment magnitude,** M_W, based on the seismic moment has been introduced. Sometimes an earthquake that is larger than another on the Richter scale has the smaller moment magnitude. Box 5.2 explains the various ways of measuring magnitude.

Despite its drawbacks, the 'Richter magnitude' is widely used because it can be computed within a few hours of an earthquake and because people are familiar with it.

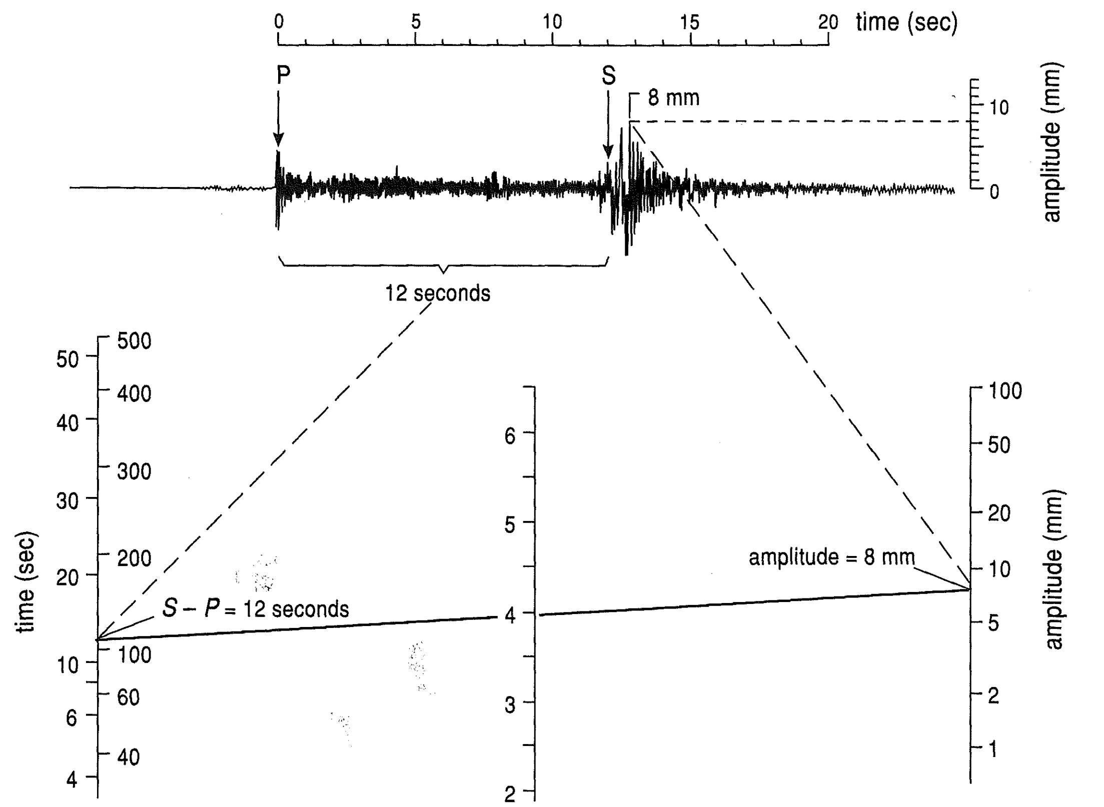


Figure 5.23 How magnitude is defined from a seismogram.

BOX 5.2 Magnitude scales

Magnitude as used today relies on the concept devised by Richter, but the formula has been modified to take account of several effects:

(i) Different types of seismometer give different amplitude seismograms. This is allowed for by calibrating the instruments to give the amplitude of oscillation of the ground. When differing distance from the epicentre is also allowed for, the resulting formula gives the local magnitude:

$$M_L = \log_{10}(\text{maximum ground amplitude}) + 3\log_{10}\Delta - 2.92 \qquad \text{Eq. 1}$$

The amplitude is measured in microns (millionths of a metre).

(ii) Richter's seismometers were at relatively small epicentral distances, where the waves with the largest amplitude are S-waves, but beyond 600 km surface waves have the largest amplitude and would give a different value. Therefore, a surface wave magnitude, M_S, is used (see Eq. 2).

(iii) Earthquakes vary in the frequencies of the waves with largest amplitude they produce, but as seismometers are not equally sensitive at all frequencies they would give different values. Therefore, they are standardised for a single frequency (0.05 Hz – a 20-sec period – for surface waves).

(iv) Particular stations give a consistently larger or smaller magnitude in comparison to the average. This is because of the local geology, which may cause the rays to be focused towards or, alternatively, deflected away from the station, or because the underlying rocks absorb the waves. A station correction term is added to allow for such variations between localities.

(v) The largest amplitudes are usually those of surface waves (except for small epicentral angles), but deep earthquakes are less effective than shallow earthquakes at generating them, and so the measured magnitude underestimates the size of deep earthquakes. To surmount this problem the largest body-wave amplitude (at a frequency of 1 Hz) is used, for this depends little upon the depth of the hypocentre. This body wave magnitude is called m_b. As using body waves instead of surface waves for shallow earthquakes would give a smaller magnitude, a constant is added to the formula to give approximately the same result, though the formulas agree only at about magnitude 6.5 (above this magnitude the value of M_S is the larger, but below it is smaller):

$$M_S = \log_{10}\left(\tfrac{A}{T}\right)_{max} + 1.66\log_{10}\Delta + 3.3 + \text{corr(depth)} \qquad \text{Eq. 2}$$

where corr(depth) is a correction made if the hypocentre is deeper than 50 km, and

$$m_b = \log_{10}\left(\tfrac{A}{T}\right)_{max} + q(\Delta, h) + \text{local correction} \qquad \text{Eq. 3}$$

where $q(\Delta, h)$ is a correction for both distance from the earthquake and for the depth, h, of the hypocentre, and ranges from about 6 to 8.

(vi) The larger an earthquake, the lower is the frequency of the waves with the largest amplitude (because of their larger area of rupture, just as a bass speaker is bigger than a treble one). As many seismometers lose sensitivity at the lowest frequencies, they underestimate the magnitudes of the largest earthquakes (m_b and M_S are affected differently, as they are measured at different frequencies).

Because seismic moment is a better measure of the size of an earthquake, the moment magnitude, M_W, was devised to match – approximately – the Richter magnitude at lower magnitudes but often gives higher values for the largest earthquakes. The largest known value of M_W is 9.6 (Chile, 1960), compared to an M_S of 8.3.

$$M_W = \tfrac{2}{3}\log_{10}M_0 - 10.7 \qquad \text{Eq. 4}$$

where M_0 is the seismic moment.

(vii) There are other limitations, such as where in the radiation pattern (Fig. 5.6) the receiver chances to be.

Clearly, no formula for magnitude is fully satisfactory.

5.9 Energies of earthquakes

Seismic waves carry energy, as is obvious from the damage they can cause. Waves are momentary deformations of rock, and if the energies of all such simultaneous deformations over the whole volume affected by the earthquake were added together, we would have the seismic energy of the earthquake. (This is less than the strain energy released by a fault rupture, for reasons given in Section 5.1.) This is a difficult sum to carry out but it has been estimated and related to the magnitude of the earthquake; the approximate relationship is shown in Figure 5.24.

An increase of 1 in magnitude means 10 times the amplitude but about 30 times the energy. The energy of some earthquakes far exceeds that of the largest nuclear bomb exploded, which is why great earthquakes are so destructive. The 1960 Chile earthquake released the equivalent of about 2500 megatonnes of high explosive.

Numbers of earthquakes and their total energy. The largest earthquakes, fortunately, occur less frequently than smaller ones; for instance, each year there are about 100,000 earthquakes of $M_L = 3$, compared to 20 of $M_L = 7$. For any region, or the Earth as a whole, a plot of numbers N of earthquakes greater than a given magnitude M (as a log) versus the magnitude gives an approximate straight line ($\log_{10} N = a - bM$), though the slopes (b) are not usually the same, as shown in Figure 5.25 (only earthquakes above a certain magnitude have been plotted, but smaller ones are even more numerous). There are many more small earthquakes than large ones, but as they have little energy the relatively few large earthquakes account for most of the seismic energy released. (Similarly for seismic moment; the 1960 Chile earthquake released a quarter of the total moment in the period 1904–1986.) The average annual energy release worldwide equals that of about 100 megatonnes of high explosive! But large as this is, it is small compared to some other forms of global energy, as explained in Chapter 17.4.1.

5.10 Earthquake damage and its mitigation

5.10.1 Causes of damage

It is hardly surprising that the huge energies released by large earthquakes produce enormous damage, but how is the damage actually caused? There are several mechanisms, both direct and indirect.

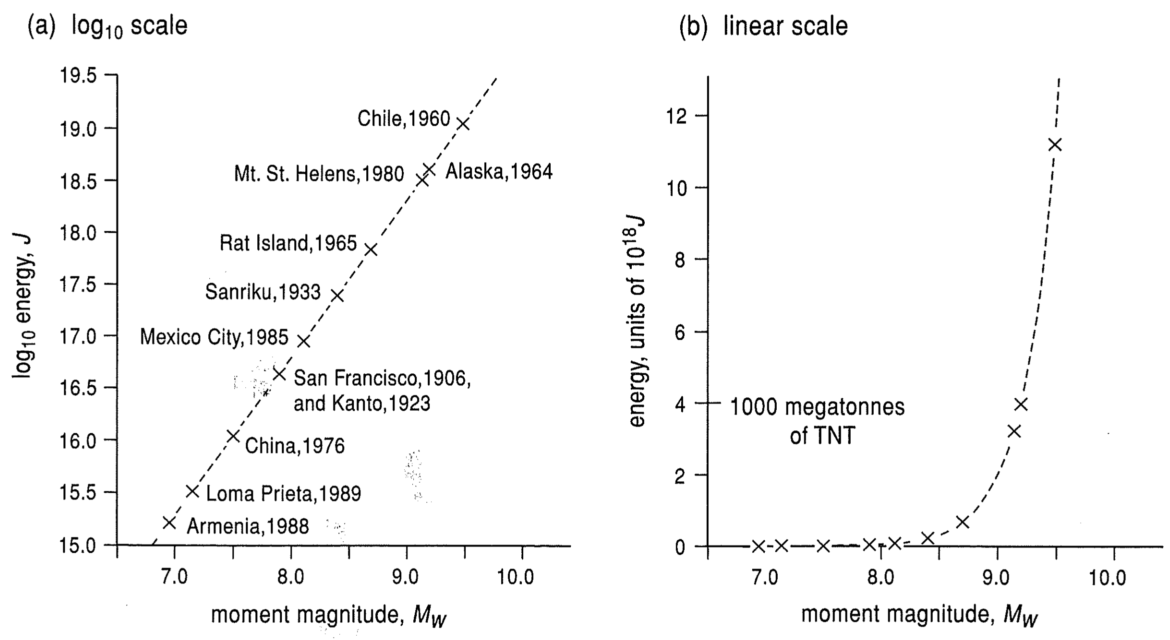


Figure 5.24 Energy and magnitude.

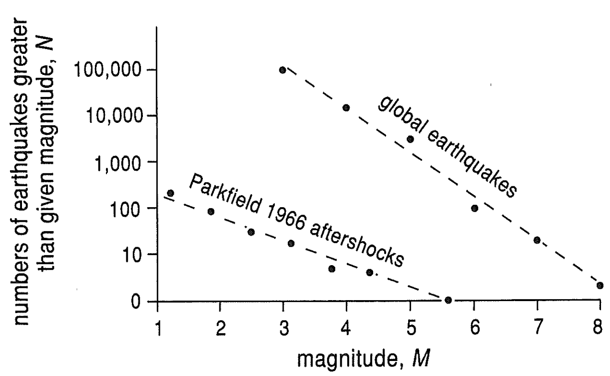


Figure 5.25 Numbers of earthquakes of different magnitudes.

Most obvious is the rupture, if it reaches the surface, for any house, road, pipe, and so on straddling the fault is likely to be ruptured. However, such damage is confined to the fault trace.

Much more extensive, and the chief cause of damage, is the shaking caused by the passage of seismic waves, particularly surface waves, for often these have the largest amplitude and continue the longest (Fig. 5.22). Aftershocks may cause extensive damage, for though their magnitudes are less than that of the main shock they may occur nearer to towns. Few traditional buildings are designed to be shaken, least of all from side to side, and many collapse; particularly at risk are masonry and adobe buildings, which lack cross bracing. More resistant are buildings with wood or steel frames, though buildings in earthquake-prone areas need to be specially designed to survive severe shaking. Some countries are zoned according to seismic risk and have building codes to match.

The amplitude of shaking depends not only upon the magnitude of the earthquake and its distance, but also on the local geology. When waves slow on entering rocks with a lesser seismic velocity, their wave trains shorten and therefore their amplitude increases; this commonly happens as waves approach the surface, for surface rocks often have lower velocities than deeper ones. In addition, unconsolidated rocks such as alluvium may be temporarily 'liquefied' by the shaking (just as sand and mud when rapidly shaken tend to flow like a thick liquid), and soil liquifaction has caused buildings to topple or collapse. Another effect is that subsurface interfaces can reflect or refract the waves so that the seismic energy is partly focused into the area or, alternatively, deflected away.

Other mechanisms of destruction are indirect. The shaking of the ground may initiate an avalanche or landslip. In turn, these may dam a river valley, forming a lake; once the lake overflows the dam, it often cuts rapidly down through the unconsolidated jumble of rocks and soil, releasing the pent-up water to cause severe flooding downstream. The collapse of buildings containing fires and the rupturing of gas pipes may lead to widespread conflagrations, which, with the general chaos and damage to water pipes, may rage unchecked. Fire was a major cause of damage in the 1906 San Francisco earthquake and in the Kobe earthquake in Japan in 1995. People have even been known to set fire to their house on discovering that their insurance covers fires but not earthquakes!

Yet another form of damage may be suffered along coasts, from tsunamis (sometimes incorrectly called tidal waves – they have nothing to do with tides). These are generated when a mass of water is abruptly displaced, usually by movement of the sea floor below shallow seas, often by an earthquake, though also by volcanic eruptions and landslips. The waves travel across the deep oceans with a wavelength of perhaps 100 km but an amplitude of no more than a metre, so that a ship does not notice their passing, being slowly raised and lowered over a period of several minutes. But when the waves reach the shallow water near a coast they 'break', just as normal-sized waves do when they cause surf. Tsunamis may reach a height of 30 m or more, sweeping far inland, sometimes carrying ships to improbable places and devastating towns on their way. Around the Pacific, tsunamis are a sufficiently frequent hazard that the Seismic Sea Wave Warning System has been set up. When a large earthquake has been detected seismically, changes of sea level are checked locally to see if a tsunami has been generated; if it has, warnings are sent by radio to vulnerable shores. Though the waves travel at perhaps 800 km/hr, they still take several hours to cross an ocean, sufficient time for warning. (But the warning does not always save lives, for the curious may go to view the freak wave and be overwhelmed!)

5.10.2 Mitigating the damage caused by earthquakes

Most earthquakes are caused by tectonic forces that are beyond our control, and there is no current prospect of preventing or reducing the severity of earthquakes. Much effort, therefore, has been put into predicting them. Ideally, we would like to predict quite precisely the size, location, and time of earthquakes, but despite all the effort and a few apparent successes there is yet no reliable method for precise prediction. That an earthquake will occur in a certain area some time in the future can often be anticipated from the buildup of strain, but not its precise details. The problem is similar to predicting the breakage of a length of string when it is suspending a bucket into which water is dripping: we know the increasing weight of water must break the string but not exactly when or where. A number of possible precursors might warn of an impending earthquake. These include changes of seismicity; variation of seismic, electrical, and other properties of the ground (measured using geophysical methods); variation in the water level in wells and its radon content (Section 16.3); and so on. However, no change has yet been found that regularly precedes an earthquake but does not occur otherwise, and seismologists are becoming increasingly pessimistic that prediction is possible.

Still, much can be done to lessen the damage. Areas prone to earthquake can be identified from a knowledge of where stresses are increasing and from historical records of earthquakes. Records can be extended to prehistorical times using 'palaeoseismology'; though past earthquakes have left no seismograms, they have often produced faulting in unconsolidated sediments, which can be recognised in geological sections. Dating these, using carbon-14 dating or other radiometric dating methods (Section 15.12.2), reveals how often earthquakes occurred. This information can be used to estimate the chance of a large earthquake occurring in the next few decades. This information can be used to draw up appropriate building codes that will allow buildings and other structures to survive the earthquake, and to plan how to deal with the damage that will still occur. Seismology can also discover blind faults – ones that have no expression at the surface – either by seismicity on the fault or finding the fault in seismic surveys, as recently occurred for a fault below Los Angeles, California.

Seismic networks have been set up in earthquake-prone areas to record earthquakes as they occur. Within minutes the location and size of an earthquake can be calculated and used to direct emergency services to likely areas of damage. Such systems can even provide warning if the hypocentre is at a distance from towns and susceptible structures. In Japan the bullet trains are automatically slowed, and in 1995 Mexico City was provided with a 72-sec warning of the magnitude 7.3 earthquake that occurred 300 km away. This warning is sufficient to shut down power stations.

Summary

1. 'Earthquake' can refer either to the shaking of the ground and associated effects at a *locality,* or to the *source* of the shaking, which is usually a sudden displacement, or rupture, of part of a fault.
2. Rupture occurs when the shearing strain across the fault becomes too large for the fault to support. The resulting elastic rebound releases strain energy and generates elastic waves.
3. Where the rupture starts is the hypocentre (or focus) of the earthquake; the point on the surface immediately above the hypocentre is the epicentre.
4. The distance of an earthquake from a receiver is determined from the delay, or lag, of the S arrival after the P arrival. The depth is found from the lag of the arrival of the pP-ray after the P-ray. Seismograms of three or more stations are used to locate the hypocentre of an earthquake.
5. The senses of first motion, compressive or dilatational, of P-wave arrivals at a number of seismic stations are used to determine fault-plane solutions, as described in Box 5.1. The fault plane can be distinguished from the auxiliary plane only by using additional information. The axes of compressive and tensile stresses (actually, maximum and minimum compressive stresses) can be found without ambiguity.
6. The dimensions of the rupture (length along strike, L, depth down dip, W) and the displacement D can be found from a combination of observation of faulting, location of aftershocks, surveying, and other ways.
7. The best measure of the size of an earthquake at source is the seismic moment, M_0.
8. The usual cause of earthquakes is tectonic stress; consequently, seismotectonics studies earthquakes to find the direction of stress or the nature of the tectonic process. This can be done qualitatively, using fault-plane solutions, or quantitatively by using seismic moments.
9. A less rigorous measure of the earthquake size at its source is the Richter magnitude, which is based on the maximum amplitude of ground motion. It has a variety of forms (Box 5.2). It is easy to determine and is widely quoted. The moment magnitude, M_W, is based on the seismic moment and is better for the largest earthquakes. Magnitude is on an open-ended log scale. An increase in magnitude of 1 means 10 times the amplitude of oscillation but about 30 times the energy. The largest earthquakes, with magnitudes over 7, have energies equivalent to many megatonnes of high explosive.
10. Larger earthquakes result from larger strained volumes, rather than greater strains. The largest earthquakes result from ruptures extending for hundreds of kilometres and involve millions of cubic kilometres of rock.
11. Intensity measures the severity of an earthquake at a *locality,* in terms of observed effects, such as shaking, and damage. The usual intensity scale is the 12-point Modified Mercalli Scale.
12. Earthquake damage results chiefly from shaking caused by surface waves and S-waves. Other mechanisms include rupturing of the surface, liquifaction of unconsolidated surface rocks, landslides, and tsunamis.
13. Earthquakes cannot yet be predicted, but areas prone to earthquakes can often be identified, with estimates of the likely maximum size and repeat interval. Local conditions that would increase or decrease the damage can be recognised.
14. You should understand these terms: earthquake, hypocentre, epicentre; elastic rebound, aftershock; fault plane, fault-plane solution, auxiliary plane, focal sphere, take-off angle, double-couple mechanism; intensity, isoseismal, seismic moment, moment magnitude, magnitude, Richter magnitude; body, surface, Rayleigh, and Love waves; seismotectonics.

Further reading

Bolt (1999) gives an introduction to most aspects of earthquakes, and Doyle (1995) includes several chapters on their different aspects. Kasahara (1981) and Scholz (1990) are more advanced: The former covers most aspects, while the latter emphasises how rocks fail, the buildup and release of strain, and seismotectonics. Coontz (1998) gives a discussion of whether earthquake prediction is possible, written at a popular level, while Kanamori et al. (1997) describe the value of seismic networks that provide information about an earthquake with minimal delay.

Problems

1. A seismic recording station receives the first S arrival 8 min after the P one. What is the epicentral angle to the source?
(i) 26°, (ii) 29°, (iii) 45°, (iv) 60°, (v) 68°, (vi) 76°.
2. The length of rupture on a fault plane associated with a large earthquake may be hundreds of kilometres long and tens of kilometres deep. How then can we refer to the hypocentre of the earthquake?
3. If the long-term average rate of displacement is roughly the same along the length of a long fault, how can some places experience large earthquakes yet others escape them?
4. A newspaper reported that an earthquake with an intensity of 7.3 on the Richter scale had caused thousands of deaths. What is their error?
5. Give at least two reasons why the P-waves recorded on a receiver due to an earthquake are not a single pulse.
6. A ray arrives at a seismic station 80° from the near-surface hypocentre. What was its take-off angle?
7. During an earthquake a building experiences horizontal shaking. It could have been due to which waves?
   (i) P-waves.
   (ii) S-waves.
   (iii) Rayleigh waves.
   (iv) Body waves.
   (v) Love waves.
   (vi) Surface waves.
8. A 3-component seismometer aligned N–S, E–W, and vertically recorded far the largest amplitude

of Love waves in the E–W direction. What is the approximate direction of the earthquake from the station?

9. Which of the following ways of measuring the size of an earthquake does not need an instrumental record:
   (i) Richter magnitude.
   (ii) M_W.
   (iii) Moment.
   (iv) Intensity.

10. Plot the data in the table, find the fault-plane solutions and deduce what type of fault was involved (strike-slip, etc.).

Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*
0	18	D	123	28	D	247	55	C
34	27	C	153	46	D	270	19	C
45	19	C	202	15	D	311	16	C
58	29	C	230	31	D	329	46	D
94	17	C	238	29	D	356	39	D

*C: compressional; D: dilatational

11. Plot the data in the table, find the fault-plane solutions and deduce what type of fault was involved (strike-slip, etc.).

Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*
46	74	C	138	67	D	259	64	D
64	38	C	140	37	D	317	58	D
89	43	D	192	54	D	318	78	C
104	68	C	198	64	C	334	51	C
128	76	C	243	71	C	351	22	D

*C: compressional; D: dilatational

12. Plot the data in the table and find the dip and strike of the possible fault planes and the slip angles.

Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*	Azimuth	Take-off angle	Sense*
9	70	D	128	85	D	276	32	C
26	69	D	150	48	D	276	68	C
41	21	C	206	63	D	298	50	D
62	66	C	226	37	D	310	81	D
86	45	C	237	22	C	335	82	D
			244	72	D	349	41	C
112	28	C	255	82	C			

*C: compressional; D: dilatational

13. When a certain locality experienced an earthquake with a Richter magnitude of 7 there was little damage. A newspaper reported that as this was only a little less than the largest known in the area, magnitude 8, citizens could be confident that recent work to make buildings, bridges, and so on safe had been successful. Why should they not be too confident?

Chapter 6
Refraction Seismology

Refraction seismology is a powerful and relatively cheap method for finding the depths to approximately horizontal seismic interfaces on all scales from site investigations to continental studies. It also yields the seismic velocities of the rocks between the interfaces, a useful diagnostic tool.

It exploits a particular case of refraction in which the refracted ray runs along an interface, while sending rays back up to the surface.

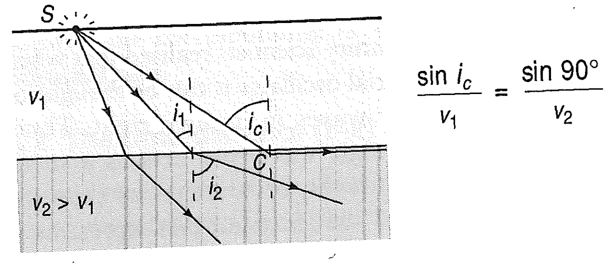


Figure 6.1 Critical angle of refraction.

6.1 Critical refraction and head waves

A ray crossing an interface downwards into a layer with a higher velocity – a common situation – is refracted away from the normal, according to Snell's Law (Eq. 4.5). Consider rays progressively more oblique to the interface (Fig. 6.1): There must come an angle, the **critical angle**, i_c, where the refracted angle is exactly 90°; that is, the refracted ray travels along the interface (if the ray is yet more oblique, *all* the seismic energy is *reflected*, called total internal reflection). The value of i_c depends on the ratio of velocities either side of the interface:

$$\sin i_c = \frac{v_1}{v_2}\sin 90° = \frac{v_1}{v_2} \qquad \text{Eq. 6.1}$$

What does a ray travelling along the interface mean: Is it in layer 1 or 2? This is best answered by considering wave fronts. As rays and wave fronts always intersect at right angles, a ray along the interface means waves perpendicular to the interface (Fig. 6.2a). A wave is due to oscillations of the rock, and as the layers are in contact, the rocks just above the interface must oscillate too. Therefore, as a wave front travels along just below the interface a disturbance matches it just above the interface. This in turn propagates waves – and therefore rays – up to the surface at the critical angle (Fig. 6.2b).

To understand why this is so, we use the concept of **Huygens's wavelets.**

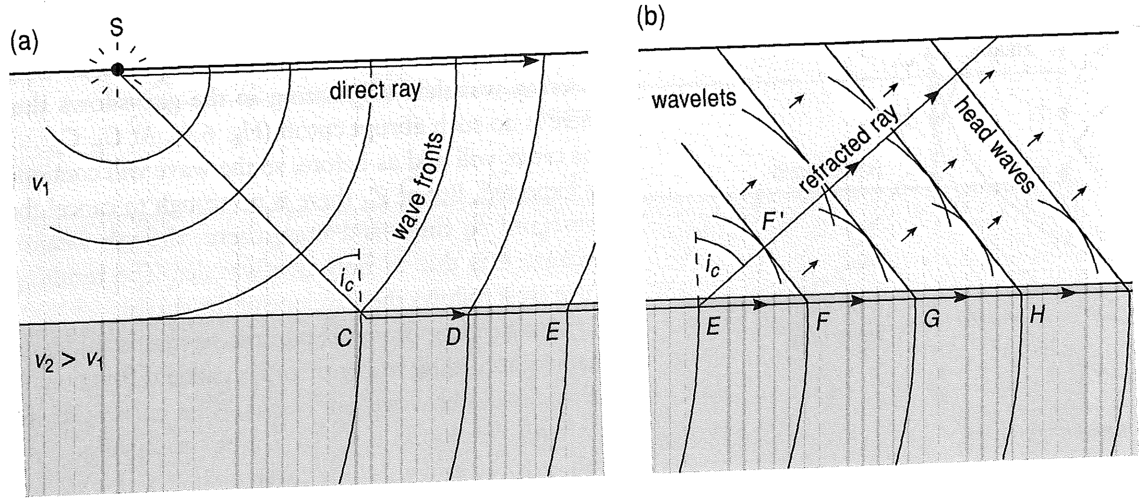


Figure 6.2 Wave fronts and head waves.

6.1.1 Huygens's wavelets

Huygens (a 17th-century scientist) realised that when a particle of a material oscillates it can be thought of as a tiny source of waves in its own right. Thus, every point on a single wavefront acts as a small source, generating waves, or wavelets as they are called. In Figure 6.3 the left-hand line represents a planar wave front travelling to the right at some instant. Wavelets are generated from all points along its length, and a few are shown. The solid half-circles are the crests of the wavelets after they have travelled for one wavelength (which is the same as that of the wave). At C_1, C_2, C_3, . . . these overlapping crests add, called reinforcement, so that they form a new crest. The succeeding troughs will have advanced half as far, shown by the dashed half-circles. At each of T_1, T_2, . . . there are two troughs and these will add, giving a trough between the two crests. But at Z_1, Z_2, . . . there is a crest and a trough, and these cancel perfectly. As there will be just as many troughs as crests along the left-hand line – assuming it is indefinitely long – then there will be perfect cancellation all along the line. Therefore, the crest has advanced from left to right. This is just what we have been using since waves were introduced in Section 4.1, which is why wavelets have not been introduced before. However, wavelets are helpful in some special cases, such as critical refraction and diffraction, the latter being considered first.

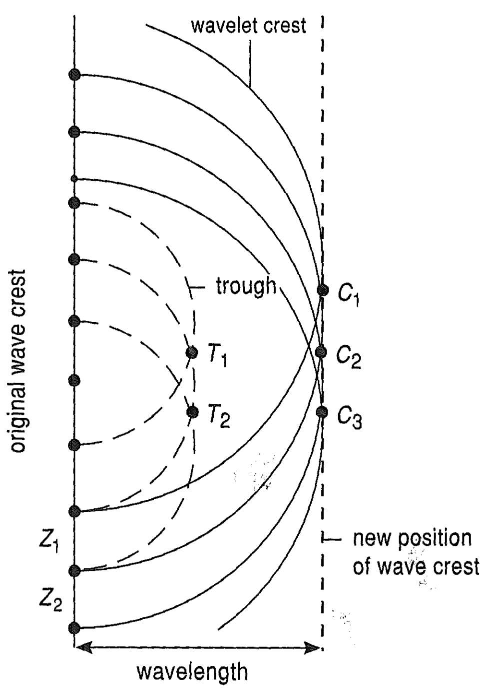


Figure 6.3 Huygens's wavelets.

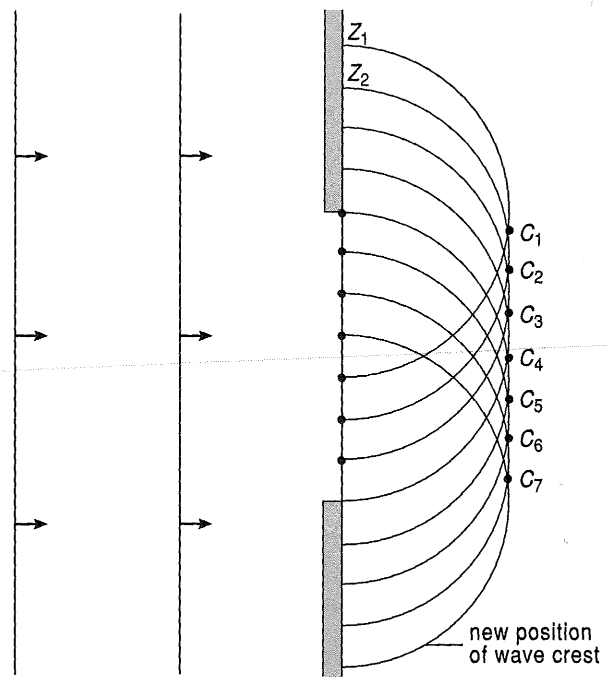


Figure 6.4 Diffraction into a shadow.

Diffraction. If waves pass through a gap, such as water waves entering a harbour mouth, it might be expected that once inside they would continue with the width of the gap, with a sharp-edged shadow to either side. But if this were so, at the crest ends there would be stationary particles next to ones moving with the full amplitude of the wave, which is not feasible. Considering wavelets originating in the gap shows that there is no such abrupt cutoff (Fig. 6.4). At C_1, C_2, . . . the crests will add as before, so the wave will continue as expected. But at Z_1 there is no trough to cancel the crest and so there is a crest there, though weaker because it is due to only one wavelet. The resulting wave crest follows the envelope of the curves.

This bending of waves into places that would be a shadow according to ray theory is called **diffraction.** It helps explain how we can hear around corners (there is often reflection as well). It will be met again in Section 6.10.1, and in some later chapters. Next, we return to the critically refracted wave of Figure 6.2.

6.1.2 Head waves

As a wave travels below the interface it generates wavelets (Fig. 6.2b). By the time a wavelet in the lower layer has travelled from E to F, a wavelet in the upper layer – travelling at velocity v_1 – has gone only a distance EF'. If we draw the wavelets produced by successive waves below the interface, at E, F, . . . , and add up their effects, the result is waves travelling up to the right (Fig. 6.2b). These are called **head waves,** and their angle i_{head} (which equals $\angle EFF'$) depends on the ratio of EF' to EF:

$$\sin i_{head} = \frac{EE'}{EF} = \frac{v_1}{v_2} = \sin i_c \qquad \text{Eq. 6.2}$$

This is the ratio of the velocities, and so is the value of the critical angle (Eq. 6.1): Therefore, head rays leave the interface at the critical angle.

The seismic refraction method depends upon timing the arrivals of head waves at receivers on the surface; the corresponding rays are usually called **refracted rays.**

6.2 The time-distance (t–x) diagram

Seismic energy can follow three main routes from the source to receivers (Fig. 6.5b): the refracted rays just described, the **direct ray** that travels just below the surface (shown in Fig. 6.2a), and **reflected rays.** They take different times to reach the receivers, as shown in the **time–distance,** or **t–x, diagram** (Fig. 6.5b).

The time it takes a direct ray to reach a receiver is simply the distance along the surface divided by the velocity, v_1, so its graph is a straight line from the origin. The slope of the line equals $1/v_1$, so the greater v_1 the shallower the slope. To find the travel-time to a receiver, follow straight up from the receiver on the lower diagram until you reach the line on the t–x diagram, then go left to read off the time on the vertical time axis.

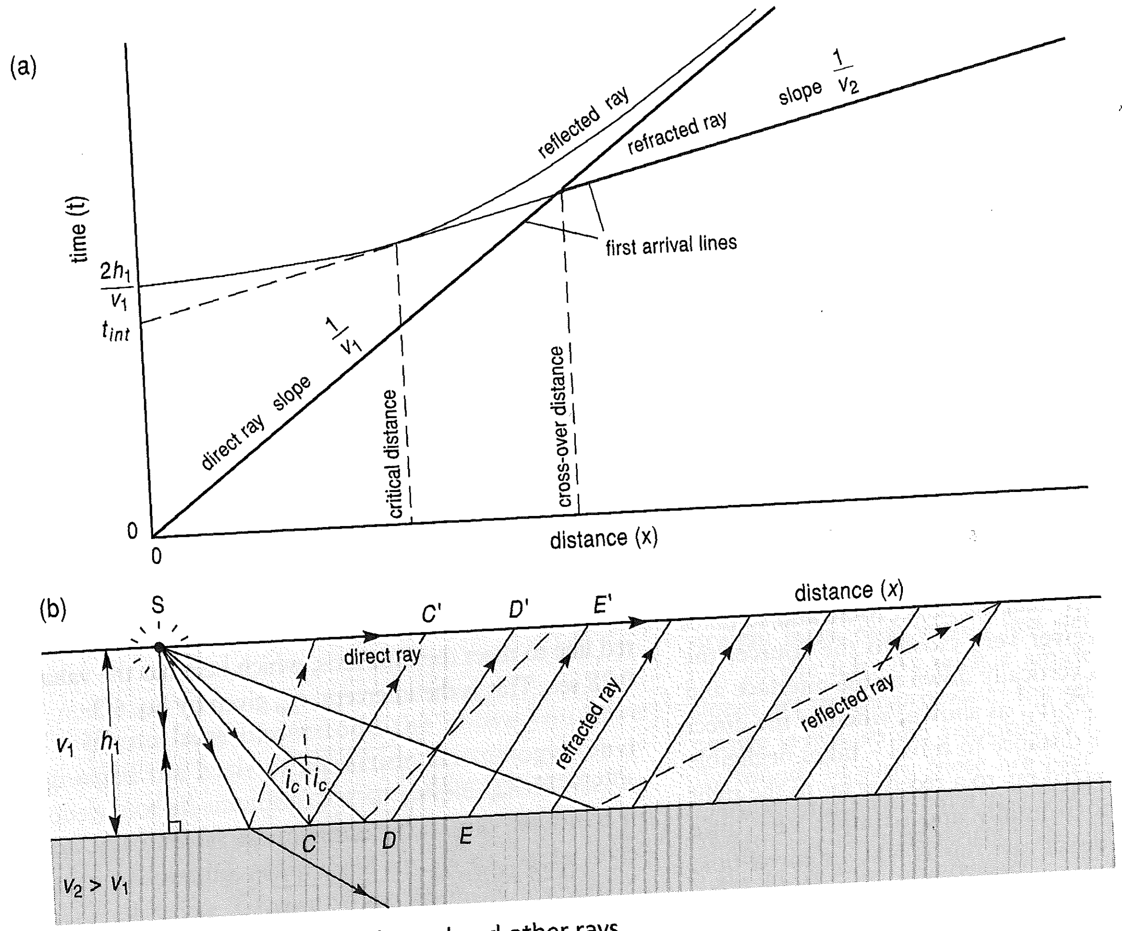


Figure 6.5 Travel-times of refracted and other rays.

Finding the time of a refracted ray is more complicated. Clearly, a refracted ray travels further than the direct ray, but part of its route is in the second layer, where it travels faster. If it travels far enough in the second layer it must overtake the direct ray (just as travelling to a distant destination via a motorway is quicker than using the shorter but slow road, even though you have to travel to and from the motorway on slow roads). All the refracted rays follow the critical ray *SC* down to the interface, and an equal distance up from the interface to the surface, *CC′*, *DD′*, *EE′*, and so on, both at velocity v_1; to these must be added the connecting part in the second layer, *CDE*. . . . Therefore their total times will differ only by the different times they spend in the lower layer. For instance, ray *SCEE′* takes longer than *SCDD′* by time DE/v_2. As a consequence, the travel-time line for refracted rays is a straight line, with slope $1/v_2$; but it does not start at the origin, because the nearest point to the source it can reach is *C′*, the **critical distance.** Nor does the travel-time line, extended back, reach the origin, because of the extra time taken travelling to and up from the interface.

The time for a refracted ray to reach a receiver is the time spent below the interface, plus the time it takes to go down to, and back up from, the interface. It works out to be

$$t = \frac{1}{v_2}\cdot x + 2h_1\sqrt{\frac{1}{v_1^2}-\frac{1}{v_2^2}}$$

time = slope · distance + intercept, t_{int} Eq. 6.3a

Equation 6.3a can also be written as

$$t = \frac{x\sin i_c}{v_1} + \frac{2h_1\cos i_c}{v_1} \qquad \text{Eq. 6.3b}$$

This will be used later.

The third route to a receiver is by reflection. This obviously is further than the direct route and so takes longer. For a receiver very close to the source, the reflected ray goes vertically down to the interface and up again, taking $2h_1/v_1$, as shown, whereas the direct ray – having little distance to travel – takes negligible time. But the reflected ray to a receiver far to the right travels almost horizontally and so takes very little longer than the direct ray; this is why its travel-time curve approaches that for the direct ray. Though the reflected route is shorter than the refracted one it always takes more time, except for the ray *SCC′*, which can be regarded as either a reflected or a refracted ray; therefore the refracted and reflected times to the critical distance, *C′*, are the same. Reflected rays are never first arrivals and are not considered further in this chapter. However, they form the basis of a most important prospecting method, reflection seismology, described in the next chapter.

Figure 6.5a describes what happens below ground, but we have to make deductions from surface measurements, in particular the times of first arrivals, as these are the easiest to recognise on a seismogram. The first arrivals for near distances are direct rays, but beyond a certain distance, the **crossover distance,** refracted rays are the fastest. Therefore, first arrival times for a range of distances trace out the two heavy lines of Figure 6.5b.

To get this information, a row of receivers – seismometers or geophones – is laid out in a line from the source, called the **shot point.** The resulting seismograms are shown in Figure 6.6a. The times of first arrival are 'picked' and their values are plotted against the distances of the corresponding receivers (receivers are often spaced more closely near the shot point, for reasons explained in Section 6.7). Then lines are drawn, for the direct and refracted travel-times (Fig. 6.6b).

These lines are used to deduce the depth to the interface and the velocities of both layers. First, the slopes of the direct and refracted lines are measured (this is easier if a convenient distance is used, as shown); their reciprocals yield the velocities of the top and lower layers:

Slope of direct line = 0.42 sec/250 m = 0.00168, so v_1 = 1/0.00168 = 595 m/s.
Slope of refracted line = 0.34 sec/500 m = 0.00068, so v_2 = 1/0.00068 = 1470 m/s.

These velocities can be useful for suggesting likely rock types (Table 6.1), but usually more useful is the depth to the interface, deduced as follows: Extend the refracted line to meet the time axis, which here has the value 0.28 sec. This is the intercept, t_{int}, given by Eq. 6.3:

$$\text{intercept, } t_{int} = 2h_1\sqrt{\frac{1}{v_1^2}-\frac{1}{v_2^2}};$$

$$0.28\ \text{sec} = 2h_1\sqrt{\frac{1}{595^2}-\frac{1}{1470^2}},$$

$$\text{so } h_1 = 91\,m$$

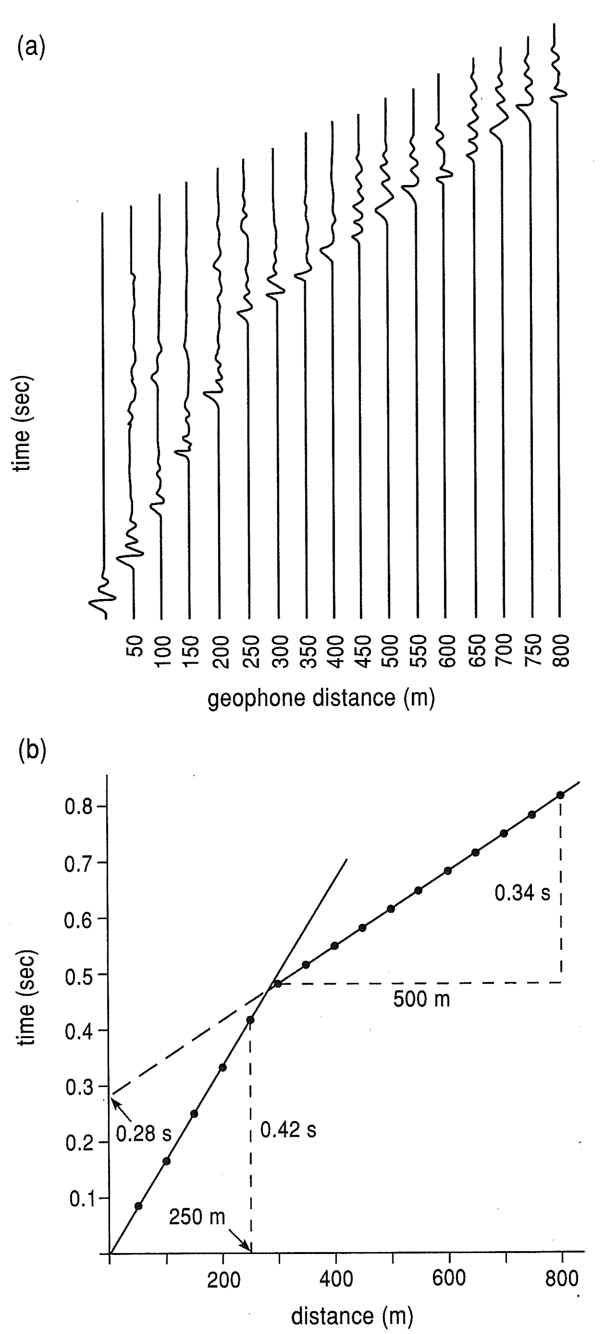


Figure 6.6 Making a time–distance diagram.

If you have understood the above you have grasped the basis of refraction seismology. Obviously, not all situations are so simple: There may be several interfaces, and interfaces may be tilted or even undulating. All of these situations can be investigated using refraction seismology, as will be explained in the following sections.

Table 6.1 Seismic velocities for rocks

Rock type	v_p (km/sec)
Unconsolidated sediments	
clay	1.0–2.5
sand, dry	0.2–1.0
sand, saturated	1.5–2.0
Sedimentary rocks	
anhydrite	6.0
chalk	2.1–4.5
coal	1.7–3.4
dolomite	4.0–7.0
limestone	3.9–6.2
shale	2.0–5.5
salt	4.6
sandstone	2.0–5.0
Igneous and metamorphic rocks	
basalt	5.3–6.5
granite	4.7–6.0
gabbro	6.5–7.0
slate	3.5–4.4
ultramafic rocks	7.5–8.5
Other	
air	0.3
natural gas	0.43
ice	3.4
water	1.4–1.5
oil	1.3–1.4

Ranges of velocities, which are from a variety of sources, are approximate.

6.3 Multiple layers

If there are several layers, each with a higher velocity than the one above – which is often the case – there will be critical rays for each interface, as shown in Figure 6.7. For each interface the critical angle depends only upon the velocities above and below it, but the ray paths down to an interface, SC_1, SC_2, SC_3, . . . , depend upon the thicknesses and velocities of all the layers above.

The *t*–*x* diagram for first arrivals is a series of straight lines, each with a less steep slope than the one to the left. As with a single interface, their slopes are the reciprocals of the velocities of the successive layers, so the velocities of the layers are easily found. In turn, these are used to calculate the critical angle for each interface, using Eq. 6.1. Depths are calculated from the intercept times but have to be done progressively. The depth, h_1, to the first interface, using Eq. 6.3b with $x = 0$, is

$$t_{\mathrm{int}_1} = \frac{2h_1 \cos i_{c_1}}{v_1}$$

giving:

$$h_1 = \frac{v_1 t_{\mathrm{int}_1}}{2 \cos i_{c_1}} \qquad \text{Eq. 6.4}$$

Its value is put in the equation for the intercept time for the second interface, which can be shown to be

$$t_{\mathrm{int}_2} = \frac{2h_1 \cos i_{c_1}}{v_1} + \frac{2h_2 \cos i_{c_2}}{v_2} \qquad \text{Eq. 6.5}$$

which gives h_2. This is repeated for successive interfaces, the equation for the interface time increasing by one term each time.

6.4 Dipping interfaces

Tilting the interface does not change the value of the critical angle, but it rotates the ray diagram of Figure 6.3 by the angle of dip, α (Fig. 6.8). As a result, rays to successive receivers C', D', E', . . . not only have to travel the additional distance CD, DE, . . . along the interface, but also an extra distance up to the surface because the interface is getting deeper. Therefore the refracted line on the t–x diagram is steeper, its slope yielding a velocity less (slower) than v_2.

This presents a problem: There is no way to tell from Figure 6.8 that the refracted line is due to a dipping interface and not a horizontal interface over a layer with a slower seismic velocity equal to 1/slope. For this reason, velocities calculated from the slopes are called **apparent velocities.** However, if the line is **reversed,** that is, repeated with the shot point at the right (Fig. 6.9), then successive head rays have *less* distance to travel up to the surface, and so the refracted line has a shallower slope, and appears faster. Therefore, *the presence of a dipping interface is shown by the **forward** and **reversed** refraction lines having different slopes* ('forward' and 'reverse' merely depend on which direction is shot first). If the interface has a small dip (less than about 5°), an approximate value of v_2 is found by averaging the forward and reverse slopes:

$$\frac{1}{v_2} \approx \frac{1}{2}\left(\mathrm{slope}_{2_f} + \mathrm{slope}_{2_r}\right) \qquad \text{Eq. 6.6}$$

v_2 can then be used to find the critical angle, as usual (Eq. 6.1).

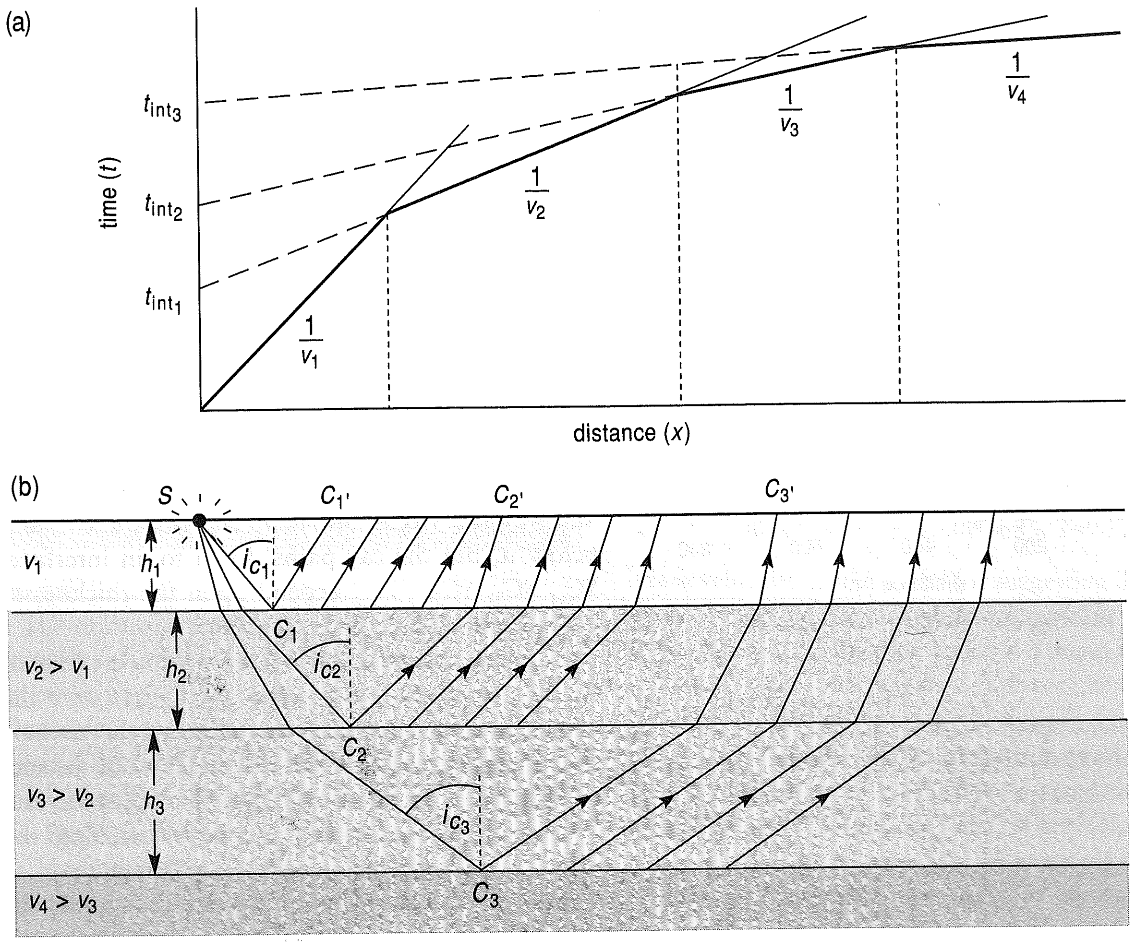


Figure 6.7 Multiple layers.

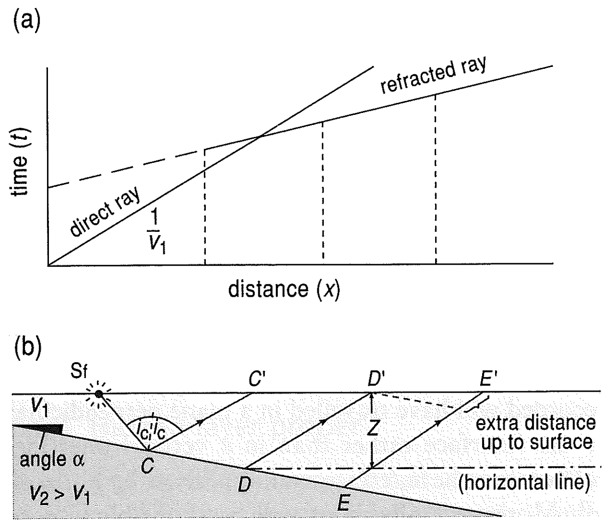


Figure 6.8 Dipping interface.

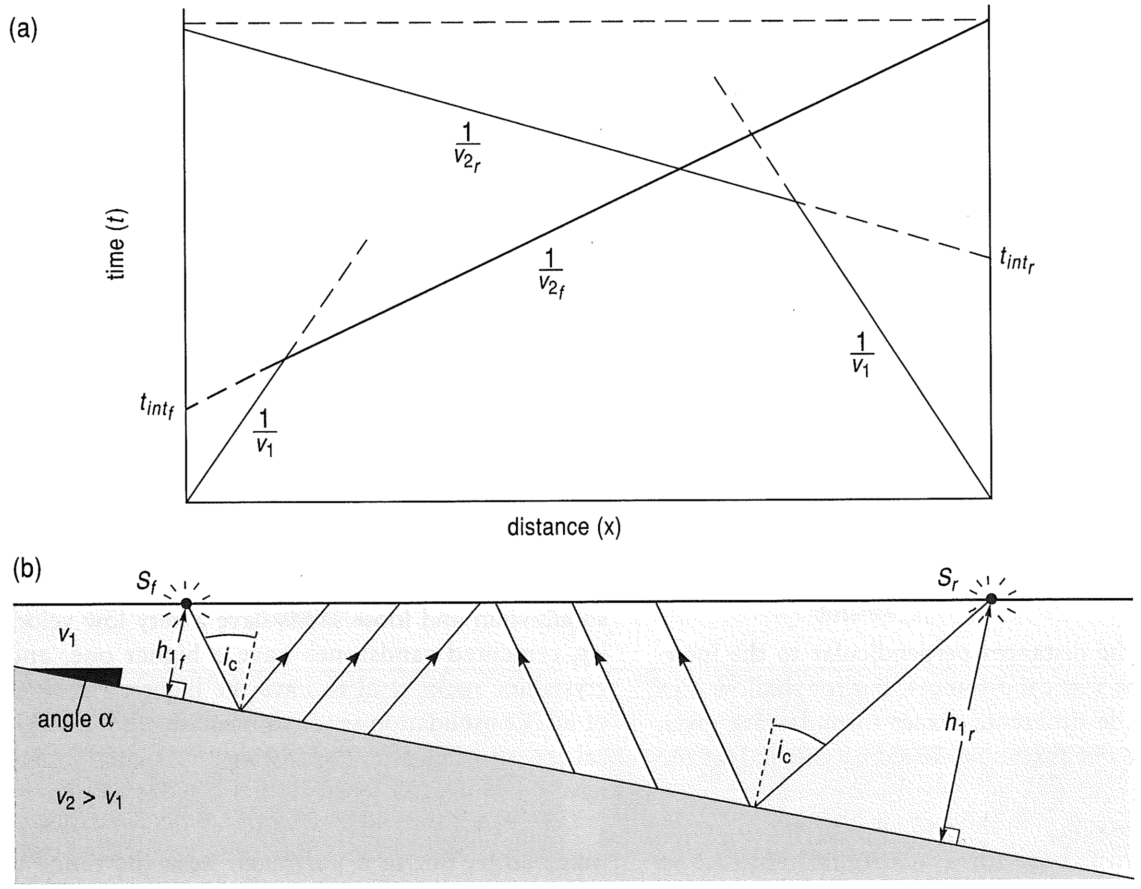


Figure 6.9 Reversed lines.

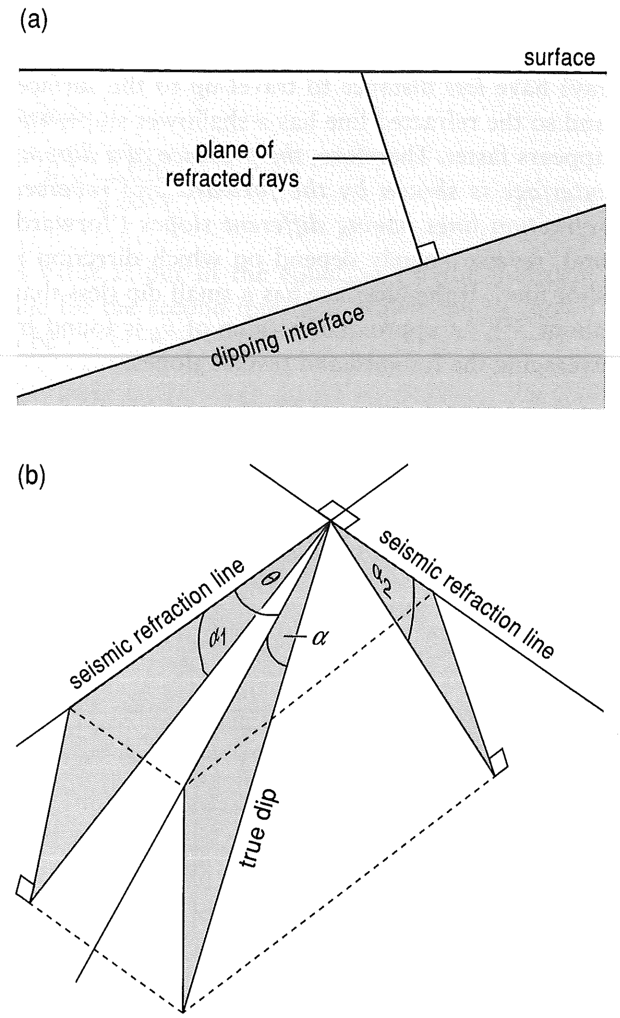


Figure 6.10 Dip in three dimensions.

The intercepts are different too, being less at the up-dip than the down-dip end. The depths to the interface are given by

$$h_{1_f} = \frac{t_{\text{int}_f} v_1}{2 \cos i_c} \qquad h_{1_r} = \frac{t_{\text{int}_r} v_1}{2 \cos i_c} \qquad \text{Eq. 6.7}$$

These are the distances perpendicular to the interface, not the vertical distances, but for shallow dips they are little different. (Exact formulas are given in some of the textbooks listed at the end of the chapter.)

In summary, t–x plots for a dipping interface compare with those for a horizontal interface as follows:

(i) The slope of the refraction line is less steep up dip, steeper down dip.
(ii) The intercept is less at the up-dip than the down-dip end.
(iii) The slopes for direct rays (first sections on *t–x* plot) are unchanged.

The true dip. It has been assumed so far that the seismic line is in the direction of the dip of the interface, but there is no reason why this should be so, unless it was known when the survey was set out. If the reversed lines are shot along strike, the rays recorded will have travelled in a plane perpendicular to the interface rather than in a vertical one (Fig. 6.10a), but the interface will appear to be horizontal. More generally, lines will be shot obliquely to the dip direction and give a value for the dip that is too low.

To find the true dip of the interface, two pairs of reversed lines at right angles can be used. Each is used to deduce a dip, α_1 and α_2 (Fig. 6.10b), from which the true dip, α, and its direction can be calculated from

$$\sin \alpha = \sqrt{\sin^2 \alpha_1 + \sin^2 \alpha_2} \qquad \text{Eq. 6.8a}$$

$$\cos \theta = \frac{\sin \alpha_1}{\sin \alpha} \qquad \text{Eq. 6.8b}$$

where θ is the angle between the dip direction and the seismic line that gave the dip component α_1.

6.5 Seismic velocities in rocks

The velocities found from travel-time diagrams give some indication of the types of rocks that form the layers, though rock types have a range of velocities. Velocities of some common rock types are given in Table 6.1.

In general, velocity increases with consolidation, so alluvium and loose sands have a very low velocity, cemented sandstones have a higher one, and crystalline rocks tend to have the highest velocities of all. Consolidation tends to increase with geological age, so a Palaeozoic sandstone, for example, may have a seismic velocity twice that of a Tertiary one.

Most rock types have a range of velocities, sometimes large, but in a particular area the range is often much less. Therefore, rock types can be identified more confidently if the velocities are measured in the area of interest – preferably in boreholes, for velocities measured on exposed rocks usually give a lower and more variable value because of weathering, opening of fractures under the reduced pressure, and pores not being fully saturated with water.

6.6 Hidden layers

There are two situations where a seismic interface is not revealed by a *t–x* refraction plot.

6.6.1 Hidden layer proper

It was explained in Section 6.2 that a refracted ray overtakes the direct ray, provided it travels a sufficient distance, but before that can happen it may be overtaken in turn by the refracted ray from the interface below. Figure 6.11a shows a 3-layer case where the second layer is revealed only by the short length of the line 'refracted ray 1' that is a first arrival. Suppose layer 2 were thinner: Then head rays from layer 3 would arrive earlier, displacing line 'refracted ray 2' downwards to the left, and no part of line 'refracted ray 1' would be a first arrival (Fig. 6.11b). Layer 2 is then **'hidden'**. Layer 2 would also be hidden if v_2 were decreased, or v_3 increased, for the first would displace crossover 1 to the right, while the second would displace crossover 2 to the left. If the second layer is hidden, the *t–x* plot will be interpreted as two layers, one with velocity v_1 over a layer with velocity v_3. The depth to the top of the third layer, calculated in ignorance of the existence of layer 2, will be intermediate between the true depths to the tops of the second and third layers.

6.6.2 Low-velocity layer

If a layer has a *lower* velocity than the one above there can be no critical refraction, the rays being refracted *towards* the normal (Fig. 6.12). There is no refracted segment corresponding to the layer, so the *t–x* plot will be interpreted as a two-layer case. The calculated depth to the top of layer 3 will be exaggerated, because the slower velocity of layer 2 means that a ray takes longer to reach it, than if there really are two layers, v_1 over v_3.

The possibility of a hidden layer of either kind can be recognised only from independent information, particularly geological sections or borehole logs. Common low-velocity situations are sand below clay, sometimes sandstones below limestones, and most sedimentary rocks beneath a lava or sill (see velocities in Table 6.1). A hidden layer proper may be inferred, for instance, by comparing the seismic models with geological sections. Depending upon the information available, some correction may be possible.

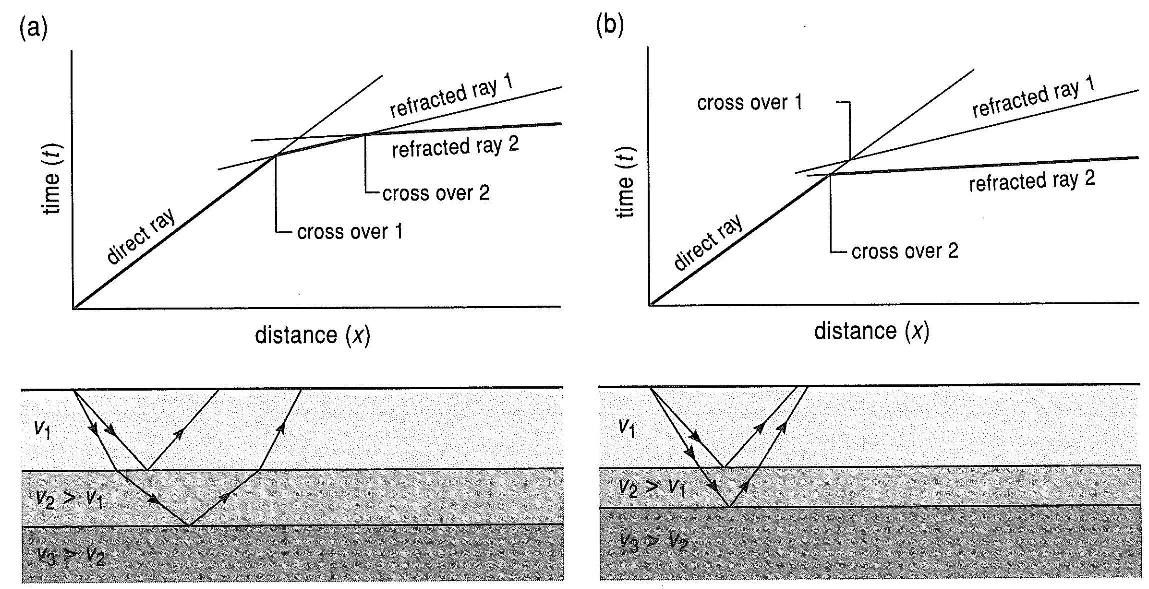


Figure 6.11 Hidden layer.

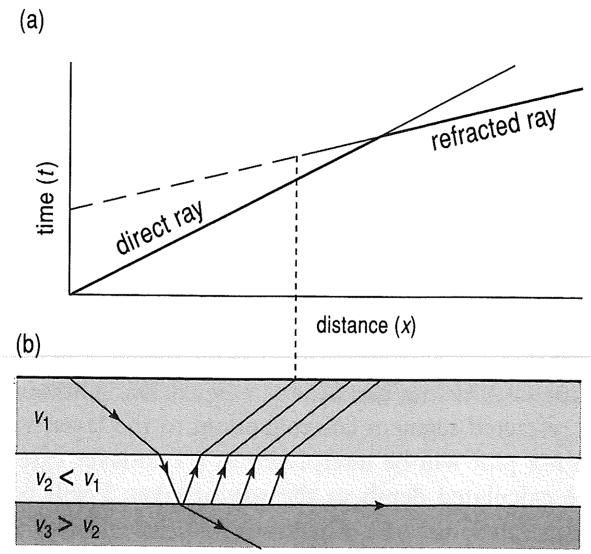


Figure 6.12 Low-velocity layer.

A final point to note is that a seismic interface is not necessarily a geological boundary and vice versa. For refracted arrivals to occur there must be an abrupt increase in velocity, that is, a seismic discontinuity. The water table may be such a interface, though it is not usually a geological boundary; conversely, Table 6.1 shows that two quite different rocks types – such as shale and sandstone – may chance to have similar velocities and so would not form a seismic interface.

6.7 Carrying out a seismic-refraction survey

Refraction surveys can be carried out on all scales from the shallow investigation of a building site to deep studies of the lithosphere that extend over hundreds of kilometres (e.g., Section 21.3.3). All require a source, a line of receivers, and a way of timing arrivals. Most modern recording is in digital form, where the seismogram is recorded as a series of binary numbers, as in a computer (described in Section 7.7.1), rather than analog form (wiggle trace) because the former gives better records and permits sophisticated processing of data to make corrections and extract the maximum amount of information, as outlined in Chapter 3.

Rather than discuss the various kinds of sources and receiver separately, it is better to describe complete systems.

Hammer seismics. Most sources send a pulse of waves into the ground. The simplest way to do this is to strike a plate on the ground with a sledgehammer. When the hammer strikes the plate a switch closes, starting the recorder, and the time of the first arrival, in milliseconds, is easily read from a screen (the duration of the flat part of the trace) or displayed as a number. Because it is so simple to strike the plate only a single geophone is needed, and it is moved progressively along a line.

The maximum workable distance depends upon the sensitivity of the system, the size of the hammer blows, the subsurface lithologies, and noise. Noise can be due to traffic and machinery, and to wind, which disturbs the ground by shaking trees, buildings, fences, and so on, or by directly shaking the geophones and cables. This last can be reduced by burying the geophones and laying the cables flat along the ground. Poor arrivals can also be improved by stacking: The plate is struck repeatedly

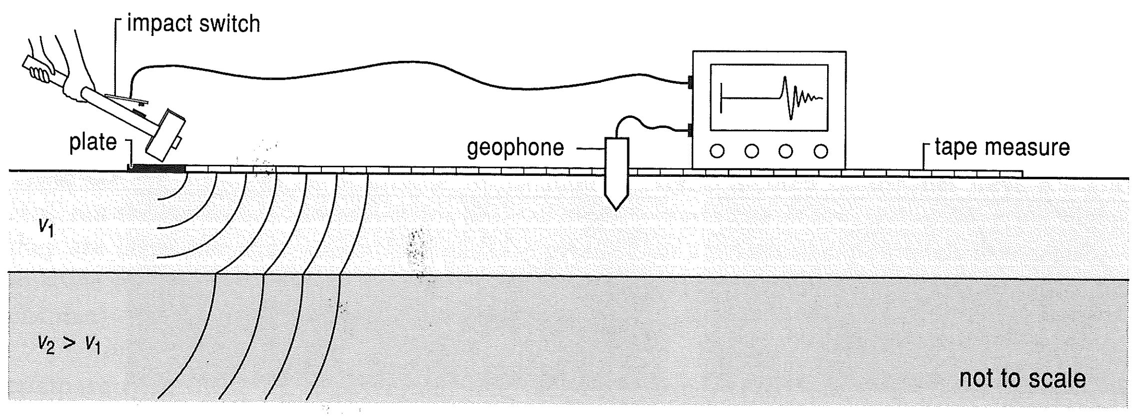


Figure 6.13 Hammer system.

with the geophone in the same position, and the arrivals are added together to average out chance noise but not the arrivals and so increase the signal-to-noise ratio. In practice, the maximum distance is, roughly, 100 m. Using a rule of thumb that the geophone line should be about ten times the depth to the interface gives a detectable depth of about 10 m, sufficient for many small-scale surveys. The range can be extended by using bigger sources such as a large dropped weight or mechanical hammers.

Explosion seismics. For larger-scale surveys a more powerful source is needed; on land a charge of explosives is commonly used. Since the object is to send seismic energy into the ground and not to produce a spectacular but wasteful blowout, the charge is buried to a sufficient depth, usually in a drilled hole. As boreholes for large charges (which can be tonnes in weight for lithospheric studies) are expensive – and the results poor if the charge is not below the water table – charges are often fired in water, making use of ponds, lakes, or the sea, as available. Firing in water is so much cheaper and generally more effective that a survey line is often chosen to take advantage of suitable bodies of water.

Because of the effort and expense of using explosives, a single receiver ceases to be practical, for it would entail a separate explosion for each receiver position. Instead, a line of geophones or seismometers is laid out. Lines up to a few kilometres in length can be connected to a central recorder by cable, but for surveys extending tens or hundreds of kilometres, cables are impracticable and each receiver has its own accurate clock, often synchronised to a master clock by radio, to time arrivals, or signals are radioed to a central recorder.

Far from the source the signal can be very weak, and high-sensitivity seismometers are essential. These are set up carefully, often in a pit to reduce noise and make contact with more solid ground. Three seismometers are often used, two horizontal instruments at right angles plus a vertical one, to record all components of the ground motion (an example is shown in Figure 5.22).

Though a large-scale survey – designed to investigate structure down to the Moho or deeper – is, in essence, no more than the hammer seismics survey greatly magnified, the logistical effort required may need years of planning and preparation.

General remarks about land surveying. Often there are near-surface layers with low velocities; in small-scale surveys these are often soil, subsoil, and the weathered top of the rock below. Though these layers may be of little interest in themselves, they often have low velocities, so the *time* spent in them may be significant even if they are thin. If receivers are spaced too far apart, these interfaces are not recognised (compare Fig. 6.14b with 6.14a), and then layers may be lumped together and the time spent in them attributed to a single layer with a faster velocity and greater thickness, and so produce an incorrect interpretation. To ensure that they are detected, receivers may need to be spaced more closely, particularly near the source (ideally, receivers would be close all along the profile, but this would be too expensive).

Surveys at sea. These generally use different sources and receivers from those used on land. As well as explosives, sources include a range of devices, which use compressed air, high-voltage discharges, or other types of energy sources. As these have their most important use in reflection seismology, they are described in Section 7.7.1. Receivers are usually **hydrophones**, immersed in the water to respond to *pressure* changes rather than water motion caused by the passage of P-waves (S-waves, of course, cannot propagate through water, Section 4.5.2). As shooting and receiving are carried out on the move, long lines, or a set of parallel lines, can be easily surveyed, an advantage offsetting the high cost of ships.

For the largest-scale surveys explosives are often used. These are dropped overboard from a moving ship and fired when safely astern, with seismic energy into the water maximised by firing at a depth that increases with the size of the charge. Two ships may be used to reverse lines. Sonobuoys may be left floating to record while the source ship sails away. Ocean-bottom seismometers that sink to the ocean floor and record like land seismometers (before rising again to the surface for retrieval), are increasingly used to improve the quality of records and provide all three components of motion.

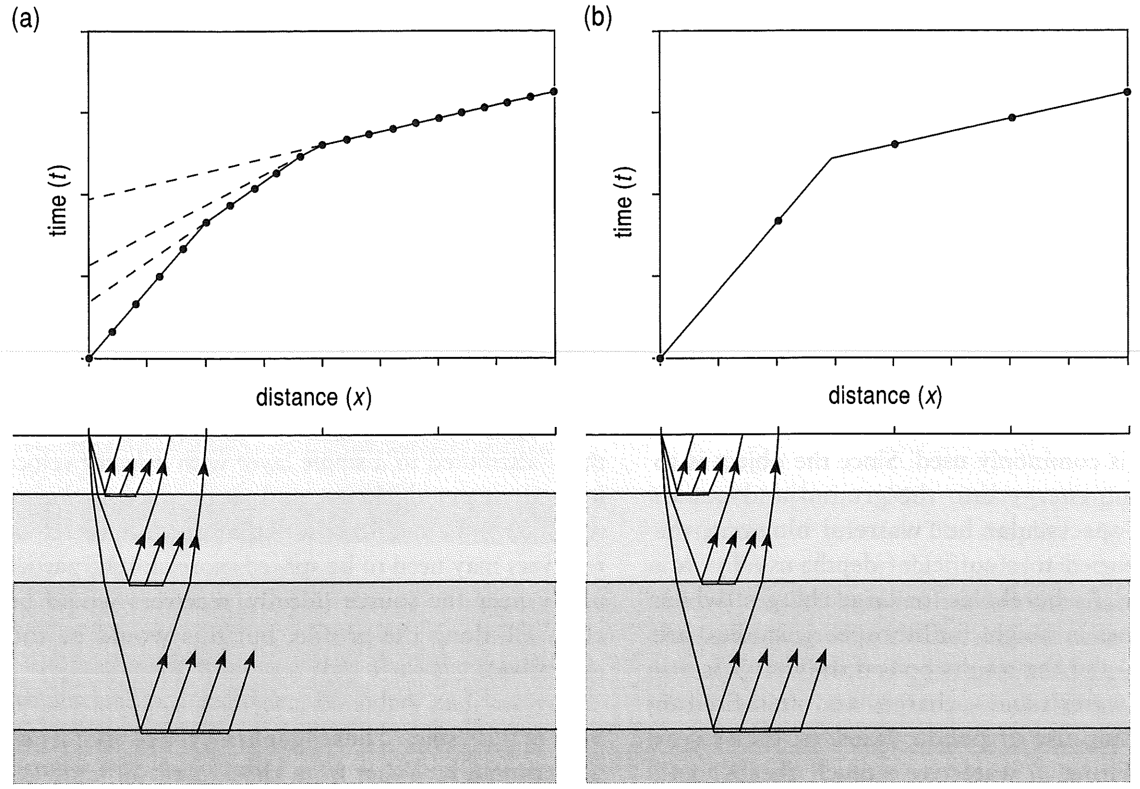


Figure 6.14 Effect of slow surface layers.

Having described how refraction surveys are carried out, we return to other layer geometries.

6.8 Undulating interfaces and delay times

If interfaces are not flat, a more sophisticated method of analysis is needed. Figure 6.15 compares the actual, undulating interface with a flat reference interface joining C_f to C_r. The forward and reverse refraction lines for this reference interface are shown dashed in the t–x diagram. As M, for instance, is closer to the surface than the reference interface, the actual travel time to M' plots below the reference line; conversely, that for N' is above it. The same argument applies to the reversed line, so the shapes of the refraction lines on the t–x plot show qualitatively how the interface differs from a flat surface. Note that the vertical separation of the forward and reversed refraction lines is about the same for the actual and reference interfaces.

These observations can be made more precise using the concept of delay times.

6.8.1 Delay times

The travel-time of a refracted ray is made up of three parts: travelling obliquely down to the interface, along the interface, and up to the receiver (Fig. 6.16a). But we can think of it as being made up in a different way (Fig. 6.16b): the time it takes to travel the distance between source and receiver, S_vR_v, just below the interface at velocity v_2, plus a term δ_S at the source end to equal the *extra* time it takes to go SC at velocity v_1 compared to going S_vC at v_2, and similarly δ_R at the receiver end. In effect, we are pretending that the ray travels all the way at v_2, but there is a delay between the shot time and when the ray starts on its way, and another after it finishes before the signal is recorded by the receiver:

$$t_{SR} = \delta_S + \frac{SR}{v_2} + \delta_R \qquad \text{Eq. 6.9}$$

δ_S and δ_R are called **delay times** (or sometimes time terms).

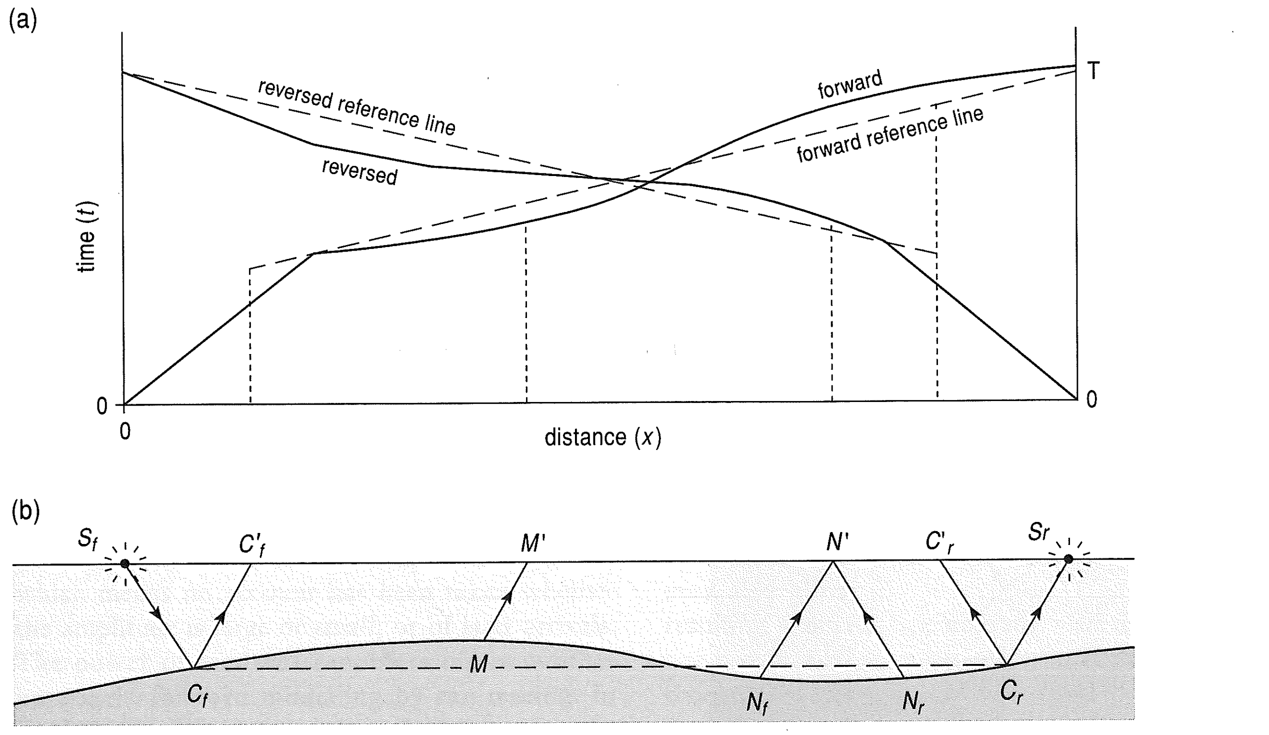


Figure 6.15 Undulating interface.

The delay time for a receiver is easily measured (Fig. 6.17a). The time, t_f, to go from one end to a receiver (path S_fCDR), and then on to the other end, t_r (path $REFS_r$), is longer than the time, t_{total}, to go from end to end (along S_fCDEFS_r) because of the extra times taken to travel from intercept to receiver, along DR and ER. As each of these extra times is just the delay time, we have

$$t_f + t_r = t_{total} + 2\delta_R \qquad \text{Eq. 6.10}$$

$$\delta_R = \frac{1}{2}\left(t_f + t_r - t_{total}\right) \qquad \text{Eq. 6.11}$$

As t_f, t_r, and t_{total} can all be read off the t–x diagram, δ_R can be calculated. This can be done for each receiver.

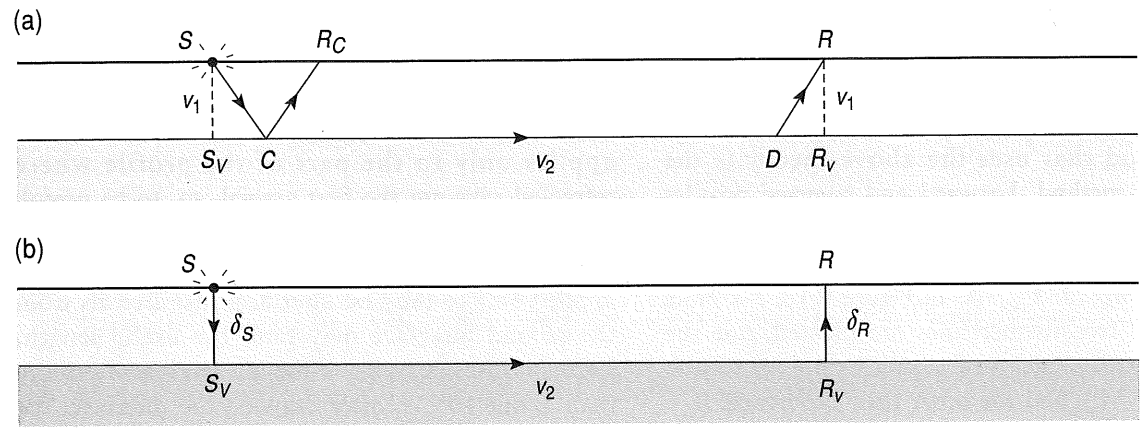


Figure 6.16 Two ways of treating travel-times.

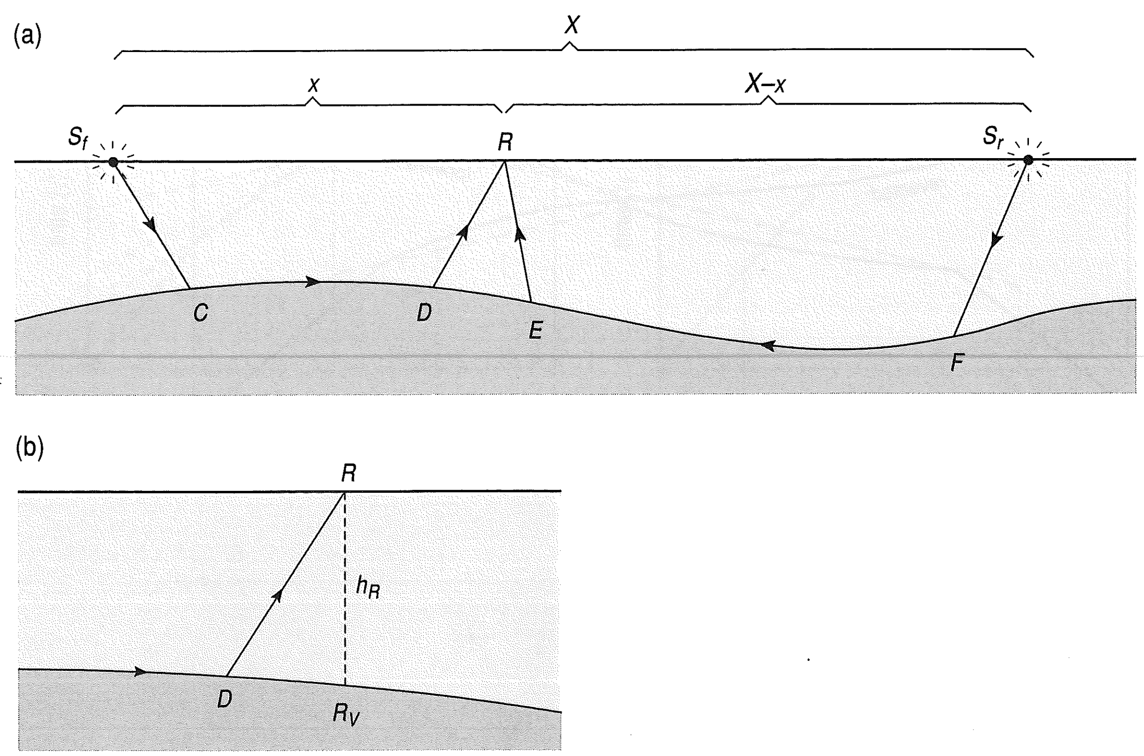


Figure 6.17 Finding delay times.

The depth to the interface can be deduced because obviously the deeper it is the larger is the delay time. The delay time is the extra time to travel from D to R, compared to D to R_v, that is $(DR/v_1 - DR_V/v_2)$ (Fig. 6.17b), and it can be shown that

$$h_R = \delta_R \frac{v_1 v_2}{\sqrt{v_2^2 - v_1^2}} \qquad \text{Eq. 6.12}$$

This can be used to find the depth to the intercept, once v_1 and v_2 have been found.

6.8.2 The plus–minus method

One method that uses the above theory is the **plus–minus method.** Forward and reversed profiles are measured in the usual way, though it is essential that forward and reverse lines are exactly the same lengths, to provide t_{total}. As in Figure 6.18, t–x plots are drawn. Two further lines are plotted, one the sum of the forward and reverse times for each receiver, $(t_f + t_r)$, and the other their difference, $(t_f - t_r)$, as in Figure 6.18a; this is why it called the *'plus–minus' method* (it is also known as Hagedoorn's method). The minus line should be straight; this is a test that the method is applicable. Its slope is $2/v_2$, from which v_2 is found. v_1 is found from the slope of the direct lines, in the usual way.

For each receiver, the end-to-end time, t_{total}, is subtracted from the added times of forward and reverse arrivals, and the result is halved to give the delay time (Eq. 6.11). This value is put into Eq. 6.12 to give the depth to the interface below the receiver. Finally, below each receiver an arc is drawn proportional to this depth (Fig. 6.18b) and a line is drawn through them to show the interface.

There are some points to note. The method applies only to the part of the profile where refracted rays are the first arrival, so, to be useful, the profile should be long compared to the crossover distances. As the method is usually applied only to the first interface, first arrivals from the second interface may limit the useful length. Lastly, it loses accuracy if the dip anywhere is more than about 10°; if, after drawing the interface, the dips are found to be steeper than 10°, the general

shape will still be correct but the depths will not be accurate.

There are other methods for dealing with uneven interfaces, but only a very general one, ray tracing, will be described.

6.9 Ray tracing and synthetic seismograms

The methods described so far for interpreting refraction data have all used inverse analysis; that is, starting with arrival times, the subsurface structure has been deduced by drawing travel-time diagrams, measuring slopes, and so on, and putting the values in equations. This works well for interfaces with simple geometries but becomes more difficult as the geometry becomes more complex. In addition, the methods use only the times of first arrivals, which means no account has been taken whether the amplitude is large or small, or of later arrivals. The power of modern computers offers another approach, forward modelling by **ray tracing.** In essence, a structure is guessed (based on a model deduced using the methods described above) and the paths of rays leaving the source at different angles are traced through the structure by using Snell's Law at each interface. The velocities can also be allowed to vary between interfaces, which causes the rays to curve. The time for each ray to return to the surface is calculated and used to construct a travel-time diagram, which is compared with the observed travel-times and thence modified as necessary. Figure 6.19 shows an example of ray tracing, with rays in the lowermost layer curving because of a significant downward increase in velocity. This example is discussed more fully in Section 21.3.3.

More sophisticated computer programs also calculate the amplitudes, taking into account loss of energy at interfaces and due to spreading out of the wave front as it travels further from the source, and may also allow for absorption. Comparing the resulting **synthetic seismogram** with the observed one provides a better test of the model than travel-times alone, for arrivals have to match the observations in amplitude as well as time.

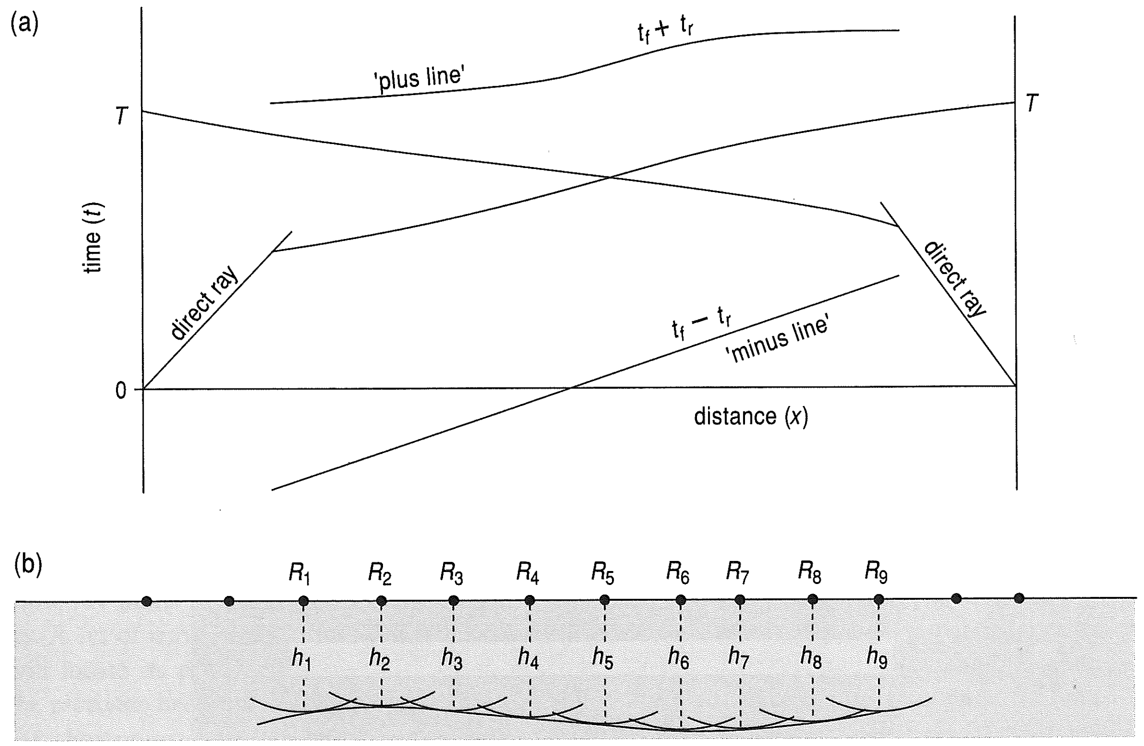


Figure 6.18 Plus–minus plots.

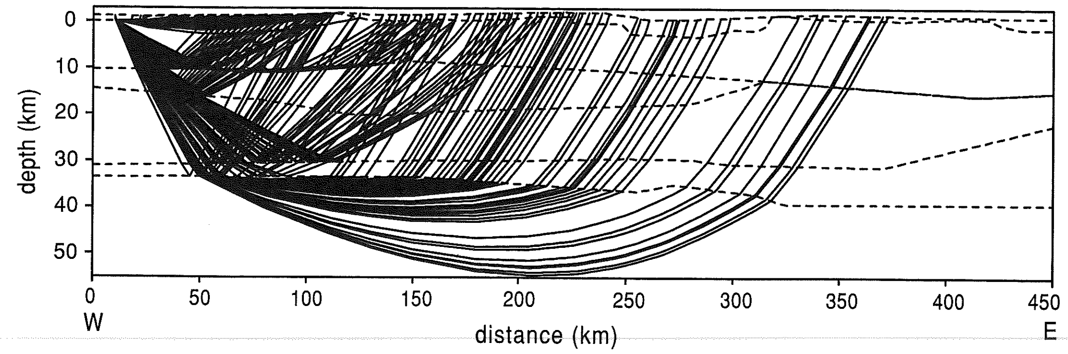


Figure 6.19 Example of ray tracing.

6.10 Detecting offsets in interfaces

Refraction seismology is mostly about measuring the depths to continuous, roughly horizontal interfaces, but it may be used to detect a fault if it offsets an interface. In Figure 6.20b a critically refracted ray travels just below the interface, producing head waves as usual along the interface from C to D. But what happens beyond the offset: Does the corner E cast a shadow preventing seismic energy reaching EF . . . ? This is best answered by considering waves. F is not in shadow because waves diffract into it, as explained in Section 6.1.1, and therefore head waves originate along EF. . . . The t–x plot (Fig. 6.20a) has an offset of the refraction line: Little energy arrives in the interval D' to E', and when arrivals resume at E' they are later than those at D' not just because it is further from the source but also because E is further from the surface.

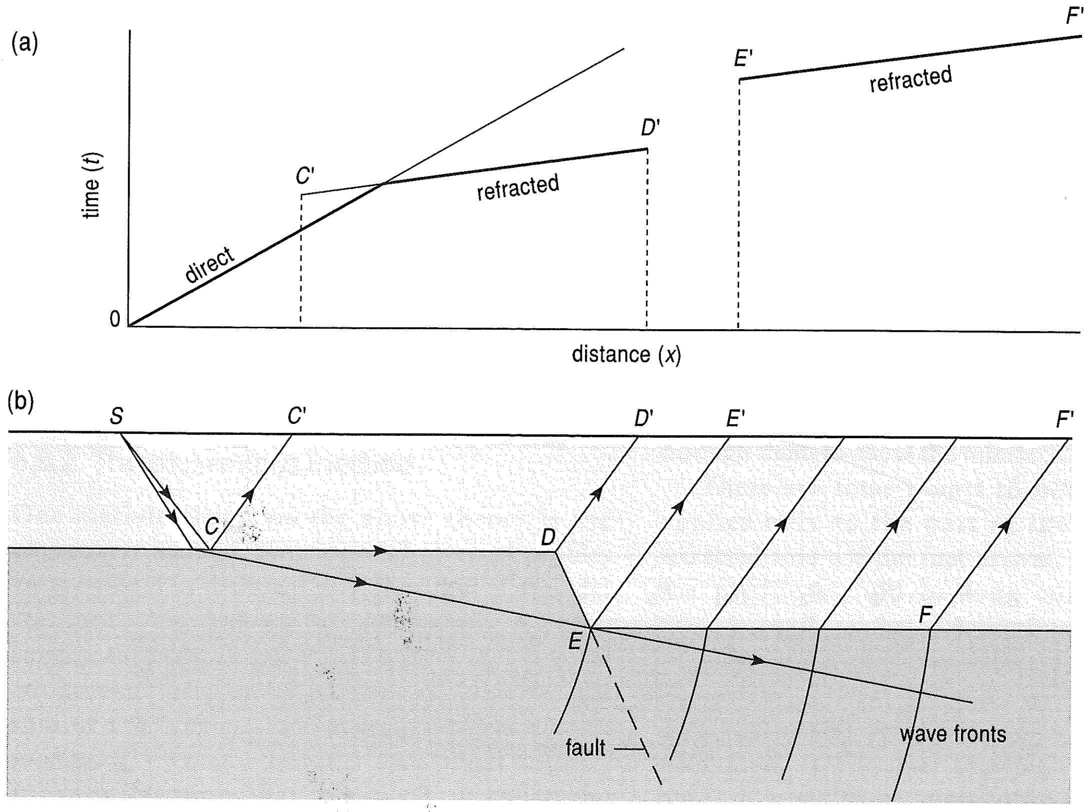


Figure 6.20 Faulted interface.

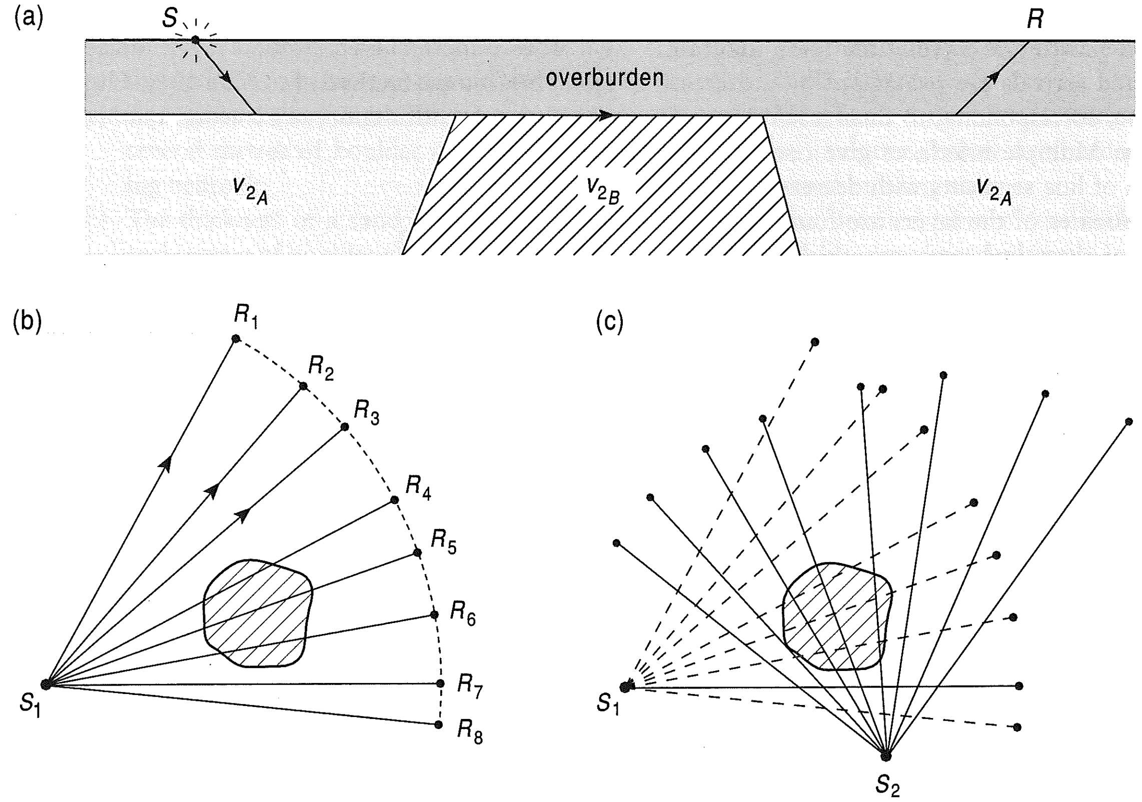


Figure 6.21 Fan shooting.

6.11 Fan shooting: Simple seismic tomography

This is a method that can detect steep-sided features even if they do not have a sharp boundary. Suppose the structure beneath the interface is as shown in Figure 6.21a, perhaps an intrusion that was truncated by erosion and then covered by sediments. Critically refracted rays travel just below the interface as usual, but those passing through the intrusion go faster and so have shorter travel-times. If receivers are arranged in an arc about the source, called **fan shooting** (Fig. 6.21b), the travel-times to R_4, R_5, and R_6 will be less than to the other receivers. This will reveal that *somewhere* along these ray paths is a body with higher seismic velocity. A set of travel-times using a different shot point will locate its position closely (6.21c). In practice, the receivers need not be all the same distance from the shot points and can be placed to record both shots without needing moving, though analysis is somewhat more complicated.

Fan shooting is an simple example of seismic tomography (Section 4.6). In the early days of oil prospecting in Texas it was used successfully to find salt domes because these were often associated with oil; the high seismic velocity of salt compared to most near-surface rocks made them fairly easy to locate.

Summary

1. Refraction seismic surveys are mainly used to detect roughly horizontal interfaces separating layers with different seismic velocities.
2. The seismic refraction method utilises head waves, which are generated by a critical ray that travels just below the interface. Head waves leave the interface at the critical angle.
3. Refracted rays arrive only beyond the critical distance but give first arrivals only beyond the crossover distance; up to the crossover distance the first arrivals are direct rays.

4. Seismic refraction results are usually interpreted by first plotting a travel–time (t–x) diagram. Refracted arrivals are recognised by a decrease of slope, corresponding to their faster apparent velocity. Multiple interfaces give rise to a succession of line segments with decreasing slopes. The velocities of the layers are found from the slopes of these line segments, thicknesses from their intercepts, using appropriate formulas.
5. The value of the velocities can help identify the rock type of a layer, but not uniquely because rocks types have a range of velocities. Ambiguity can be reduced by comparing values with the velocities of known rocks in the area.
6. Survey lines are reversed to allow the dips of interfaces to be detected and measured. The down-dip refraction line has a smaller intercept but greater slope than the up-dip line.
7. An interface is not detectable if:
   (i) The layer below has a lower velocity than the layer above (low-velocity layer).
   (ii) Though it has a higher velocity, its velocity and thickness are such that no refracted rays from the interface arrive before those from the interface below (hidden layer).
8. An undulating interface may be investigated using delay times. One method is the plus–minus method.
9. A general and powerful method for interpreting the data from complex structures is modelled by ray tracing, which uses a computer to adjust models until calculated travel-times and traces agree with those observed.
10. An offset in an interface produces an offset in the refraction line. Steeply sided bodies – such as an intrusion – can be found using fan shooting, a simple tomographic method.
11. Seismic interfaces need not be geological boundaries, and vice versa.
12. Seismic refraction surveys require some kind of source and one or more receivers, often laid out in a line. Sources range from a hammer blow to a ton or more of explosive detonated in a borehole or underwater. Receivers are geophones or seismometers on land, hydrophones in water.
13. You should understand these terms: shot point, hydrophone; direct ray, refracted ray, head ray, reflected ray; critical angle and distance, crossover distance; time–distance diagram, forward and reversed profile, apparent velocity; low-velocity layer, hidden layer; delay time, plus–minus method; fan shooting; Huygens's wavelets, diffraction; ray tracing, synthetic seismogram.

Further reading

The basic theory of seismic refraction is covered in standard textbooks on applied geophysics, such as Doyle (1995), Kearey and Brooks (1991), and Reynolds (1997) – which also gives case histories of environmental interest, Robinson and Coruh (1988), and – at a more advanced level – Telford et al. (1990). Milsom (1996) gives advice for carrying out surveys.

Problems

1. Explain, with the aid of a ray diagram, how direct, refracted, and reflected rays can reach a receiver on the surface, given that the subsurface consists of a layer overlying higher-velocity material. Explain why not all the rays can be received at all distances from the shot point.
2. Compare the relative arrival times of direct, refracted, and reflected rays (i) close to the shot point, (ii) at the critical distance, (iii) at the crossover distance, (iv) a long way from the shot.
3. Explain why it is important to determine the structure and velocities of thin, near-surface layers even though the interest is in deeper layers.
4. Sketch the t–x diagram for reversed profiles for an interface that is horizontal. Show how the diagram changes when the interface is tilted about a point midway between shot points.
5. A seismic survey is to be carried out to measure the approximate thickness of overburden, which is thought to be about 4 m thick. Deduce approximate distances from the shot point at which geophones should be placed to measure the depth of the overburden, given that it has a velocity of 1.4 km/sec and overlies sandstone with a velocity of 2.4 km/sec.
6. Explain why a hammer refraction seismic survey is often carried out using only one geophone, whereas in explosion seismics it is usual to use a line of them.

7. Consider whether refraction surveys could be used in the following cases, and if so, how surveys would need to be carried out to determine:
   (i) The approximate depth to bedrock, below several metres of boulder clay and overlying bedrock.
   (ii) The thickness of a sand layer beneath several metres of boulder clay.
   (iii) Whether a thin igneous sill several metres thick underlies an area.
   (iv) The position of a normal fault in sandstones beneath shale, beneath overburden.
   (v) The dip of sedimentary strata beneath unconsolidated deposits.
   (vi) The dip of the contact between a granite and the country rock which can be seen on the surface.
   (vii) Whether a sinkhole filled with clay exists in limestone, below overburden, on a proposed building site.
   (viii) The detailed form of the bedrock surface of a proposed road cutting.
8. A reversed refraction line is shot along the direction of strike of an interface that dips at about 10°. How do ray paths differ from the case of (a) across the strike, (b) a horizontal interface? How can the three cases be distinguished seismically, if necessary by carrying out further surveys?

Chapter 7

Reflection Seismology

Seismic reflection is the most important tool we have for detailed imaging of approximately horizontal layering within the Earth, in three dimensions if required. It can reveal structural features such as folding and faulting. It is extensively used by the oil and gas industry to search for oil and gas fields, and to exploit them successfully.

An important branch of reflection seismology is seismic stratigraphy. On a regional scale it has been developed as sequence stratigraphy, revealing stratigraphic correlations that date different parts of a sedimentary succession, and even to establish eustatic sea-level changes. On a local scale, it can reveal lateral changes of lithology that may form stratigraphic oil traps.

Seismic reflection is a form of echo sounding, detecting 'echoes' from seismic interfaces. It can yield results that are the closest of any geophysical technique to a conventional geological section. However, there are several reasons why a seismic section must not be accepted uncritically as a geological section.

7.1 Seismic-reflection sections and their limitations

Reflection seismology, in essence, is echo or depth sounding. Carried out at sea, it is like conventional depth sounding, except that the echoes of interest come not from shoals of fish or the sea floor but from within the solid Earth.

It is most easily explained by describing a simple example (Fig. 7.1). A ship sails along emitting pulses of seismic energy, which travel downwards, to be partially reflected back up from the sea floor and from interfaces in the rocks, called **reflectors** (Fig. 7.1a). When a reflected pulse reaches the surface it is detected by seismic receivers. At the same instant as the pulse is emitted, a pen begins to move steadily across a roll of paper (Fig. 7.1b); it is connected to the receiver so that every reflected pulse produces a wiggle on the trace, as shown. After a short interval, during which the ship has moved along, the process repeats, but as the paper is moving slowly along, each trace is slightly to one side of the previous one. The wiggles on the separate traces line up to show the interfaces.

As Figure 7.2 shows, the result, a **seismic section,** can give a very direct picture of the subsurface structure, but it is not a true vertical section, for several reasons:

(i) The vertical scale is not depth but time (usually the time to reach the reflector *and return:* the **two-way time, TWT**). Since velocity varies with depth, times cannot be easily converted into depths, as they can for measuring water depths.

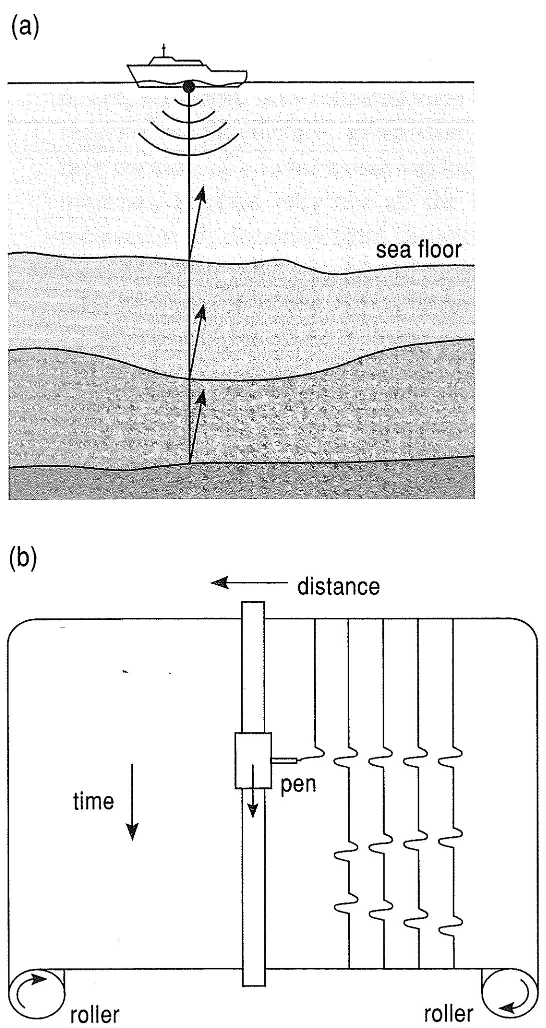


Figure 7.1 Simple-reflection profiling at sea.

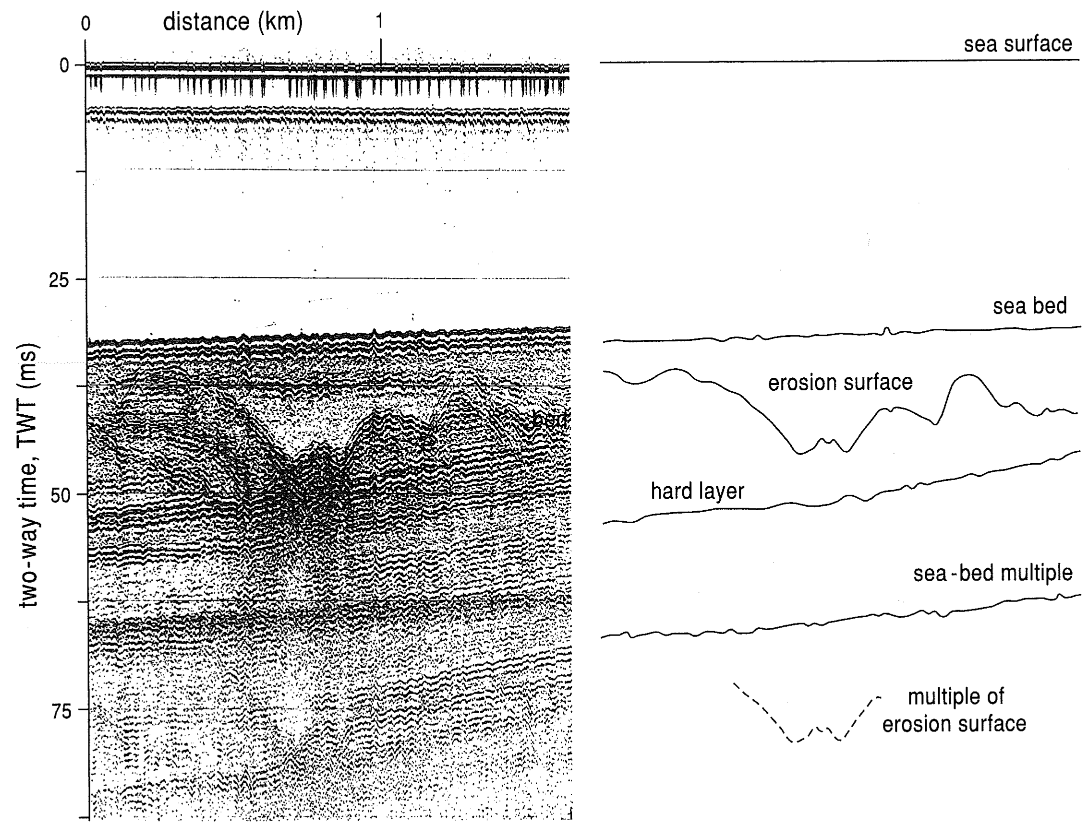


Figure 7.2 Seismic section of a buried channel.

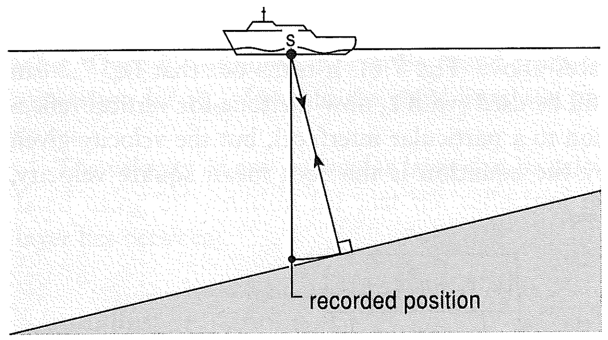


Figure 7.3 Nonvertical reflection.

(ii) Reflections may not come from directly below the source, since they reflect at right angles to the interface (Fig. 7.3), but the recording takes no account of this.

(iii) There may be multiple reflections in addition to the primary reflections (Fig. 7.4), and these produce spurious reflectors on the record, as shown in Figure 7.2.

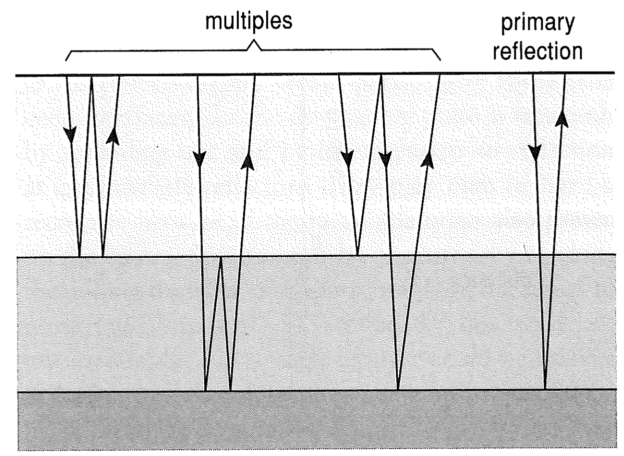


Figure 7.4 Multiple reflections.

(iv) Further reasons, which will be pointed out at appropriate places in the chapter.

How these difficulties can be recognised and largely overcome is explained in the sections that follow.

7.2 Velocity determination using normal moveout, NMO

To deduce the velocities – and hence depths – of the layers, we need a number of receivers along the line, offset to one or both sides of the shot point, so that most rays do not travel vertically (Fig. 7.5). The shortest reflected ray path (for a horizontal reflector) is the vertical one; the rays that reach receivers to either side travel progressively further, taking extra time, as shown by the curve (called a hyperbola) on the time-distance (t–x) diagram above. The *extra* time, Δt, depends on the velocity (as well as the thickness) of layer 1. It is calculated as follows: The vertical TWT, t_0, for the first reflector is just twice the depth divided by the seismic velocity, $2h_1/v_1$. The TWT, t, to reach offset receiver R is SAR/v_1, or $2SA/v_1$. SA is found by applying Pythagoras's theorem to right-angle triangle SVA:

$$t=(t_0+\Delta t)=\frac{2}{v_1}\sqrt{h_1^2+\left(\frac{x}{2}\right)^2} \qquad \text{Eq. 7.1}$$

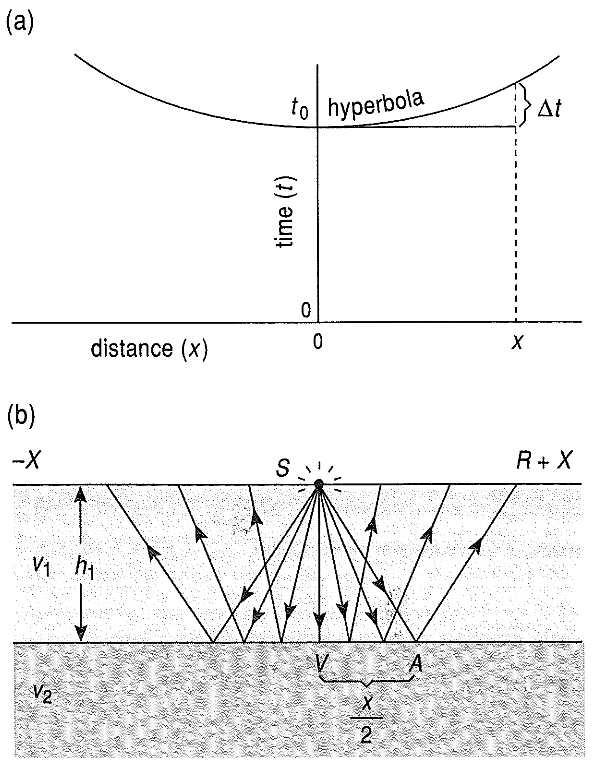


Figure 7.5 Offset receivers.

Squaring both sides, and replacing $2h_1/v_1$ by t_0, we have

$$t^2=(t_0+\Delta t)^2=t_0^2+\frac{x^2}{v_1^2} \qquad \text{Eq. 7.2}$$

If the offsets are small compared to the thickness of the layer, so that the rays are never more than a small angle away from vertical, as is often the case, the extra time Δt is approximately

$$\Delta t=t-t_0\approx\frac{x^2}{2v_1^2t_0} \qquad \text{Eq. 7.3}$$

The later time of arrival of receivers offset from the source, for a *horizontal* reflector, is called **normal moveout, NMO.**

This equation can be rearranged to give v_1:

$$v_1\approx\frac{x}{\sqrt{2\,t_0\Delta t}} \qquad \text{Eq. 7.4}$$

Values of t_0, Δt, and x are measured from the t–x curve.

Multiple layers. The preceding formulas are true only for the topmost reflector. For deeper reflectors we have to allow for refractions through the interfaces above (Fig. 7.6). It turns out that Eq. 7.2 can still be used (with t_0 now the time for vertical reflection to a particular interface), but the velocity given by the equation is the root mean square velocity, v_{rms}:

$$v_{\mathrm{rms}}=\sqrt{\frac{\left(v_1^2\tau_1+v_2^2\tau_2+\mathrm{L}\right)}{\left(\tau_1+\tau_2+\mathrm{L}\right)}} \qquad \text{Eq. 7.5}$$

where τ_1, τ_2, . . . are the one-way times spent in each layer (Fig. 7.6b). A special sort of average, v_{rms} is the velocity that a single layer, with thickness equal to the depth to the interface, would need to give the same TWT and NMO as observed.

However, to be able to deduce the thickness of each layer we need to know the velocity of each layer. These are found from the values of v_{rms} for successive interfaces. Firstly, v_1 is found using Eq. 7.2. For the second interface v_{rms} is

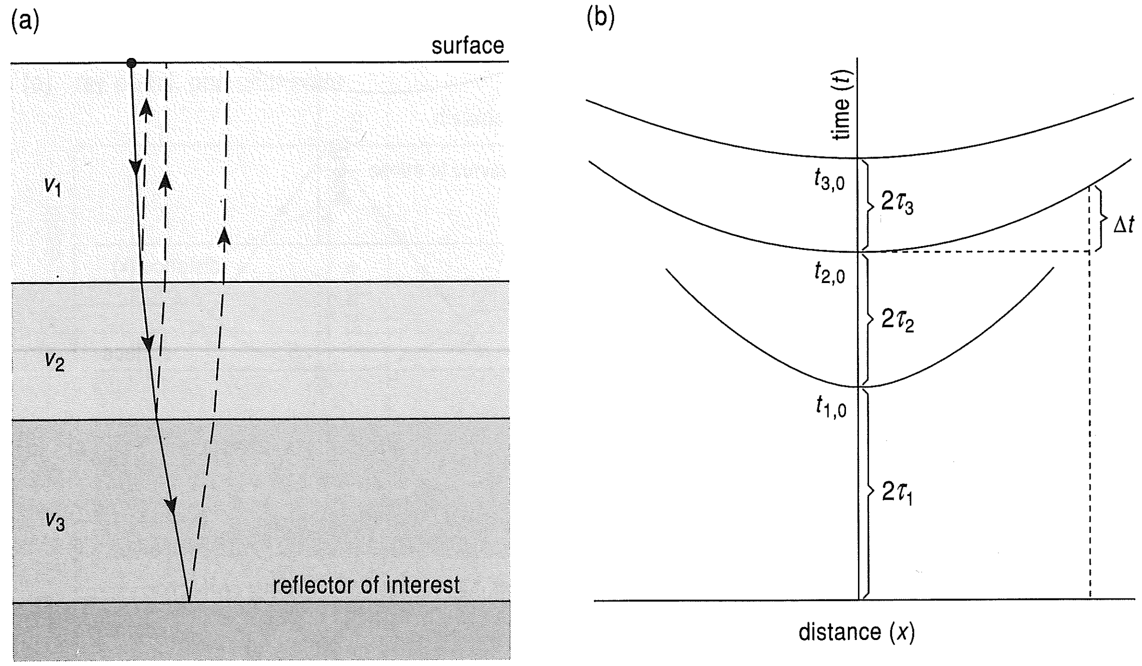


Figure 7.6 Offset ray paths with several layers.

$$v_{\mathrm{rms}}=\sqrt{\frac{v_1^2\tau_1+v_2^2\tau_2}{\left(\tau_1+\tau_2\right)}} \qquad \text{Eq. 7.6}$$

As v_2 is the only value not known it can be calculated. Once v_2 is known, v_3 can be found from the expression for v_{rms} down to the third interface, and so on.

The velocity of any particular layer can be calculated using information for just the reflectors the layer lies between:

$$v_{\mathrm{layer}}=\sqrt{\frac{\left(v^2_{\mathrm{rms}.B}t_n-v^2_{\mathrm{rms}.T}t_{n-1}\right)}{t_B-t_T}} \qquad \text{Dix's Formula} \quad \text{Eq 7.7}$$

where t_T and t_B are the TWTs to the top and base reflectors. This is known as Dix's Formula or Dix's Equation, and the velocity found is known as the **interval velocity.** Deducing the velocity of the various layers is called **velocity analysis.**

The thickness of a layer, h, is found by multiplying its velocity by the time taken to travel vertically through it, which is half the difference of the TWTs to its top and base interfaces:

$$h=v_{\mathrm{layer}}\left(\frac{\tau_B-\tau_T}{2}\right) \qquad \text{Eq. 7.8}$$

7.3 Stacking

Often reflections are weak, particularly those from deep interfaces, for the downgoing pulse is weakened by spreading out and by losing energy to reflections at intermediate reflectors. They may then be hard to recognise because of the inevitable noise also present on the traces. Sections can be improved by stacking, the adding together of traces to improve the signal-to-noise ratio (Section 2.3). Section 6.7 described how repeated traces for a single receiver could be stacked; in reflection seismology the shot is not repeated but instead the records of the line of receivers are used. Before they can be stacked, each trace has to be corrected by subtracting its moveout, so changing the hyperbolas of Figure 7.6b into horizontal lines. But moveout is difficult to calculate when arrivals are hard to recognise. To surmount this impasse, finding velocities and stacking are done together. A value for v_{rms} is guessed and used to calculate the NMO at each receiver according to its distance from the shot point (using Eq. 7.3), and the times are displaced by the value of the NMO. If the velocity guessed was correct, the arrivals of the various traces will line up so that, when added together they produce a strong arrival; but if an incorrect velocity was guessed, the traces will not line up so well and then their total will be lower. Therefore, a range of velocities are used to calculate moveouts, and the one that results in the largest addition is used. This is repeated for each interface, for they have different values of v_{rms}. This method works despite noise on the traces, because the noise usually differs from trace to trace and so tends to cancel when stacked.

7.4 Dipping reflectors and migration

If a reflector is dipping, both its apparent position and dip on a seismic section are changed. This is most easily understood if we return to the single receiver system of Figure 7.1. A pulse sent out from S_1 (Fig. 7.7) reflects from P_1 and not from vertically below the shot point. But the recording device is just a pen moving across a roll of paper (or its electronic equivalent), so the reflected pulse is plotted as if it were vertically below S_1, at R'_1. Similarly for S_2, S_3, . . . , though the differences become smaller up dip. Consequently, the reflector appears to be *shallower* and with a *less steep* dip than is actually the case. This distortion cannot be recognised using only a single receiver, but a line of receivers gives rise to a hyperbola offset from the shot point (Fig. 7.8) because the shortest time is not for the ray that travels straight down from the shot point, but the one the returns vertically up from the interface. The hyperbola is displaced up dip by an amount $2h \sin \alpha$, where α is the dip of the interface (Fig. 7.8).

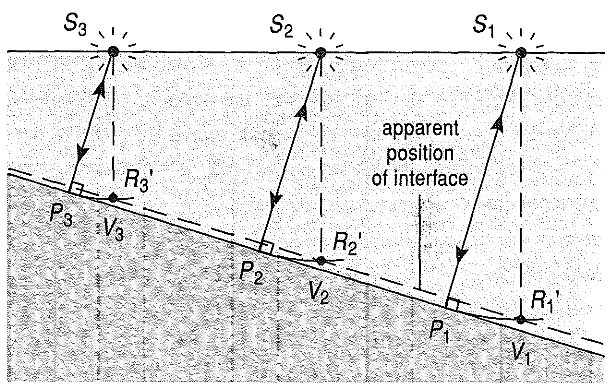


Figure 7.7 Displacement of a dipping reflector.

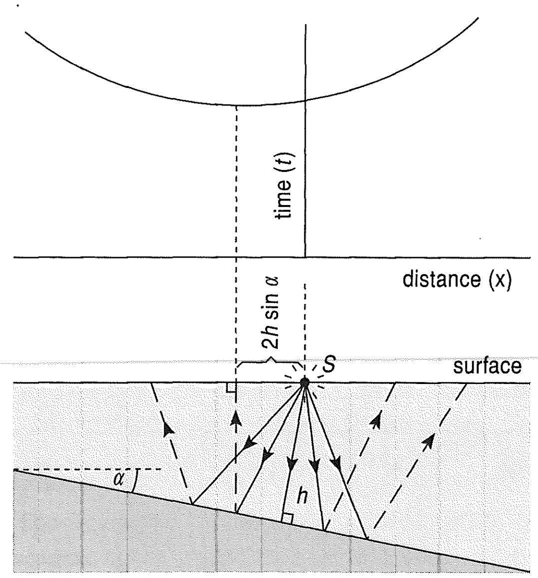


Figure 7.8 Moveout for dipping reflector.

If the reflector is curved, distortions may be more complex, for there may be more than one path between the reflector and a particular shot point. If the reflector is concave, as in a syncline, there will be three ray paths to the receiver, as shown in Figure 7.9a, and as they generally have different lengths – SP_1, SP_2, and SP_3 – they produce more than one arrival. The vertical ray will not be the shortest or the first to arrive if the bottom of the syncline is further from the shot point than the shoulders (i.e., the centre of curvature of the reflector is below the surface). As the shot point moves along, 1 to 10 (Fig. 7.9b), the shortest path switches from one side of the syncline to the other. The three arrival times for each of the transmitter positions 1 to 10 together produce a **'bow tie'** on the section (Fig. 7.9c). If the centre of curvature is above the surface, a syncline produces a simple syncline in the section, though somewhat narrowed; for an anticline, the effect is to broaden it (Fig. 7.9d and e).

Correcting for the displacement of the position and shape of a reflector that is not horizontal is called **migration.** This is complicated and requires such large amounts of computer time that sometimes it is not carried out; in this case the section will be marked 'unmigrated'. If so, anyone interpreting the structure has to be aware of the possible distortions, as comparison of Figure 7.10a and b shows. These are discussed further in the next section.

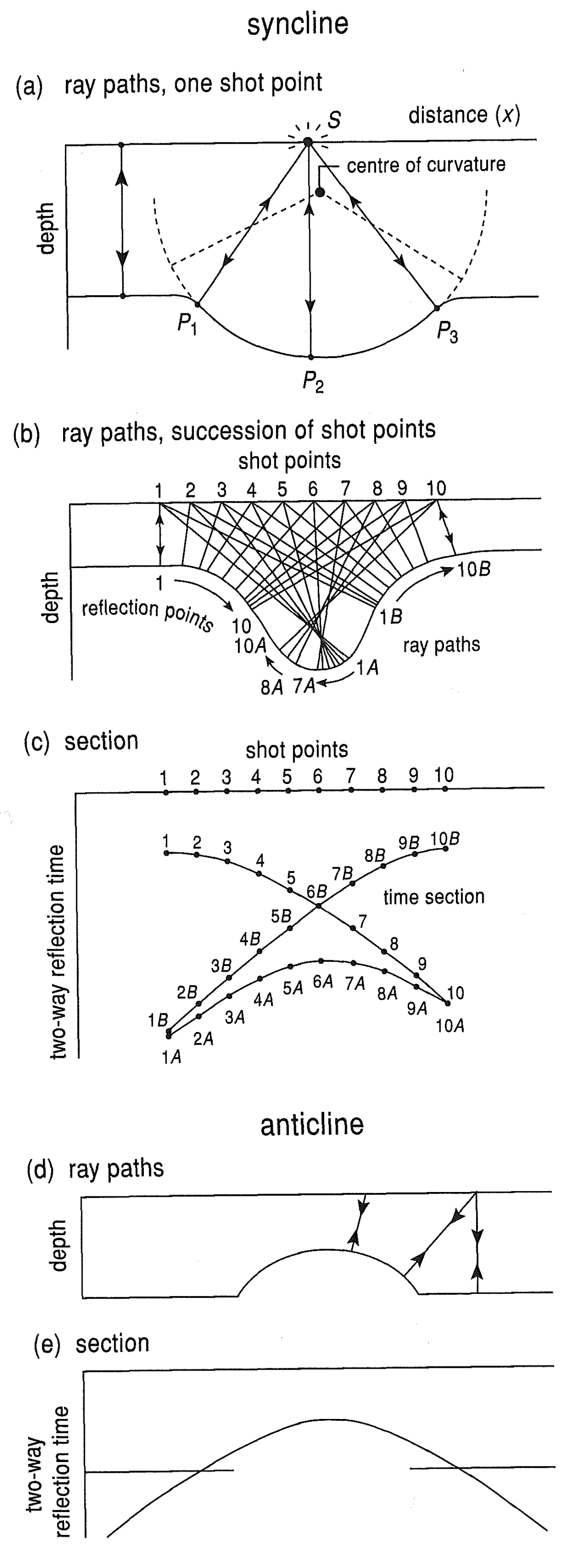


Figure 7.9 Distortion of synclines and anticlines.

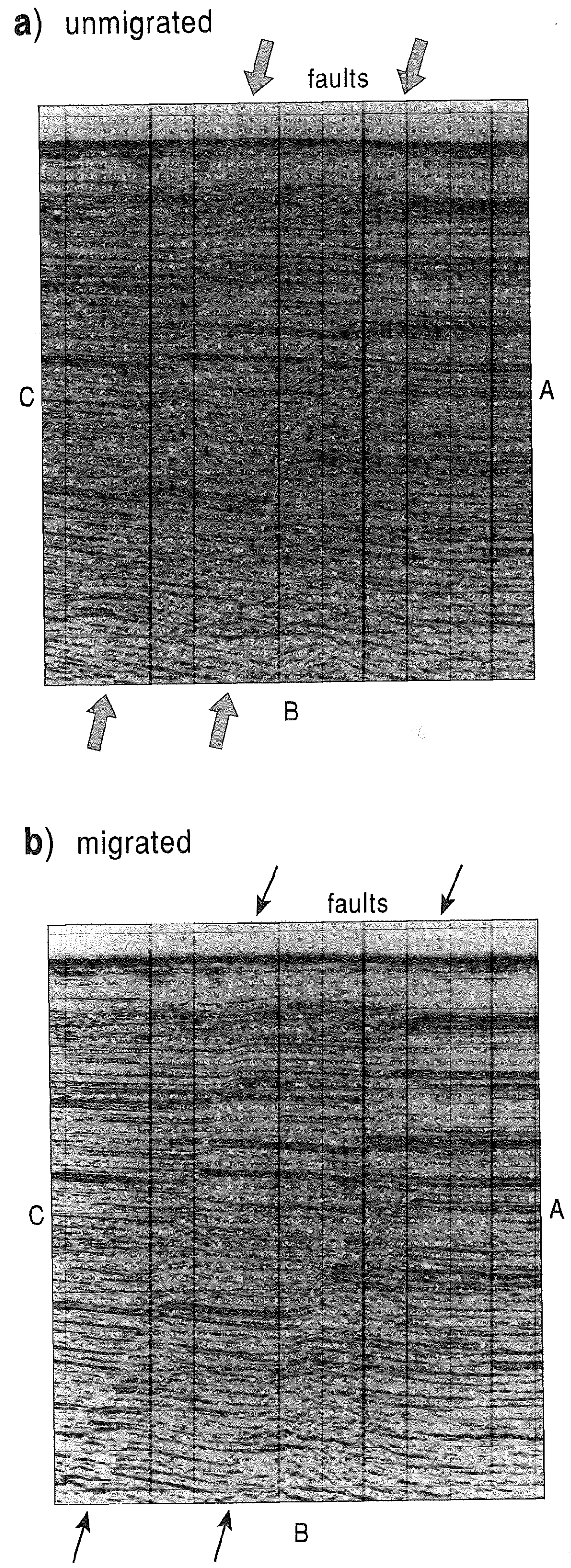


Figure 7.10 Unmigrated and migrated sections.

7.5 Faulted reflectors: Diffraction

You should understand by now that seismic reflection is about locating interfaces that are continuous and roughly horizontal. But an interface may not be continuous; for instance, it may have been offset by a fault. Such an abrupt end to a reflector produces diffracted waves, which are modelled using Huygens's wavelets, as explained in Section 6.1.1.

Consider first just a very short length of reflector (Fig. 7.11a). It acts as a point source; for example, when water waves encounter a pole sticking out of a pond, circular wave fronts travel outwards from the pole as if they were generated by the pole. As the source–receiver system approaches a buried reflector point it receives reflected waves from the point reflector, and these are recorded as if coming from below the transmitter–receiver position, at a distance equal to that of the reflector, as indicated by the arcs. Because this distance changes as the transmitter–receiver passes over the point, the resulting section shows a curve, known as a **diffraction hyperbola** (its amplitude decreases away from the centre of the hyperbola, because of the increasing distance of the reflector).

Turning to the stepped reflector (Fig. 7.11b), the Huygens's wavelets produce a horizontal, upward-moving reflected wave except near the ends, which act rather like point reflectors and produce diffraction hyperbolas. The overall result is that the two parts of the reflector are shown offset, but it is hard to tell exactly where the fault is located (Fig. 7.10a). Migration removes diffraction effects and reveals features more clearly (Fig. 7.10b).

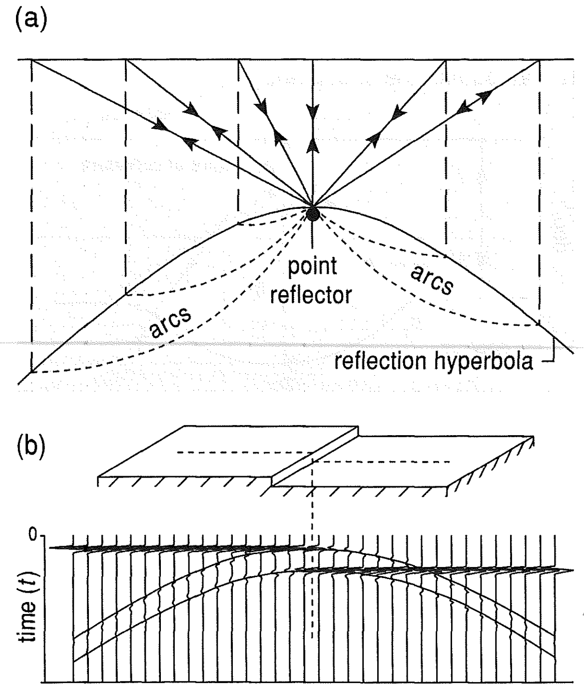


Figure 7.11 (a) Point and (b) stepped reflector.

7.6 Multiple reflections

Some **multiple reflection** paths were shown in Figure 7.4, and some spurious arrivals were shown in Figure 7.2. The strongest multiple reflections are those with the fewest reflections and, of course, from the strongest reflectors. A common one is the double reflection from the sea floor (Fig. 7.2).

The positions of multiples on the seismogram can be anticipated from the positions of the primary reflectors higher up the seismogram, but it may be difficult to recognise a primary reflector that happens to lie 'beneath' a multiple (i.e., has nearly the same TWT as that of a multiple from a shallower interface). However, these can be distinguished because the moveout for the primary is less than for the multiple and so it stacks using a higher velocity.

We have now considered the three reasons given in Section 7.1 for why a 'raw' reflection seismic section (i.e., as recorded) is not a true geological section (additional reasons will be given later). Next, we see how a reflection seismic survey is carried out.

7.7 Carrying out a reflection survey

7.7.1 Data acquisition

As surveys on land and on water are carried out differently, they will be described separately as complete systems.

Marine surveys. Pulses can be generated by a variety of devices. Commonly used is an air gun, in which air at very high pressure (over 100 times atmospheric pressure) is released suddenly. Less widely used is the sparker, in which a high-voltage electrical charge is discharged from a metal electrode in the water, like a giant spark plug. Other devices use a pulse of high-pressure water (water

gun), or explode a mixture of gas and oxygen (e.g., Aquapulse). Each of these sources has the ability to produce pulses that can be repeated at the required rate. The various types differ in the amplitudes of the pulses (dependent on their energy) and more particularly in the duration of the pulses they produce. The importance of a pulse's duration is explained in Section 7.8.2.

The reflected pulses – only P-waves, for S-waves cannot travel through water – are received by a line of hydrophones, which are essentially microphones that respond to the small changes of *pressure* produced by the arrival of a pulse, rather than to water movement. The hydrophones are usually mounted at regular intervals inside a **streamer** (Fig. 7.12), a flexible plastic tube filled with oil to transmit pulses from the water to the hydrophones. The streamer is towed behind a ship at a fixed depth below the surface. As the source has to be supplied with energy from the ship, it is placed in front of the streamer, and therefore the receivers are offset only to one side of the source. Commonly, several sources and many streamers are towed by a single ship, to cover an area more rapidly and thoroughly.

Land surveys. Explosive is probably the most common *impulsive* source used on land. Small charges (up to a few kilograms) are loaded into holes drilled through the top soil and weathered layer (typically a few metres thick), and each is fired in turn by an electrical detonator. Other sources include dropping a weight of 2 or 3 tonnes, or an adapted air gun.

The signals are received by geophones (described in Section 4.2). To improve the signal-to-noise ratio, each geophone may be replaced by a cluster of several, set out within an area only a few metres across, close enough for them to all to receive the signal from a reflector at practically the same time; their outputs are added together to increase the signal as well as partly cancelling any random noise, which differs between geophones. Moving the system along on land is much harder than on water, which makes land surveys about 10 times as expensive (the ratio for 3D surveying is even higher: See Section 7.9). A completely different kind of source, Vibroseis, will be described in Section 7.7.4.

Data recording. The output of each hydrophone or geophone (or cluster) is connected to an amplifier and forms one **channel** to the recorder, which nowadays records on magnetic tape.

The magnetic recording is usually digital (this is similar to that used in compact discs; the older 'wiggle trace' paper and magnetic tape recordings are called 'analog'). The signal is sampled at regular intervals, perhaps only a millisecond apart, and the height of signal at each interval is given as a number (Fig. 7.13). It is these numbers – in binary form, as used by computers – that are recorded and later changed back into the familiar wiggle trace. Digital recording both gives better records and make it easier to process them by computer.

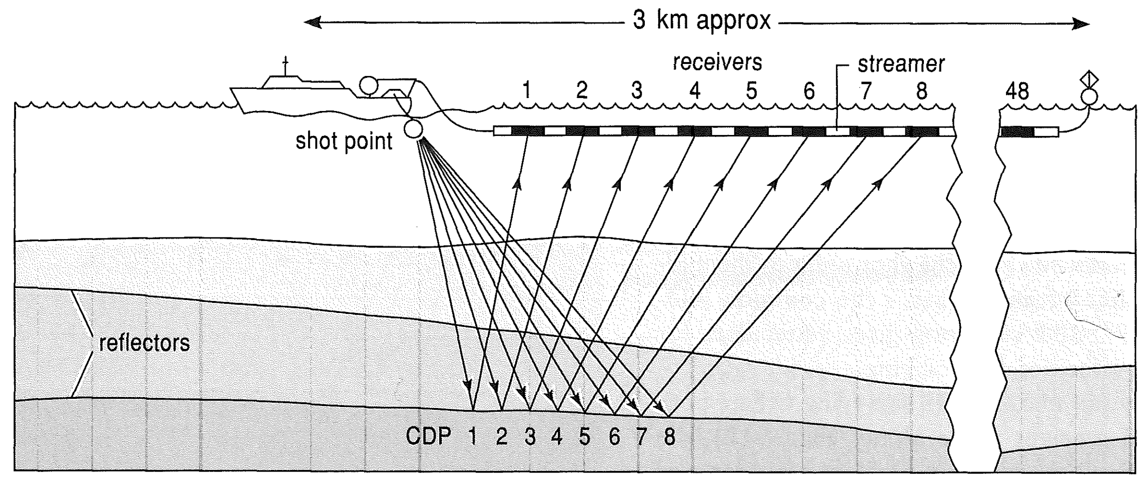


Figure 7.12 Seismic streamer for marine surveys.

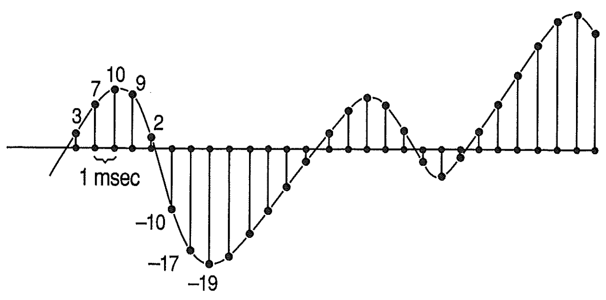


Figure 7.13 Digitisation of a seismogram trace.

Reflected rays are not the first arrivals, for they arrive after the direct ray (Section 6.2), and usually after the **ground roll,** the oscillations caused by surface waves (Section 5.7). Though surface waves travel more slowly than body waves, at small offsets from the source they arrive before reflected rays because of their shorter path, and then their large amplitude can make the reflected arrival difficult to recognise. Recording systems reduce the effects of direct rays and ground roll by using geophone arrays and filters but will not be described further.

7.7.2 Common-depth-point (CDP) stacking

Stacking was described in Section 7.3 as a way of adding together traces to deduce the velocities of layers from the moveouts and to improve the signal-to-noise ratio. The reflections used come from different parts of the interface (Fig. 7.14a); however, it is possible to stack using rays that have all reflected from the same part of an interface (Fig. 7.14b) by using pairs of shot points and receivers that are symmetrical about the reflector point, rays S_1R_3, S_2R_2, S_3R_1. Shots S_1, S_2, . . . are fired successively, so the traces have to be extracted from recordings of successive shots. The reflection point is called the **common depth point, CDP** (also known as the common reflection point, or – for horizontal reflectors with the shot point in the middle of the line of geophones – the common midpoint, CMP). CDP surveying gives better data for computing velocities and stacking.

To carry out marine CDP surveying, where both source and receiver move along (Fig. 7.12), the source is 'fired' whenever the system has advanced *half* the receiver spacing, so that reflections from a particular common depth point due to successive shots are received by successive receivers. Thus, for shot position S in Figure 7.14c, R_1 receives the reflection from CDP_3, R_2 from CDP_2, . . . ; then when the shot has reached S', R_2 has advanced to R'_2 and receives the reflection from CDP_3, while R'_1 receives one from CDP_4. The traces are sorted by computer into gathers with the same common depth point. When CDP surveying is carried out on land, using a line of regularly spaced *fixed* receivers, the shot point is advanced a distance equal to the receiver spacing.

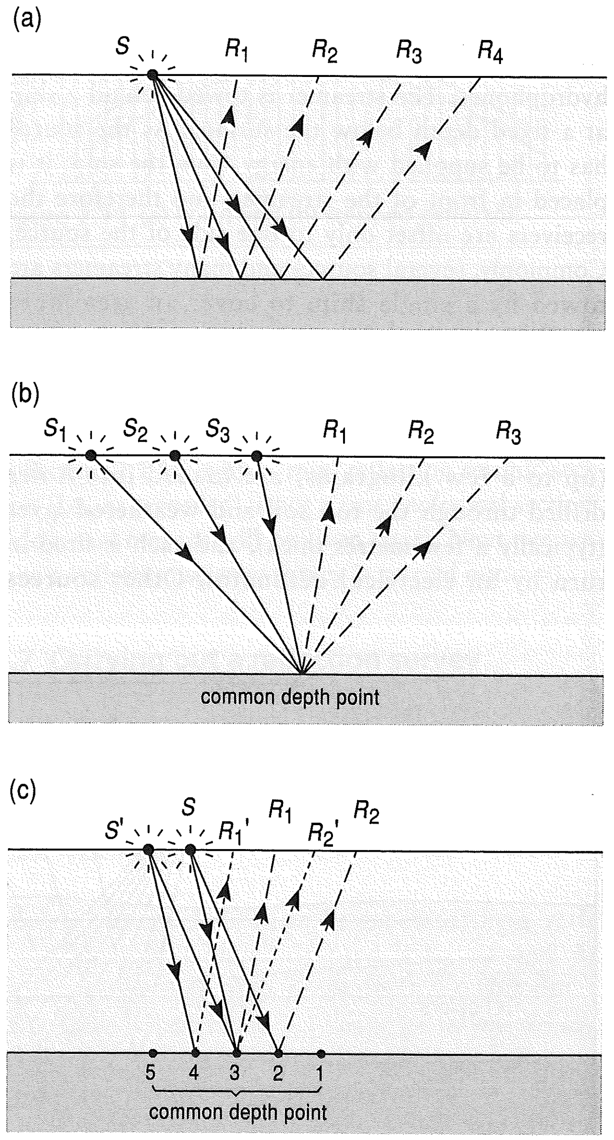


Figure 7.14 Common-depth-point stacking.

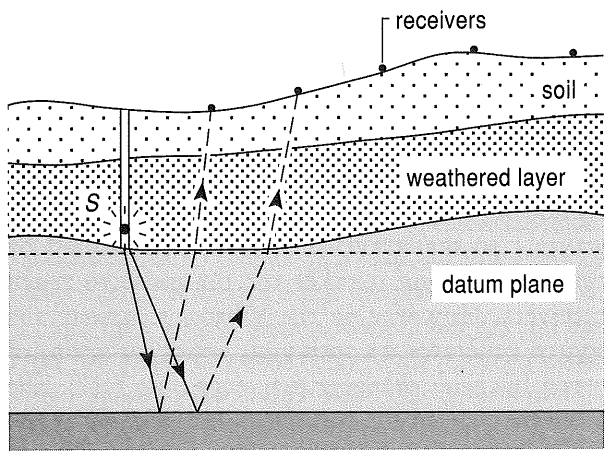


Figure 7.15 Static corrections.

The number of channels that are added together, or gathered, is termed the **fold** of the stacking; in Figure 7.14b it is threefold. The larger the number of receivers used the greater the fold, and 240-fold stacking is commonly used.

Velocity analysis, gathering, stacking, and migration are routinely applied to the 'raw' data. Additionally, in land surveys, any significant topography has to be corrected for, and allowance made for the effect of the topsoil and weathered layers; as pointed out in Section 6.7, though they are usually thin, the time spent in them is not negligible because of their slow seismic velocities. These corrections are called statics, and the **static correction** corrects or reduces the data as if it had been obtained with shot and receivers on some convenient horizontal datum plane beneath the weathered layer (Fig. 7.15). If the variation of static correction along a profile time is not properly allowed for, time in the surface layers may be interpreted as spuriously thick layers at depth where the velocities are much higher.

7.7.3 Data display

After the data have been processed, a single trace is produced for each common depth point. These are seldom shown as simple wiggle traces because other forms show arrivals more clearly. The most common modification is to shade the peaks but not troughs of the wiggles, to give a **'variable area display'** (Fig. 7.16). Reflectors then show up as dark bands, as already displayed in Figure 7.10. Sometimes this idea is carried further, by shading the peaks more heavily the taller they are or only showing peaks and troughs over some value, so omitting minor wiggles which it is hoped are not important.

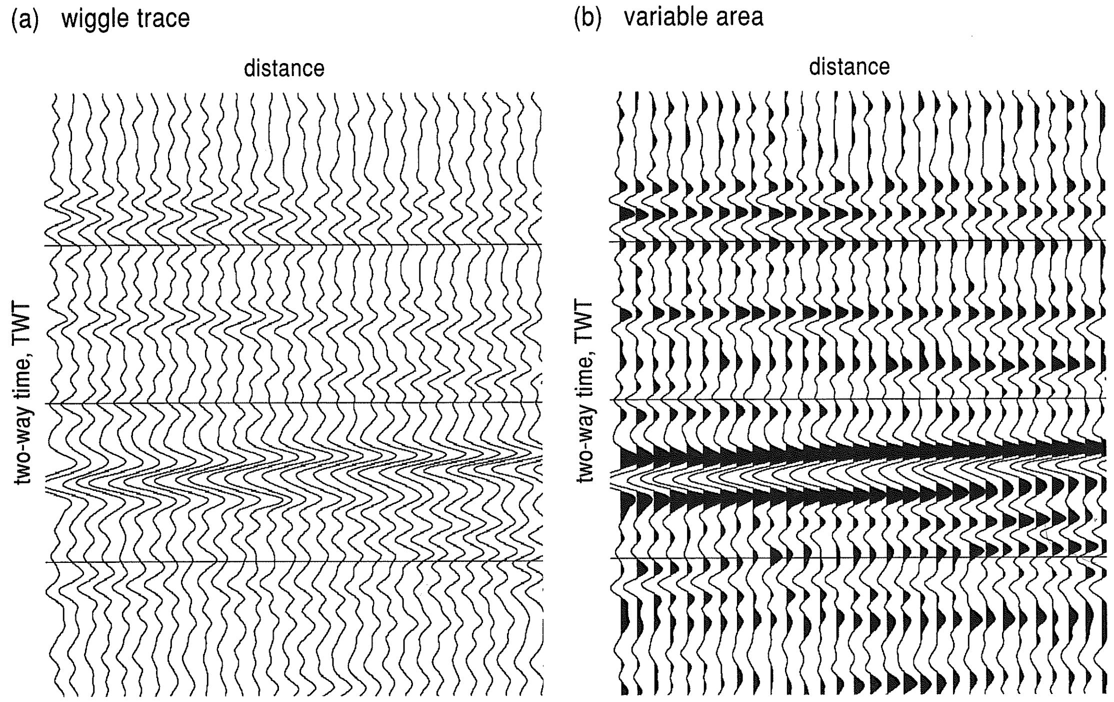


Figure 7.16 Wiggle and variable area displays.

Reflections from deeper reflectors become progressively weaker, for wave-front amplitudes are lessened by spreading and loss of energy due to reflections from shallower interfaces. To compensate, later parts of the traces are progressively amplified so that the average amplitude of the trace is kept constant; this is called **equalisation.** However, sometimes true amplitude is retained to emphasise the presence of a strong reflector, such as a gas–liquid interface. The most common display uses an equalised wiggle trace with variable area shading.

Colour is often used to add further information, such as the interval velocity, particularly in 3D displays (Section 7.9).

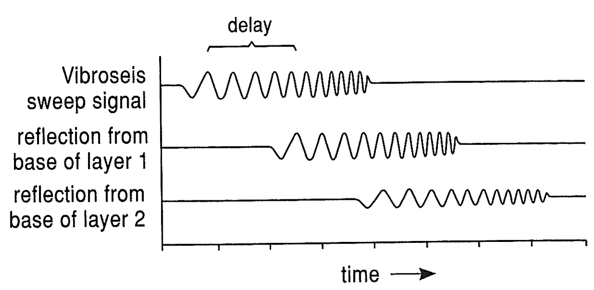


Figure 7.17 Vibroseis wave trains.

7.7.4 Vibroseis: A nonimpulsive source

All the types of seismic investigation described so far in this subpart have used impulsive sources, such as explosions, air guns, and hammer blows, and even the first arrivals from earthquakes. As explained in Section 4.1, a pulse is used – rather than a series of waves – so that travel-times can be measured by timing how long it takes for the pulse to reach receivers. However, in the **Vibroseis** system, the source generates a continuous series, or train, of waves but with *changing* frequency (Fig. 7.17). The receivers pick up the waveform, but delayed by the time taken to travel from source to receiver. To find the travel-time to any individual receiver, the recorded signal is shifted in time until all the peaks and troughs match those of the source. Because the shape of the wave is known, this can be done, even when there is much noise, by using data processing techniques and is an effective way of reducing noise.

The energy in each wave need not be large because many waves, rather than a single pulse, are used. This is an advantage when shaking of the ground has to be kept small, as in built-up areas. The waves are generated by a special vehicle with a plate between the front and rear wheels (Fig. 7.18) that is pressed rhythmically against the ground, reaching a momentary force of up to 15 tonnes.

Figure 7.18 A truck for Vibroseis surveying.

The frequency is varied, or swept, in the range 10 to 100 Hz, with the sweep lasting for up to 30 sec; the longer the signal, the more energy is put into the ground. The energy is often further increased by using several vehicles operating exactly in unison. Such sources have been used to produce reflections from the Moho, 30 km or more deep. The vehicle advances a short distance – often along a road, for this provides a convenient flat surface – and the process is repeated, to build up a seismic section. Geophones are laid out in the usual way. Once the delay of the recorded signal with respect to the source has been determined as above, processing follows as for pulse sources.

Vibroseis is the most widely used source on land, but has limited use in water, for it is difficult to generate a powerful enough signal to penetrate deeply. However, a system – called Chirp – has been used to investigate sediments to shallow depths, using frequencies in the range 1 to 12 kHz, which gives good resolution. It has also been used in marine archaeology to detect buried wrecks.

7.8 What is a reflector?

7.8.1 Strengths of reflected and transmitted pulses

Rays are reflected when they meet an interface, or seismic discontinuity, with an abrupt change of seismic velocity. But how abrupt need it be to produce a reflection, and what determines how strong the reflection is?

The ratio of the reflected and incident *amplitudes* (Fig. 7.19) $a_{\text{refl}}/a_{\text{incid}}$, is called the **reflection coefficient,** R; similarly, $a_{\text{trans}}/a_{\text{incid}}$ is the **transmission coefficient,** T. These coefficients depend on the densities, ρ, of the rocks, as well as on their velocities, v. They can be shown to be

$$R = \frac{\rho_2 v_2 - \rho_1 v_1}{\rho_2 v_2 + \rho_1 v_1}, \qquad T = \frac{2\rho_1 v_1}{\rho_2 v_2 + \rho_1 v_1} \qquad \text{Eq. 7.9}$$

The quantity ρv is called the **acoustic impedance** of the layer. Broadly, the bigger its contrast between the two sides the stronger the reflection. Equation 7.9 is true only if the ray is perpendicular to the interface. If it is oblique the formula is more complicated and gives different values for R and T; for instance, for angles greater than the critical angle

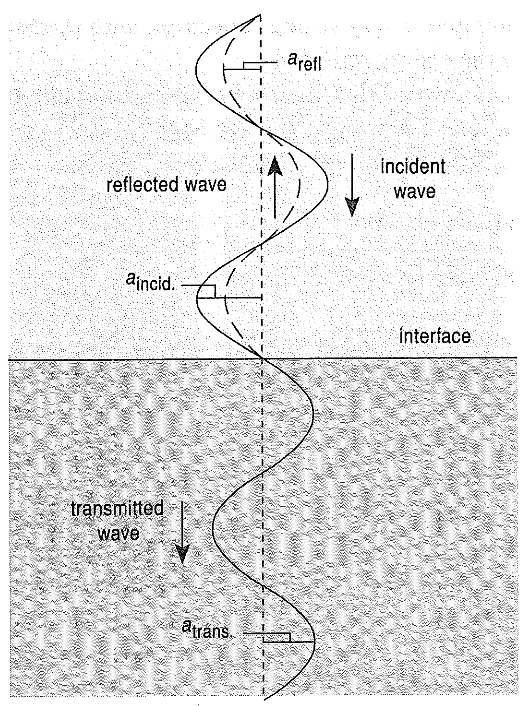


Figure 7.19 Reflection and transmission of a wave.

there is total internal reflection (Section 6.1), in which case R is 1. We shall consider only rays near to normal. As the velocities of S-waves differ from those of P-waves, they have their 'own' coefficients, and sometimes S-waves are reflected more strongly than P-waves. The ratios of *energies* reflected and transmitted are R^2 and T^2; $R^2 + T^2 = 1$.

As an example, we shall calculate the reflection coefficient of sandstone overlying limestone. These have the ranges of properties shown in Table 7.1.

Table 7.1 Seismic velocities and densities either side of an interface

Rock	Range of velocity, v_p	Range of density, ρ
Upper layer: sandstone	2 to 6 km/sec	2.05 to 2.55 Mg/m³
Lower layer: limestone	2 to 6 km/sec	2.60 to 2.80 Mg/m³

Suppose that the properties of the sandstone are at the bottom of their ranges, while those of the limestone are at the top. Then

$$R = \frac{(2.80 \times 6) - (2.05 \times 2)}{(2.80 \times 6) + (2.05 \times 2)} = 0.608$$

This would give a very strong reflection, with $0.608^2 = 0.37$ of the energy reflected.

Suppose instead that the rocks have these values: sandstone, $v = 3.3$ km/sec, $\rho = 2.4$ Mg/m^3, and limestone, $v = 3.0$ km/sec, $\rho = 2.64$ Mg/m^3. Then

$$R = \frac{(2.64 \times 3) - (2.40 \times 3.3)}{(2.64 \times 3) + (2.40 \times 3.3)} = 0$$

In this case, though there is a lithological boundary, there is no seismic reflector. Of course, acoustic impedances are unlikely to be identical, but they may be similar enough to produce only a weak reflection. (You may have realised that another choice of values could have given a negative reflection coefficient: This will be considered in Section 7.8.2.)

These calculations illustrate that the boundary between two lithologies need not be a detectable seismic interface, as was pointed out earlier. Conversely, a seismological interface need not be a geological boundary; an example is an oil–gas interface, discussed next.

Bright spot. The interface in a hydrocarbon reservoir between a gas layer and underlying oil or water produces a strong reflection, which appears on the section as a **'bright spot'**. It is also horizontal – a 'flat spot'. So the detection of a strong, horizontal reflector deep within sediments is strong evidence for the presence of gas (Fig. 7.20). Its detection is not always easy, though, because variability in the thickness or velocity of the layers above can decrease or increase the TWT to the interface, an effect called **pull-up** or **-down.**

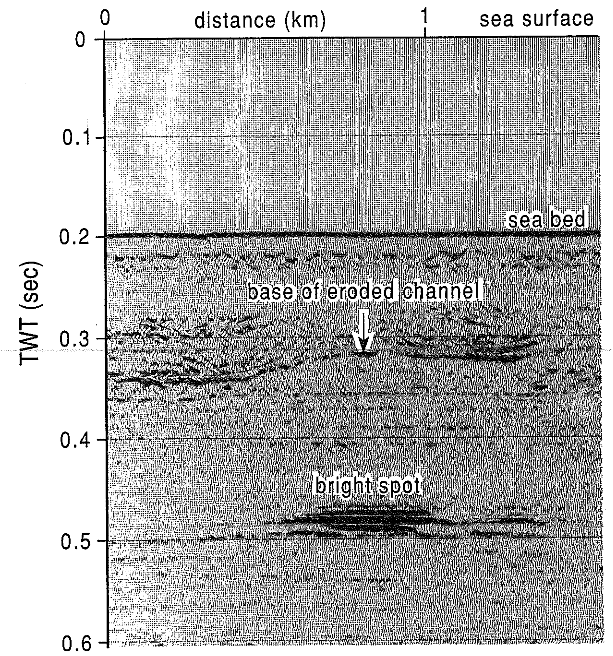


Figure 7.20 Bright spot.

7.8.2 Vertical resolution: The least separation at which interfaces can be distinguished

Suppose two interfaces could be brought progressively closer together, as shown by the series of diagrams in Figure 7.21a. The reflected pulses would overlap more and more, as shown by the 'combined' curves, until at some separation the pulses could not be distinguished, or resolved apart, and then the two interfaces would appear to be one. Figure 7.21a also shows that the shape of the combined pulse changes markedly with separation of the interfaces and may not look like the incident pulse.

Quite often the two interfaces are the two sides of a thin layer sandwiched within another lithology, such as a shale layer within sandstone. In this case *one* of the interfaces must have a positive reflection coefficient, R, and the other a negative one, because the values in Eq. 7.9 are exchanged. A negative value for R means that the reflected pulse is inverted compared to the incident wave, so one of the two reflected pulses is inverted compared to the other, as shown in Figure 7.21b (where the reflection from the lower interface is inverted). In this case, the closer the interfaces and the more the pulses overlap, the more completely they cancel, until there is no reflection at all! (This must be so, for bringing the interfaces together means shrinking the middle layer to nothing, resulting in a continuous medium, which obviously cannot produce reflections.) The shape of the combined reflection pulse depends on the actual shape of the pulse produced by the source, which is likely to be more complicated than the ones shown.

This addition or cancellation of waves or pulses (introduced in Section 6.1.1) is called **interference,** and it is termed constructive if the total is larger than either alone (Fig. 7.21a), destructive if the total is less (Fig. 7.21b).

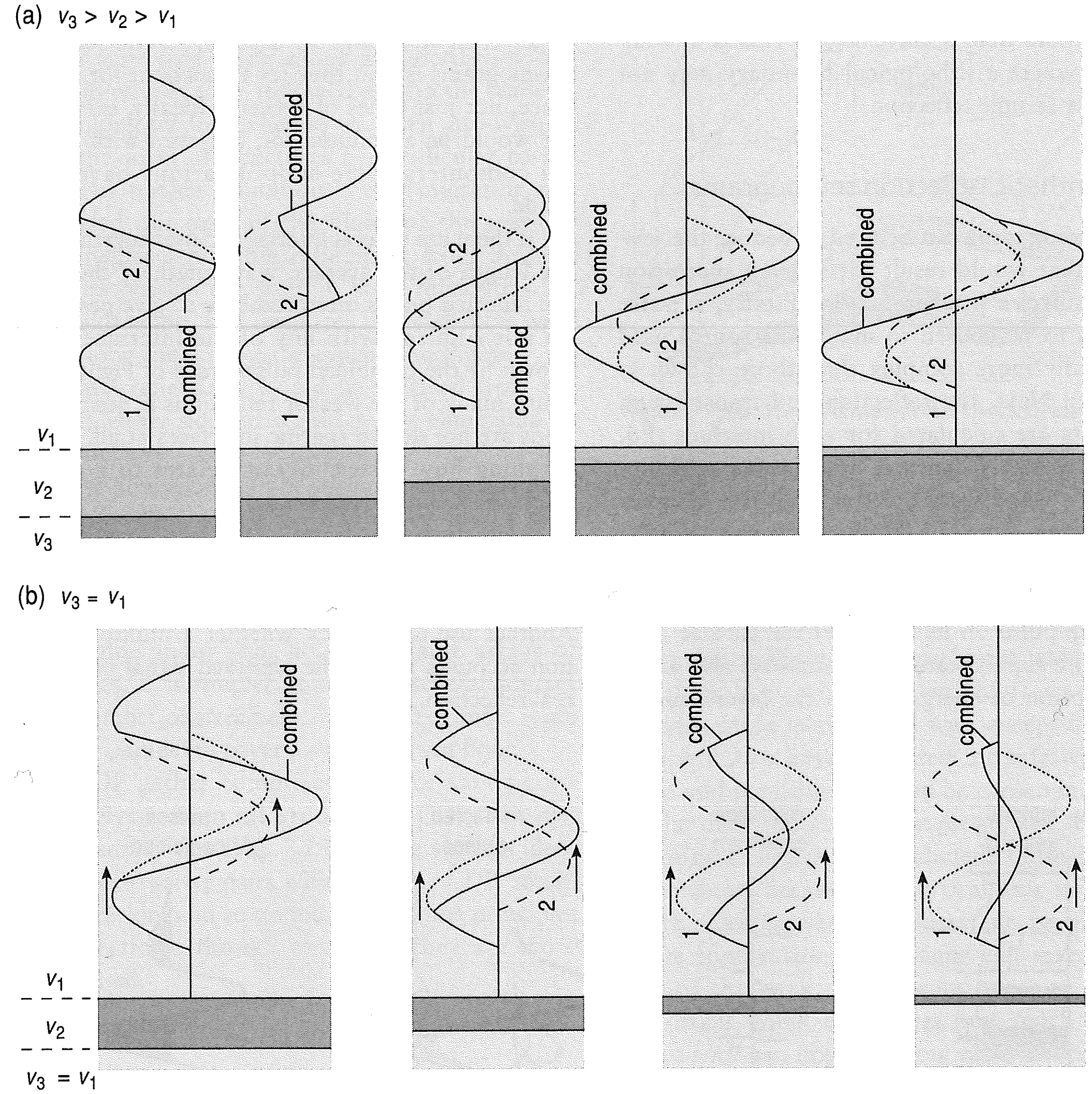


Figure 7.21 Reflections from close interfaces.

In practice, two pulses are difficult to distinguish when they are less than about a half of a wavelength apart. Since the pulse reflected from the lower interface has to travel further by twice the separation of the interfaces, interfaces cease to be resolvable when they are less than half a wavelength apart; that is, $\delta h < \lambda/4$.

Because **vertical resolution** depends upon the wavelength of the pulse, it can be improved by using a shorter pulse, which partly depends on the seismic source used; for instance, sparkers produce a shorter pulse than air guns. However, very short pulses are more rapidly attenuated by absorption (Section 4.5.5), so they may not reach deeper reflectors. Often, a compromise has to be made between resolution and depth of penetration. The same problem is met in ground-penetrating radar (GPR), and an example is given in Section 14.8.2.

Another situation where there may be no reflection is when an interface is a gradual change of velocities and densities extending over more than about half a wavelength, rather than being abrupt. We can think of the transition as made of many thin layers with reflections from each interface; the many reflected pulses overlap and cancel if they are spread out over more than a wavelength. This is another situation where a lithological boundary may not show up by seismic reflection.

7.8.3 Synthetic reflection seismograms

The preceding ideas can be used to deduce the seismogram that would result for a given succession and to improve interpretation. Firstly, a pulse shape has to be chosen to match the source used (it is usually more complex than those shown in this book). Next, the reflection and transmission coefficients are calculated for each interface (Eq. 7.9); the necessary densities and seismic velocities are usually found from borehole logs (see Chapter 18). Then the pulse is 'followed down the layers', its energy lessening by spreading and by reflection at each interface; this is repeated for each reflection of the pulse on its return to the surface (Fig. 7.22). More sophisticated computations also allow for absorption of energy within the layers, multiple reflections, and the effects of diffraction from edges. The advantage of using synthetic reflection seismograms is that they try to account for all of a trace, not just times of arrivals. Ideally, every wiggle would be accounted for, but even with powerful computers some simplifications have to be made.

If there are several interfaces closer together than the length of the average wavelength of the pulse, the reflected pulses often combine to give peaks that do not coincide with any of the interfaces, as is shown by the combined pulses trace of Figure 7.22. Thus, many of the weaker reflections in seismic sections are not due to specific interfaces at all. Understanding how a seismogram relates to a section improves its interpretation.

Synthetic seismograms – only feasible when there is good borehole control – are used to improve interpretation of seismic sections between boreholes. Another use is to check whether a multiple reflection accounts for all the observed signal or hides a real reflection.

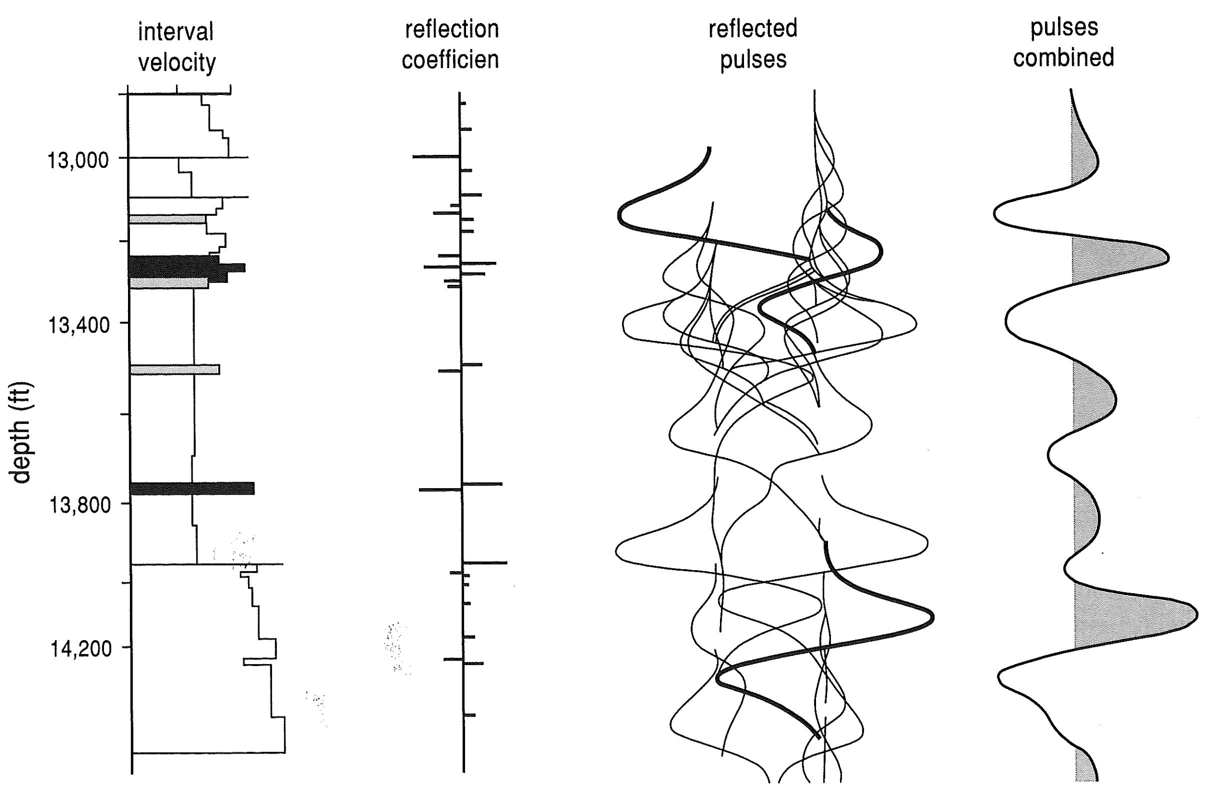


Figure 7.22 Production of a synthetic trace.

7.9 Three-dimensional (3D) surveying

Seismic surveying along a line as described so far allows interfaces to be followed only in the plane of the resulting section, whereas we often also wish to know the form of structures perpendicular to the section. A further limitation is that migration can be carried out only in the plane of the section, which neglects possible reflections from dipping reflectors outside the section. Such reflections are called **sideswipe.** For instance, if the section were taken along the *axis* of the syncline in Figure 7.9a, additional reflections would be received from the limbs on either side but would not be eliminated by migration.

These limitations can be surmounted by shooting a full **three-dimensional,** or **3D, survey,** ideally with receivers occupying the points of a regular grid on the surface (Fig. 7.23), as little as 50 m apart, with shots being fired at all grid points in turn. On land, this requires many shot holes and large arrays of geophones. It is simpler at sea, for several sources and a number of streamers – up to 20 or more – are towed in parallel to cover a swathe of sea floor.

A CDP gather comprises pairs of shot points and receivers from all around the CDP, not just along a survey line (Fig. 7.23), where suffixes identify source-receiver pairs with the same CDP. Stacking, migration, and so on follow the same principles as for sections, though obviously they are more complicated.

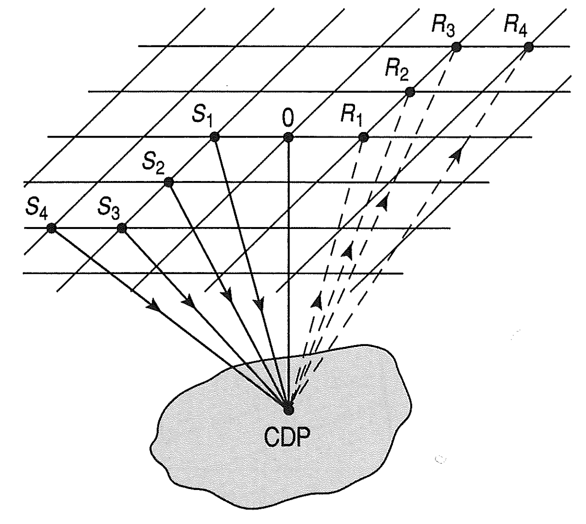


Figure 7.23 Part of a 3D survey grid and gather.

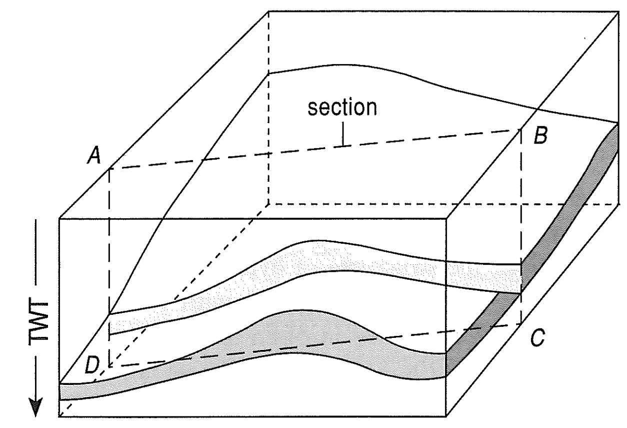


Figure 7.24 Three-dimensional block of data.

We can think of the reflectors revealed by this technique as being embedded in a block (Fig. 7.24). If there are many reflectors, this way of showing the results would be confusing; instead, the block can be 'sliced' with the aid of a computer, with colour normally used to emphasise features. Vertical slices (e.g., *ABCD* in Fig. 7.24) are similar to sections made in the usual way, except that migration will also have corrected for sideswipe reflections and their direction can be chosen independent of the direction the survey progressed. Horizontal sections can also be produced; these are called 'time slices' because the vertical axis is TWT, not depth. Alternatively, interfaces can be picked out. More sophisticated processing can reveal properties of the rock, such as its porosity, which is important, for it determines how much hydrocarbon a rock can hold and how easily it will release it (Chapters 18 and 22). Plate 6 shows a block of the subsurface from part of the Forties oil field in the North Sea, with the layers revealed by their porosity; in this case, higher porosity reflects a higher proportion of sand compared to shale. The low-porosity layer acts as a barrier to upward migration, so any oil will be just below it. Plate 7 shows a single layer from this part of the block. Its highest-porosity parts are due to a sandy channel present when the layer formed and forms a stratigraphic trap. This sort of information is essential when deciding where to drill extraction wells. A case study of the Forties field is described in Section 22.6.

A development of 3D surveying is **time-lapse modelling** (sometimes called 4D modelling). By repeating surveys at intervals, the extraction of hydrocarbons can be followed and unextracted pockets of oil detected. An example is given in Section 22.6.2 and Plate 8.

Though 3D surveying is much more expensive, both for data acquisition and reduction, than linear surveying, it can pay for itself in the increased understanding of the structures of hydrocarbon reservoirs that it gives and in the precision with which it permits holes to be drilled in often intricate structures. By the mid-1990s 3D surveying had become the norm for marine surveys, and it is increasingly being used on land, though this costs up to 50 times as much as the same area of marine surveying.

7.10 Reflection seismology and the search for hydrocarbons

By far the biggest application of reflection seismology is in the hydrocarbon industry, for which the method has been developed, though it has other applications.

7.10.1 The formation of hydrocarbon traps

The formation of hydrocarbon reservoirs requires the following sequence: (i) Organic matter – usually remains of minute plants and animals – is buried in a **source rock** that protects them from being destroyed by oxidation; this is often within clays in a sedimentary basin. (ii) The organic matter is changed by bacterial action operating at temperatures in the range 100 to 200°C – due to the burial (Section 17.3.2) – into droplets of oil or gas. (iii) The droplets are squeezed out of the source rock by its consolidation. Being lighter than the water, which is the dominant fluid in the sediments, they tend to move up, but also often sideways, perhaps due to deformation of the basin. (iv) The hydrocarbon is prevented from leaking to the surface, where it would be lost, by an impervious **cap rock**, such as shale. (v) To be extractable, the hydrocarbon has to be in a **reservoir rock** that is porous and permeable, usually a sandstone or a carbonate. (vi) To be commercially useful, it has to be concentrated into a small volume. This requires a **trap,** which can be of several kinds.

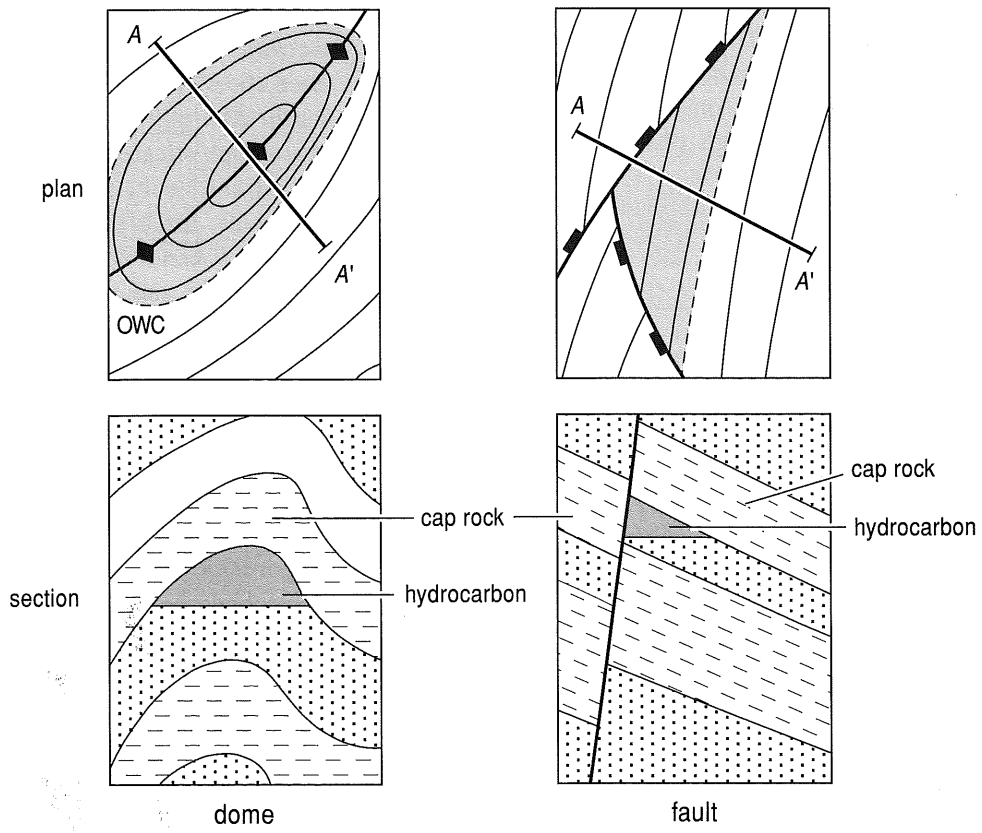


Figure 7.25 Two types of structural trap.

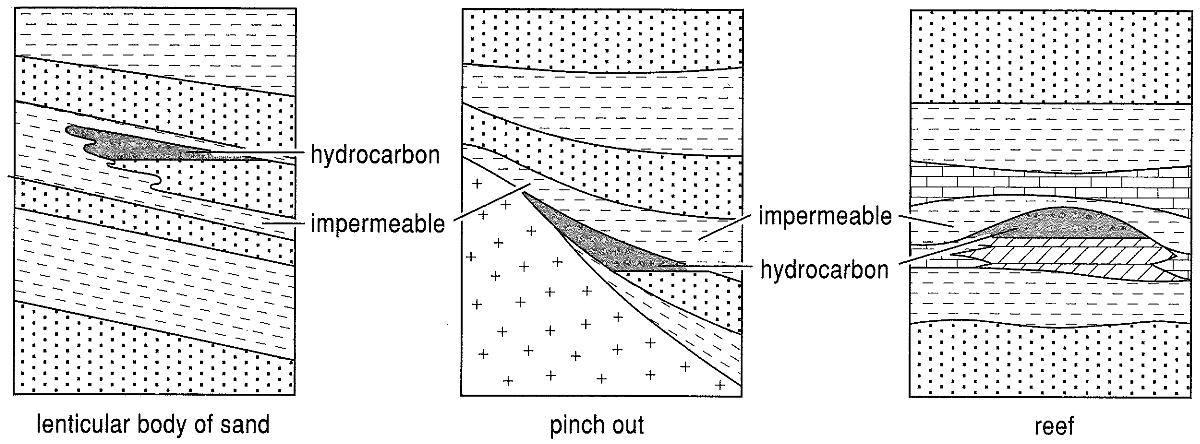


Figure 7.26 Three types of stratigraphic trap.

Structural traps result from tectonic processes, which produce folds, domes, faults, and so on. Figure 7.25 shows domal and fault traps, in plan and section. **Stratigraphic traps** are formed by lithological variation in the strata at the time of deposition, such as a lens of permeable and porous sandstone, or a carbonate reef, surrounded by impermeable rocks (Fig. 7.26). **Combination traps** have both structural and stratigraphic features, such as where low density salt is squeezed upwards to form a salt dome, both tilting strata and so causing hydrocarbons to concentrate upwards, and also blocking off their escape (Fig. 7.27). To find these traps seismically requires recognising how they will appear on a seismic section.

7.10.2 The recognition of hydrocarbon traps

Assuming that investigation indicates that traps filled with hydrocarbons may exist, how can they be located by seismic reflection? Easiest to recognise are structural ones (Fig. 7.25), where the tilting and/or termination of the reflectors reveals their presence. Stratigraphic traps which have tilted reservoir rocks terminating upwards in an unconformity are also fairly easy to spot, but other types are generally more difficult to detect, and other clues have to be pressed into service. The presence of a gas–oil or gas–water interface can give a particularly strong and horizontal reflector (bright and flat spot) (Fig. 7.20), as mentioned in Section 7.8.1, and also as the upper surface of a gas layer has a negative reflection

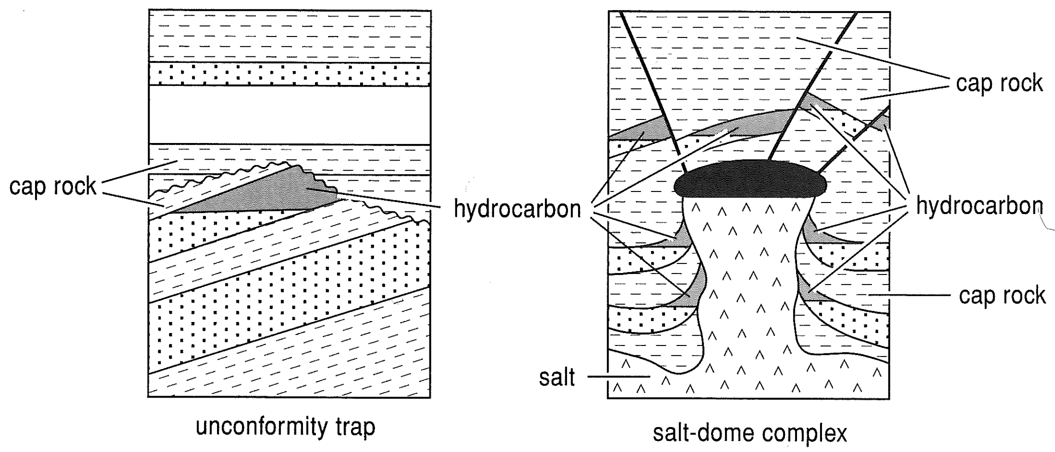


Figure 7.27 Two types of combination trap.

coefficient, it may be possible to recognise it from the inversion of the trace. Carbonate reefs also often produce a fairly strong reflector, which is often higher over the reef because reefs are often formed on a local topographic high and also generally are compacted less than other sediments following burial. The different velocities of reservoir and cap rocks may lead to a recognisable difference in interval velocity (Section 7.2) as well as a reflection. Other clues can come from more subtle changes in the character of the seismic section, which depends on a combination of the value of the reflection coefficient, the shape of the wiggles on the traces, and so on, and how these change laterally. Particular types of traps, such as sand lenses and river channels, can show up when contour diagrams are made, which is best done using 3D surveying. A model based upon 3D data is shown in Figure 22.14.

Because traps now being exploited are often smaller and harder to recognise than previously, seismic surveys must be of high-quality (closely spaced stations, high resolution sources, and so on), with static corrections and migration carried out precisely. Case histories of the finding and development of gas and oil fields are given in Chapter 22.

7.11 Sequence stratigraphy

Sequence stratigraphy (also called seismic sequence analysis and seismic facies analysis) grew out of seismic stratigraphy, which is the building up of a stratigraphy using seismic sections. As seismic data may cover large parts of a sedimentary basin, this provides a scale of investigation not possible from surface outcrops and a continuity not provided by boreholes unless they are very closely spaced. It can also allow recognition of global sea-level changes and so can provide a dating tool, valuable when palaeontological dating is not possible because of the absence of suitable fossils. It can also reveal stratigraphic hydrocarbon traps, described in the previous section. Sequence stratigraphy is possible because chronostratigraphic boundaries – surfaces formed within a negligible interval of time, such as between strata – are also often reflectors.

A stratigraphic sequence is a succession of strata with common genesis *bounded by unconformities,* or correlated unconformities. It is typically tens to hundreds of metres thick and spans an interval of hundreds of thousands to a few millions of years. How a sequence is built up and terminated depends on the interplay of deposition and changes in sea level. The formation of an extensive sedimentary basin requires subsidence to continues for millions of years, but superimposed on this subsidence are variations in sea level due to local changes of subsidence rate or relatively rapid global changes in sea level, called eustatic changes (discussed shortly). What concerns us here are *relative* sea-level changes while the sediments are being deposited.

Consider deposition along a coastline where sediments are being supplied by a river, during a period when relative sea level is changing. If the sea level remains constant (Fig. 7.28a) sediment is deposited from the river where the flow first slackens pace, at the shelf edge. Each bed forms over and beyond the previous one, to give a prograding succession. The top end of each bed thins towards the land – causing it to disappear from the seismic section, for the reason given in Section 7.8.2 – this is a **toplap;** at its lower edge each bed downlaps onto the existing rocks. If the sea level rises steadily (Fig. 7.28b), there is deposition on top of earlier beds as well as beyond, producing successive near-horizontal layers; where they butt up against the sloping coastline there is onlap. Toplap grades into onlap as the sea level changes from being constant to rising, altering the balance between prograding and aggrading of sediment, as shown in Figure 7.28c and d, where the arrow shows the direction of growth of sediments.

An unconformity – which defines the boundary of a sequence – is formed when the sea level falls fast enough for erosion rather than deposition to occur. If the rate of fall is moderate, the sea retreats across the shelf truncating the *tops* of beds. If it is rapid, sea level falls to partway down the slope before the tops can be removed, then erosion begins to cut laterally. The eroded material, of course, is deposited elsewhere, supplementing the sediments being supplied by the river. Deposition and erosion may alternate cyclically if relative sea level fluctuates. Over a cycle, deposition varies from fluviatile facies, through coastal and offshore to basin facies though these cannot all be distinguished in seismic sections without the aid of borehole data. Beds that end – by lapping of various kinds or truncation – are termed discordant; they appear in seismic sections as reflectors that cease laterally and are important for working out the history of deposition and erosion. Figure 7.29 shows how a geologic section of lithostratic units – the arrangement of strata in space – converts into a chronostratigraphy – the arrangement of strata in time. Strata are numbered successively.

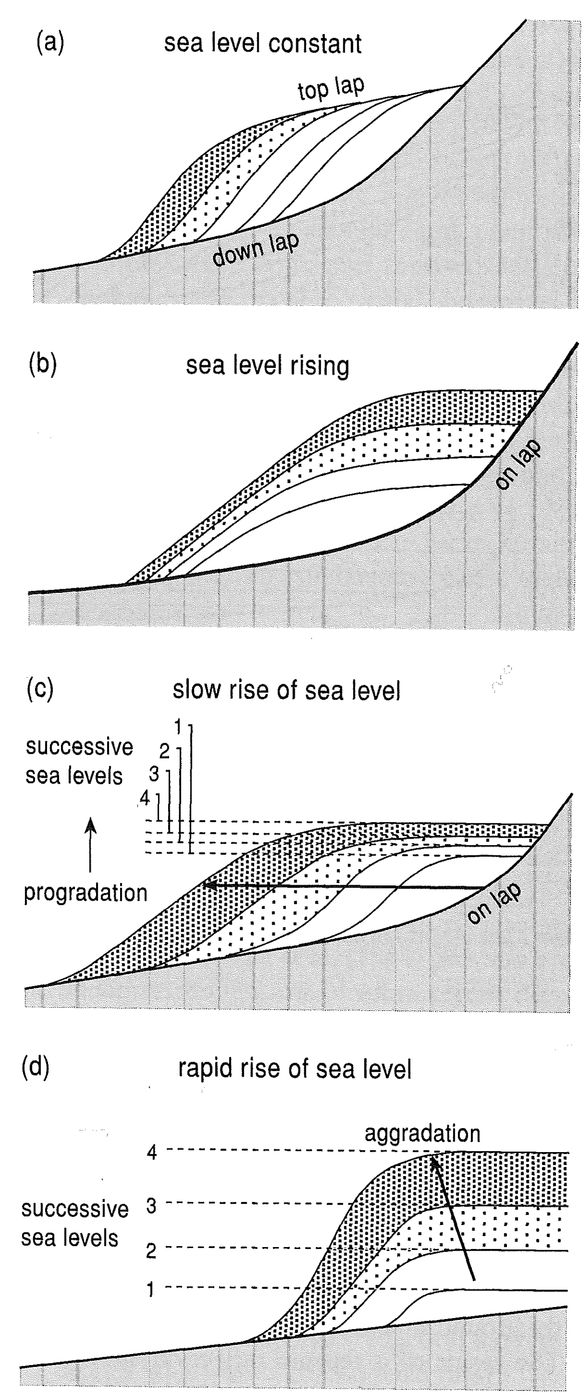


Figure 7.28 Progradation and aggradation.

Sequence stratigraphy and eustasy. Sequence stratigraphy provides a record of the changes in local relative sea level. By identifying the types and volumes of the various facies within a sequence the amounts of rise and fall can be estimated, and their times can often be deduced from fossils taken from boreholes. It was found that the records from widely separated basins were similar, and this was attributed to global sea-level changes, rather than, say, changes in local subsidence rates. Some of the falls were rapid – up to 1 cm a year – and continued for thousands of years, and though some of these rapid falls could be attributed to global glaciations, others occurred when there were no glaciations, casting doubt on the reality of global changes. However, subsequent investigations have tended to confirm the reality of the global sea-level changes, though the mechanism is not known.

7.12 Shallow-reflection seismic surveys

Reflection seismology needs to be modified for interfaces less than a few hundred metres deep. This is for several reasons, including proportionately larger ground-roll and the likely need for higher resolution (Section 7.8.2), for shallow layers are often thin. High resolution requires a short though not powerful pulse; this can be provided by a small explosion, such as by a 'buffalo gun', which fires a shotgun cartridge in a steel tube placed in a hole drilled to about a metre's depth. The geophones and recorders have to be able to respond to the higher frequencies generated, typically about 400 Hz, compared to less than 100 Hz in hydrocarbon exploration. Processing of the results is similar to that used for normal depth surveys.

Shallow-reflection seismic surveying has been applied to small-scale geological problems, such as investigating glacial structures below the surface (Harris et al, 1997) or imaging a buried river channel (Brabham and McDonald, 1992). It has also successfully mapped lignite deposits within clay sequences in Northern Ireland (Hill, 1992). This last could not be done using refraction surveys, because the lignite and clay have closely the same *velocity,* so layers are hidden (Section 6.6). However, as lignite

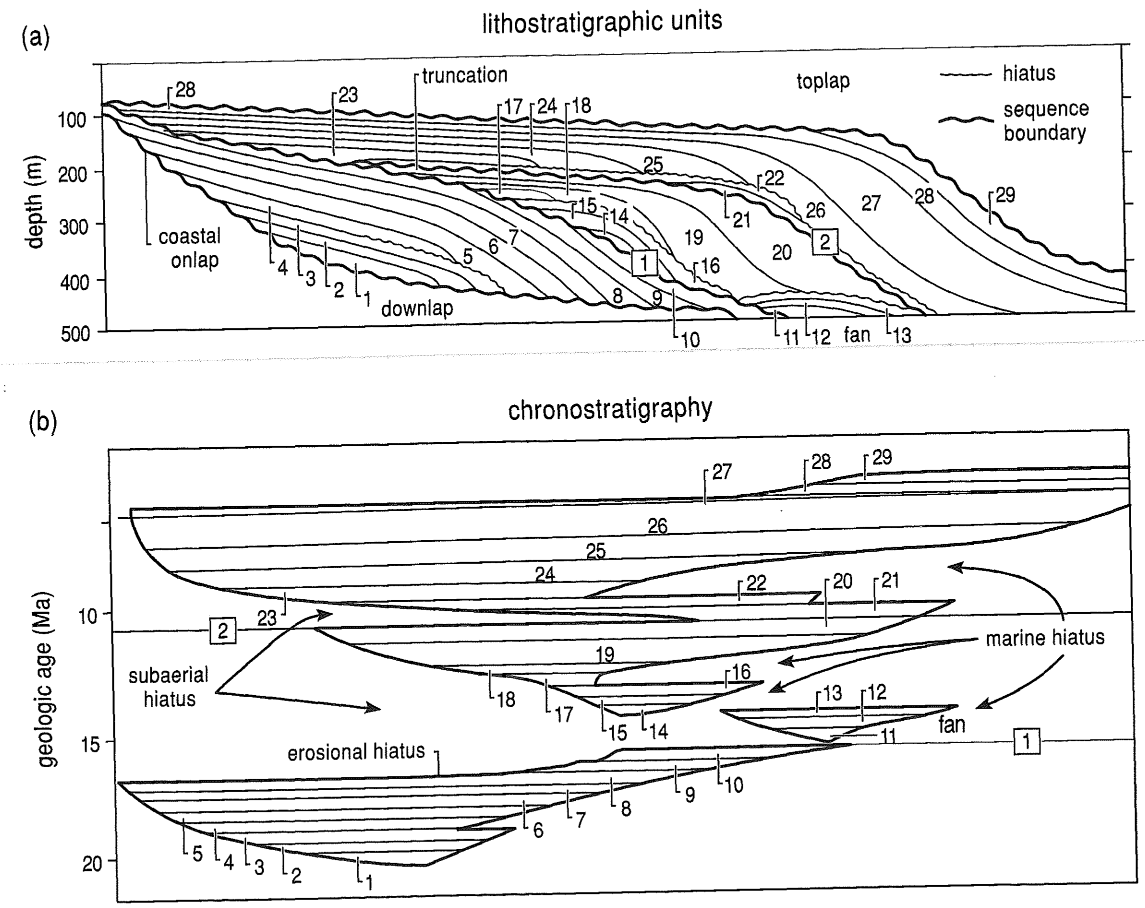


Figure 7.29 An idealised section in space and time.

has a considerably lower *density* than clay, the reflection coefficient (Section 7.8.1) is quite large, with a negative value for clay overlying lignite.

Summary

1. Reflection seismology is a form of echo or depth sounding for detecting interfaces below the ground surface, called reflectors.
2. The result of a seismic reflection survey – a reflection seismic section – is (with GPR, Section 14.8) the closest of any geophysical technique to a geological section. However, a 'raw' seismic section differs from a geological section for reasons that include these:
   (i) It gives the two-way time, TWT, rather than depth to the reflectors
   (ii) Reflections from a dipping reflector are displaced.
   (iii) Some apparent reflectors may be due to multiple reflections.
   (iv) Reflectors may not correspond to lithological boundaries, and vice versa.
3. Reflection surveying requires a source and one or more receivers. Impulsive sources for shipborne surveys are most commonly air guns or sparkers; on land usually explosives are used. Sources differ in energy and frequency (wavelength). Receivers for shipborne and land surveys are respectively hydrophones and geophones.
4. The Vibroseis source is nonimpulsive and 'sweeps' through a smoothly changing range of frequencies (or wavelengths). Its advantages are that (i) an intense source is not needed, (ii) it is quick to deploy, and (iii) it has a good signal-to-noise ratio. Therefore, it is much used in built-up areas, often along roads. It is mainly limited to use on land.

5. Velocities can be determined from the moveout. The velocity found is the rms velocity, v_{rms}, which is used to find interval velocities, using Dix's Formula, or Dix's Equation (Eq. 7.7).
6. Signals are stacked both to maximise the signal-to-noise ratio, and to find v_{rms}.
7. Migration is used to correct for the distortion in dip and depth resulting from dipping and curved reflectors, and to eliminate diffraction effects from sharply truncated reflectors.
8. The reflection and transmission coefficients of an interface depend on the difference of the acoustic impedances, ρv, on either side of the interface (Eq. 7.9). A negative reflection coefficient occurs when the acoustic impedance of the layer below an interface is less than that of the layer above it; it inverts the reflected pulse.
9. Interfaces are not resolved if they are less than about ¼ wavelength apart. Resolution is increased by using a source with a shorter pulse length (higher frequency), but at the expense of poorer depth of penetration. A transitional interface will not produce a reflection if the transition is significantly thicker than about ¼ wavelength. Many reflections on seismic sections are formed by interference of the reflections from several closely spaced interfaces.
10. Shallow-reflection surveys require high-frequency sources, receivers, and recorders to resolve shallow layers that are often thin. They find use in site investigation as well as for small-scale geological problems.
11. Synthetic seismograms, constructed using information on velocities and densities provided by borehole logs, can improve interpretation of sections away from the borehole.
12. Three-dimensional reflection surveys collect data from a grid. They permit migration perpendicular to a recording line as well as along it, eliminating sideswipe and allowing structures to followed in both directions. Results are presented as sections and time slices, and in other ways.
13. A major use of reflection seismic surveying is to find and exploit a variety of hydrocarbon traps.
14. Sequence stratigraphy allows extensive stratigraphies to be built up by recognising the various types of discordant endings of beds. The results are used to correlate successions within a sedimentary basin and to provide dates. On the largest scale they provide a record of eustatic sea level changes.
15. You should understand these terms: reflector, seismic section; two-way time (TWT); normal moveout (NMO), common depth point (CDP), channel, fold; rms and interval velocities; multiple reflections, 'bow tie', sideswipe, pull-up and -down, migration; variable area display, equalisation; acoustic impedance, reflection and transmission coefficients; vertical resolution, interference, diffraction hyperbola; three-dimensional surveying, time lapse; ground roll; static correction; bright spot; streamer; Vibroseis; source, reservoir and cap rocks; structural, stratigraphic, and combination traps, sequence stratigraphy.

Further reading

Reflection seismology is introduced in general books on applied geophysics, such as Kearey and Brooks (1991), Parasnis (1997), Reynolds (1997), and Robinson and Coruh (1988), and Doyle (1995) devotes several chapters to it; Telford et al. (1990) is more advanced. Among books entirely devoted to the subject, McQuillin et al. (1984) provides a good introduction though not of more recent developments, while Sheriff and Geldart (1995) is at a more advanced as well as more recent level.

Brown et al. (1992) contains a chapter on sequence stratigraphy, and Emery and Myers (1996) is devoted to the topic. Hill (1992) discusses a number of shallow-reflection case studies.

Problems

1. Why do marine seismic reflection surveys not record (a) S-waves? (b) refracted rays?
2. How does a migrated reflection seismic section differ from an unmigrated one? In what circumstances would they be the same?
3. How can a primary reflection be distinguished from a multiple one?
4. Will a migrated section correct for 'sideswipe'?
5. Is the dip of a reflector in an unmigrated seismic section more or less than its actual dip? Explain why with the aid of a sketch.

6. What is the main way in which the Vibroseis system differs from other data acquisition systems. Name two advantages that it has over other methods of land surveying.
7. What are the main purposes of stacking?
8. How can a reflection coefficient be negative? How can it be recognised?
9. How may synclines and anticlines appear in an unmigrated seismic section?
10. A succession consists of alternating sandstones and shales, with the top layer being sandstone. Calculate how the amplitude diminishes for reflections from each of the top four interfaces (ignore spreading of the wavefront and absorption), if the densities and velocities are as follows: sandstone 2 Mg/m^3, 2.6 km/s; shale 2.3 Mg/m^3, 2.8 km/s.
11. Seismic sections are not always what they appear. Explain how an apparent reflector may (a) have an incorrect slope, (b) may have an incorrect curvature, or (c) may not exist at all, while (d) three horizontal reflectors spaced equally one above the other may not be equally spaced, in reality.
12. What determines vertical resolution? Why does less than the required resolution sometimes have to be accepted?
13. Explain why a reflector on a seismic section need not correspond to a particular interface.
14. Why is a very strong horizontal reflection usually indicative of a gas–water interface? Why may a gas–water interface not always appear as a horizontal reflector?
15. A strong reflector that lies below several layers with different seismic velocities has the same TWT as the base of a single layer elsewhere. What do the total thicknesses above the two reflectors have in common?
16. Explain why a seismic interface may not be a lithological boundary, and vice versa. Give an example of each.
17. In what ways does shallow seismic land surveying differ from deep surveying?
18. A thin, horizontal layer of shale (v_P = 2.8 km/sec) lies within sandstone (v_P = 2.5 km/sec). What is the minimum thickness of shale that can be resolved in a Vibroseis survey? (Use an average frequency.)
19. Explain how (a) an interface may show up by seismic reflection but not seismic refraction, and (b) vice versa.

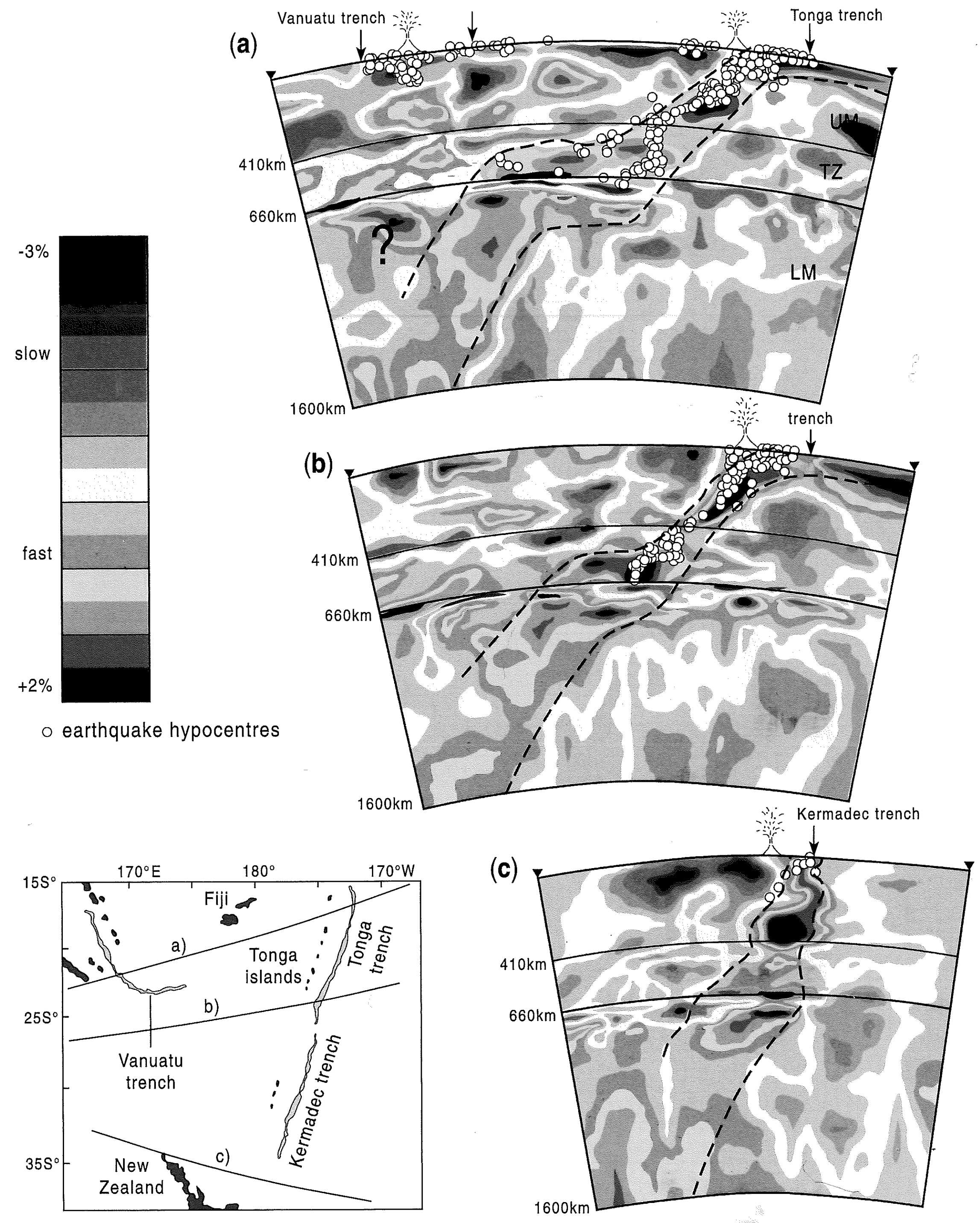


Plate 1. Three tomographic sections of the Tonga-Kermadec Trench. The subducting plate is revealed as a zone of higher velocity, believed to be because its temperature is lower than the surrounding mantle, and also – in the upper part of the plate – by earthquake epicentres shown by white dots. The plate is deflected by the 660-km discontinuity, the more so the faster the subduction rate. See Section 20.9.4 for details.