ONE • THE STRUCTURE OF EARTH AND ITS CONSTITUENTS

1-1. EARTH'S INTERIOR: RADIAL STRUCTURE, CHEMICAL COMPOSITION, AND PHASE TRANSFORMATIONS

Inferring Earth's composition is a prerequisite to understanding its evolution and dynamics as well as those of planets like it. One might think that the composition of Earth can be easily inferred from the rocks that we can see on Earth's surface. However, it immediately becomes obvious that these rocks cannot be the major constituent of Earth's interior because the densities of typical rocks on Earth's surface, such as granite or basalt, are so small, even if the effects of compression on density are taken into account. Therefore, materials in the deep Earth (and most other planets) are different from those on the surface. What materials are there, and how do we infer the composition of the deep interior of Earth (and other planets)? You may want to drill into Earth, but the deepest drilled hole in the world is in the Kola peninsula in Russia, which is only ~ 12 km deep (remember that Earth's radius is 6,370 km). Although some kinds of volcanoes bring materials from the deep mantle, this sampling is usually limited to ~ 200 km deep. Therefore, our inference of Earth's internal structure must be based largely on indirect information. In this connection, both geochemical and geophysical observations are particularly relevant. In *geochemistry,* the scientist measures the chemical composition of various materials, then uses the chemical rules that govern the distribution of various chemical elements to infer the chemical composition of Earth. In *geophysics,* one measures physical properties such as seismic wave velocities and density, and infers the composition and structure of Earth's interior based on the physical principles that control the variation of physical properties with thermodynamic conditions (pressure and temperature). In this chapter, I will first summarize these basic observations, and then explain several models of Earth that have been proposed on the basis of geo-

chemical and/or geophysical lines of inference. Although the inference of composition through such a process is indirect and not unique, some aspects of Earth's interior are now well understood. However, a number of problems remain controversial, including the nature of chemical heterogeneity in Earth's mantle.

1-1-1. Geochemical Models

An obvious starting point for inferring the chemical composition of Earth is the composition of rocks that we can collect on Earth's surface. We have a large data set of the composition of these rocks. Rocks from the shallow regions, the crust, are typically basalt or granite (rocks made mostly of quartz, pyroxenes, and plagioclase) or rocks that have been modified from these rocks by later chemical reactions (metamorphism). These rocks have a high silica content and small densities. Occasionally, we find denser rocks, peridotite or eclogite (rocks made mostly of olivine, pyroxenes, and spinels or garnets), which contain a smaller amount of silica and higher densities than basalt or granite. These rocks are obviously the candidates of mantle materials. However, petrological and geochemical studies show that the rocks that we can see on Earth's surface may not be representative of the bulk of Earth's mantle. It is important to note that Earth's shallow region has undergone extensive chemical differentiation through partial melting (melting of only some components) and the composition of each layer in the shallow regions is likely to reflect these differentiation processes. Consequently, the deeper portions of Earth's mantle likely have a composition different from that of rocks that we can collect on Earth's surface. One needs a theory for chemical differentiation in Earth to infer the chemical composition of Earth from the composition of rocks in near surface regions.

Ted Ringwood, at Australian National University (ANU) at Canberra, Australia, was a leader of the study of Earth's interior through a geochemical, petrological approach. He was perhaps the greatest Earth scientist from Australia. He made a number of fundamental contributions to knowledge of the structure and evolution of Earth and other terrestrial (Earth-like) planets. After obtaining a doctorate in geology at Melbourne in 1956, he studied with Francis Birch at Harvard. He returned to Australia in 1959 to take a position at the newly formed institution at ANU, and remained there until his premature death in 1993. Most of Ringwood's predictions or models of Earth's structures and evolution were based on simple principles, but he had an ingenious sense of synthesis, and

most of his models have survived the test of time. Ringwood proposed a model of Earth's chemical composition based on a theory of chemical differentiation in Earth (1975). The starting point of his model is the notion that the formation of basalt through partial melting of mantle materials is the most important volcanic activity on Earth. The mantle materials must then be able to produce basalt (more precisely, mid-ocean ridge basalts [MORB]) by partial melting. From this line of argument, Ringwood proposed a hypothetical rock called *pyrolite*, which produces mid-ocean ridge basalt by partial melting, and he proposed that the majority of the mantle must be composed of pyrolite. Pyrolite is rich in magnesium and iron, similar to rocks brought from the mantle, but its silica concentration is slightly higher than typical mantle samples and it has more calcium, aluminum, and other elements. In this model, typical mantle samples on the surface are interpreted as residue from the partial melting of pyrolite. Ringwood suggested that the majority of the mantle (including the upper and lower mantle) is composed of materials with a chemical composition similar to pyrolite. In this hypothesis, the depth variation of density and the elastic properties in the mantle must be explained as a result of phase transitions, compression, and thermal expansion of the same material.

Another model is that Earth has the same chemical composition as the average composition of the solar system. It is generally believed that the Sun and the planets in the solar system were formed as a result of the collapse of a putative primitive solar nebula. Therefore, the composition of the sun and the other materials of the solar system should be approximately the same. The composition of the outer layer of the sun can be inferred from the analysis of its optical properties (indeed, the element helium, He, was discovered by the analysis of the optical spectrum of the Sun, hence its name [*helios* means "the Sun" in Greek]). Another source of information on the composition of the solar system comes from the composition of meteorites. Meteorites are considered to be fragments of materials that failed to become planets. Among various types of meteorites, carbonaceous chondrite is a unique type, which is made of a mixture of various materials including metallic iron, silicates, and organic materials. The age of this type of meteorite, inferred from the composition of radiogenic isotopes, is the oldest (~ 4.56 billion years) among the ages of various objects in the solar system. Therefore, this type of meteorite is considered to be a remnant of the primitive solar system. The chemical composition of carbonaceous chondrite agrees well with that of the Sun except for the volatile elements. Consequently, the composition of the car-

TABLE 1.1
Chemical Composition of Earth (wt%) (after Ringwood, 1975)

	Continental crust	Upper mantle	Pyrolite model	Chondrite model (1)	Chondrite model (2)
MgO	4.4	36.6	38.1	26.3	38.1
Al_2O_3	15.8	4.6	4.6	2.7	3.9
SiO_2	59.1	45.4	45.1	29.8	43.2
CaO	6.4	3.7	3.1	2.6	3.9
FeO	6.6	8.1	7.9	6.4	9.3
other oxides	7.7	1.4	1.2		5.5
Fe				25.8	
Ni				1.7	
Si				3.5	

Note: In chondrite model (1), the light element in the core is assumed to be Si. Chondrite model (2) is a model of chemical composition of the mantle corresponding to the model of core shown in chondrite model (1).

bonaceous chondrite is considered to be representative of the composition of the solar system. Some scientists consider that Earth has a chemical composition that is similar to that of (carbonaceous) chondrite (except for volatile components). This model is often referred to as the *chondrite model.*

Table 1-1 shows chemical compositions corresponding to the pyrolite model and the chondrite model. One significant difference between these models is the ratio of (Mg + Fe)/Si. In the chondrite model, the amount of silicon in the mantle is greater than that in the pyrolite model. Based on various sources of information, on the other hand, the chemical composition of the upper mantle is estimated to be very similar to pyrolite or have slightly less silicon than pyrolite. Therefore, if Earth has the chondritic chemical composition, the amount of silicon in the deep mantle must be greater than in the shallower mantle. One possibility is a silicon-rich lower-mantle model; some people consider that the lower mantle consists mostly of $(Mg,Fe)SiO_3$. Thus, while the ratio of (Mg + Fe)/Si is con-

stant throughout the entire mantle in the pyrolite model, the chondrite model has a smaller $(Mg + Fe)/Si$ ratio in the lower mantle than in the upper mantle.

More complicated geochemical models have also been proposed. Don Anderson at Caltech (e.g., 1989), Eiji Ohtani at Tohoku University in Japan (1985), and Carl Agee, then at Harvard (1993), proposed mantle models containing several chemically different layers. Their thinking is based on a possible scenario for Earth's formation. Based on the results of Apollo missions, many Earth scientists believe that there was an extremely voluminous melting event due to high temperature caused by energy released via high-velocity collisions that occurred during the formation of the planets. This putative extensively molten region is called the *magma ocean*. They proposed that upon cooling of magma ocean, various minerals would have been solidified and would have sunk or floated to form chemical layering.

I must emphasize that these models have many uncertainties; their role is not to give some definitive picture of Earth structure, but to provide testable hypotheses. This point is particularly important in relation to *geochemical* models. Some rules of geochemistry are very well established and nearly independent of physical conditions. A good example is the rule of the radioactive decay of elements, which causes a temporal variation of isotope compositions. Also, the rule of partitioning some elements between liquids (magmas) and solids (minerals) is relatively well established. However, the physical processes that govern the distribution of chemically distinct materials in Earth are highly dependent on material properties that depend strongly on physical and chemical conditions, and these results should not be interpreted dogmatically to reach conclusions on chemical composition. For example, it is common to all models that the mantle basically consists of $(Mg,Fe)_2SiO_4$ (olivine and its high-pressure polymorphs) and $(Mg,Fe)SiO_3$ (pyroxene and its high-pressure polymorphs), and there is no doubt that $(Mg,Fe)O$ is richer in the mantle than in the crust. For more detailed issues such as the depth variation of the $(Mg + Fe)/Si$ ratio, however, we cannot draw conclusions only from these models. For example, Earth as a whole did not necessarily have the average chemical composition of the solar system. In the primitive solar system, its chemical composition was likely spatially variable, and it is unlikely that Earth and carbonaceous chondrites were formed from the same part of the primitive solar system. Furthermore, even if there was a magma ocean, it is not obvious that its solidification resulted in chemical layering. Brian Tonks and Jay Melosh (1990), at the University of Arizona at

Tempe, have argued that crystallization from the magma ocean could result in a nearly homogeneous chemical composition because of stirring and mixing by vigorous convection during solidification.

So far we have discussed only major element compositions. The distribution of *trace elements* (elements that occur only in a minor content) is also important in relation to material circulation in Earth's interior. Because trace elements exist only in a small amount by definition, their influence on physical properties such as density is often small. However, since their distribution varies greatly by partial melting (because most of trace elements have different sizes or different electric charges than the major elements do), the distribution of trace elements provides an important clue as to the processes of partial melting and the resultant chemical evolution of Earth. Important trace elements are *incompatible elements,* such as rare earth elements, which concentrate in the melt phase when rocks partially melt. If we examine the concentration of trace elements in igneous rocks such as basalt, we can find the chemical composition of their parental rocks. From these studies, we now know that there are several source regions in the mantle for igneous rocks. In some regions, rocks there do not contain very much of these incompatible elements. We infer that these regions have undergone extensive chemical differentiation by partial melting that has removed most of the incompatible elements. In other regions, in contrast, rocks contain large amounts of incompatible elements. These regions are inferred to have undergone a lesser degree of partial melting, or a large amount of materials enriched in these elements (sediments, crustal rocks, etc.) have been added in these regions. The differences in chemical composition lead to different *isotope ratios* because different elements have different radioactive decay schemes. In this way, we can infer the chemical *evolution* of different regions in Earth. It is now agreed that there are at least two chemically distinct regions in Earth that have been isolated from each other more than one billion years (Hofmann 1997). The distribution of these source regions is controversial and is the subject of active research (see chap. 4 for detail).

Hydrogen occupies a unique position among the trace elements. In most of the geochemical literature, hydrogen is not treated as a *trace* element, but its quantity in normal minerals is low, and hydrogen is dissolved in melts more than in minerals. Therefore, hydrogen behaves like an *incompatible element.* Like other trace elements, its influence on density is small, but its influence on some physical properties is large. In particular, hydrogen affects melting behavior (melting temperature and the composition of molten materials) and significantly reduces the resistance of ma-

terials against plastic flow. Through its effects on plastic properties, hydrogen can also change the nature of seismic wave propagation. I will discuss these issues in some detail in chapter 2.

1-1-2. Geophysical Models

Any model for the interior of Earth (or other planets) must be consistent not only with geochemical observations, but also with geophysical observations. Important geophysical observations include density and elastic properties. The average density of Earth can be easily calculated from geodetic observations of its size and the total mass. The average density of Earth is estimated to be 5,515 kg/m^3. The density of a planet provides an important constraint on its chemical composition. In fact, the average density is almost the only clue to estimate the composition of planets other than Earth, and it is actually sufficient to make a rough guess as to their composition. For example, the density of silicates (minerals consisting mostly of silicon, oxygen, and other elements, such as quartz and olivine) is 2,600–3,400 kg/m^3, and the density of iron (more precisely, iron-nickel alloy) is 7,800 kg/m^3. From these values, we can reach an important conclusion, that Earth consists mainly of silicates and iron. Similarly, we can conclude that the Moon (its density is 3,344 kg/m^3) consists mainly of silicates, and that Ganymede, one of Jupiter's satellites (its density is 1,936 kg/m^3), is composed of silicates and ice. Although we did not consider density variations due to compression and thermal expansion in Earth's deep interior, we can ignore them at this level of discussion. Density variations by pressure and temperature are at most 10–20%, and density differences due to chemical compositions are far more significant.

The moment of inertia is also an important parameter that can be determined by geodetic observation to constrain the internal structure of planets. If a mass M is located at a distance R from an axis of rotation, then the moment of inertia with respect to this rotational axis is MR^2. For a given mass, the moment of inertia is large if the mass is located far from the axis of rotation. Therefore, the moment of inertia depends on the mass distribution within a body, and is small if the mass is concentrated toward the center of a planet. The ratio $C \equiv$ (moment of inertia)/(total mass) $\times$ (radius)^2 is a nondimensional number that depends on how mass is concentrated toward the center of a planet. If the mass distribution is uniform, this ratio is 0.4, and if mass is completely concentrated at the center, it is 0.

The estimation of the *moment of inertia* is, however, not so straight-

forward. Consider a case in which mass (density) changes with depth but its distribution is spherically symmetrical. In this case, we can tell from Gauss's theorem that gravity outside a planet is the same as that in the case of all mass concentrated at the center. Therefore, the depth variation of mass cannot be determined by the measurement of gravity outside a planet. However, in a real world, mass distribution in a planet shows a slight deviation from spherical symmetry due to deformation by the effect of rotation (centrifugal force) and tidal force. Because of this, the moment of inertia can be determined using solely the observations outside a planet. If mass distribution differs slightly from spherical symmetry, objects rotating around a planet (including artificial satellites) are affected not only by the central force but also by the torque. As a result, the orbit of a satellite is not fixed at a perfect elliptical orbit, but the orbital plane moves slowly around the equator of a planet. By measuring this movement of the orbital plane, we can estimate how mass distribution in a planet deviates from the spherical symmetry. At the same time, this torque affects the motion of the planet itself and causes *precession*. Precession is significant when the cause of torque (the mass of other planets) is large. Thus, while the effect of artificial satellites on the precession of a planet can be ignored, the existence of large celestial bodies close to a planet in consideration has a significant influence on precession. The period of precession depends on the magnitude of torque applied to a planet and its moment of inertia. Once the magnitude of torque is estimated from the orbit of the satellite, therefore, the moment of inertia can be calculated by the period of precession. In the case of Earth, it has been found that $C = 0.3308$. From this, we can infer that the mass of Earth is concentrated toward the center; namely, the existence of a heavy core is implied. Precession has not been observed for most of the other planets. In these cases, by measuring deformation due to centrifugal and tidal forces, we can calculate the moment of inertia. Table 1-2 summarizes basic geodetic data on Earth.

In the case of Earth, earthquakes occur frequently, and this allows us to obtain detailed information about the elastic properties and the density of its interior. Earthquakes, which are nothing but a disaster for our daily lives, illuminate the dark interior of Earth for researchers. In fact, almost all of the chapters of this book deal with seismological observations, which give us the most detailed information on Earth's interior. Figure 1-1 shows how seismic waves propagate through Earth. Body waves (seismic waves that propagate through the bulk of Earth), which have been studied since the earliest time of seismology, propagate three-dimensionally from a hypocenter (the location where an earthquake oc-

TABLE 1-2
Basic geodetic data for Earth

equatorial radius	6,378 km
polar radius	6,357 km
flattening	1/298.26
total volume	1.083×10^{21} m³
total mass, M	5.9737×10^{24} kg
average density	5,515 kg/m³
moment of inertia around the spin axis, I	8.036×10^{37} kg m²
$C = I/MR^2$ (R: average radius)	0.3308

curs). Most of Earth's interior is solid, and there are two types of seismic waves (elastic waves): compressional and shear waves. The velocities of these two waves are determined by the elastic properties and density of a material, and they satisfy the following relations:

$$V_p = \sqrt{(K + (\tfrac{4}{3})\mu)/\rho}, \; V_s = \sqrt{\mu/\rho}, \tag{1-1}$$

where V_p and V_s are the velocities of compressional and shear waves (p stands for *primary* and s stands for *secondary*), K is bulk modulus or incompressibility, μ is shear modulus or rigidity, and ρ is density. By knowing the time and place of an earthquake (which can be inferred from the travel times at various stations), the seismic wave velocities of a region through which seismic waves propagate can be determined from the travel times that seismic waves take to arrive at observational points (seismological stations). Seismic waves also reflect and refract at various boundaries. From reflection coefficients, we can obtain the information not only on seismic velocities but also on densities.

Seismic waves that reach to a boundary undergo reflections and refractions to cause a certain type of wave that propagates along the surface. These are called *surface waves*. Analysis of surface wave propagation is more sophisticated than that of body waves. Progress in computer technology and in theoretical treatment has made it possible to analyze in
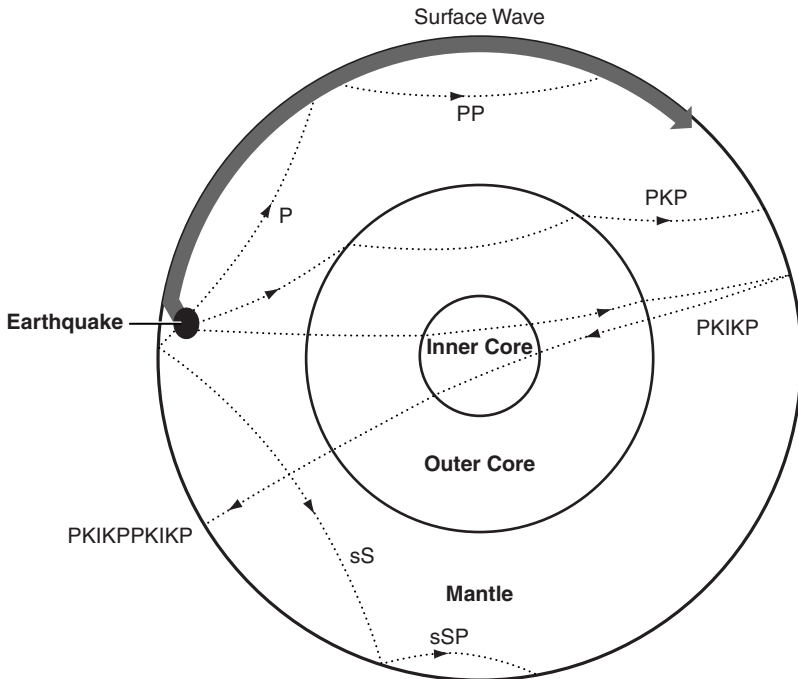
Fig. 1-1. The propagation of seismic waves. The broken lines show the propagation paths of body waves. Body waves are named (PP, SKS, etc.) based on their paths. The thick solid line denotes the propagation path of the surface wave.

detail the seismic wave propagation through complex structures. The surface wave study of Earth's internal structure is especially important for the study of the upper mantle: there is a low seismic velocity zone in the upper mantle (chapter 2), and body waves avoid traveling through the low-velocity zone. The velocity of surface wave varies with wavelength. This is called the dispersion relation. Dispersion occurs because elastic properties vary with depth and because waves with a longer wavelength are sensitive to elastic properties at greater depths. Thus, from the dispersion of surface waves, the structure of Earth can be inferred. One of the advantages of surface wave study is that it provides information on low-velocity zones. Since surface waves propagate two-dimensionally, they can propagate for long distances without much attenuation. Therefore, when a big earthquake occurs, surface waves traveling around Earth many times give us a large amount of data regarding the relatively shallow part of Earth.

Traveling surface waves can cause a shaking of Earth as a whole. This phenomenon is referred to as *free oscillation*. Earth rings just like a big bell; because its *tone* depends on the elastic properties and density of Earth, as happens in the case of a regular bell, we can investigate the interior of Earth by analyzing the *tone* of Earth. The principle is similar to that used in the study of surface waves. Free oscillations have various *modes* (like harmonic overtones for musical instruments), and each mode is sensitive to elastic properties and density at a different depth. The structure of Earth can be determined by comparing observations and models for various mode frequencies. The first observation of Earth's free oscillation was made during the Chilean earthquake in 1960, and free oscillations have been used for the study of Earth's internal structure ever since. One of the advantages of studies using surface waves and free oscillations is that it is not necessary to know the time and location of the earthquake. In body wave studies, the uncertainty of hypocenters (locations of earthquakes) can result in large errors.

Since the study of surface waves and free oscillations uses lower-frequency waves than those of body waves, it is sometimes called *low-frequency seismology*. Although this method can accurately determine large-scale structures, it cannot determine small-scale structures because low-frequency waves have long wavelengths, and hence their propagation is insensitive to small-scale features. It is therefore important to use a range of data, including body waves, surface waves, and free oscillations, to constrain the structure of Earth.

By integrating various kinds of seismic data, Adam Dziewonski, at Harvard, and Don Anderson proposed a standard model for the internal structure of Earth. In this model, a range of seismological observations are included in the analysis, and the corrections for the effects of using different frequencies are made based on the physical model of elastic and non-elastic deformation (box 1-1). This model is called PREM (Preliminary Reference Earth Model) (fig. 1-2) (Dziewonski and Anderson 1981). For their fundamental contributions to the study of the structure of Earth's interior through seismology they received the Crafoord Prize, equivalent to the Nobel Prize, in 1998 from the Swedish Academy. By comparing the densities and elastic constants of various materials, the likely composition of Earth can be inferred from the distribution of density and elastic constants within Earth as they are given by PREM. For a rough estimate of its chemical composition, we can ignore the effects of pressure and temperature on density. The previous conclusion that Earth is composed of silicates and iron can be derived from this kind of ap-

Box 1-1. Anelasticity and Physical Dispersion

One of the important points of PREM is the incorporation of a range of seismological observations in a physically consistent fashion. To a very high degree, the propagation of seismic waves can be understood based on the theory of elastic waves. However, when the details of wave propagation are investigated, deviations from perfect elasticity can be noted. This is due to the fact that seismic waves are low-frequency waves (~ 0.001 to ~ 1 Hz) that propagate through rather hot materials: the temperatures in most portions of Earth exceed half of melting temperature. Anelasticity causes dissipation of energy as heat; consequently, the amplitude of elastic waves decreases with time (and hence distance), leading to seismic wave attenuation. When a material responds to an external force with some energy dissipation, then elastic constants of that material become dependent on frequency. At infinite frequency, there is no time for a viscous element to respond, and there is no effect of anelasticity: elastic constants are the same as the case without any anelasticity. In contrast, at lower frequencies, there is more time for viscous components to affect the response of a material; hence, the elastic constants (seismic wave velocities) decrease with the decrease of frequency. This is called physical dispersion (an example of what is known the *Kramers-Kronig relation* in physics). Its importance was first pointed out by Gueguen and Mercier, in France (1973), and first demonstrated in seismology by Hiroo Kanamori and Don Anderson (1977). Anelasticity therefore provides a link between seismological observations and rheological properties.

proach. When we want to consider elastic properties in addition to density, we cannot ignore the effects of temperature, and especially of pressure, anymore. Elastic constants vary by several times for the range of pressure expected in Earth's interior. I will discuss the behavior of materials under high pressure in the later part (sec. 1-1-3) of this chapter.

Based largely on the studies on seismology and high-pressure mineral physics, our understanding of Earth's structure has made significant progress in the last few decades. First, I will explain the first-order approximation of Earth's structure, then summarize our current understanding of temperature distribution within Earth, which is closely related to geodynamics. Recent progress in the study of the fine structure of Earth
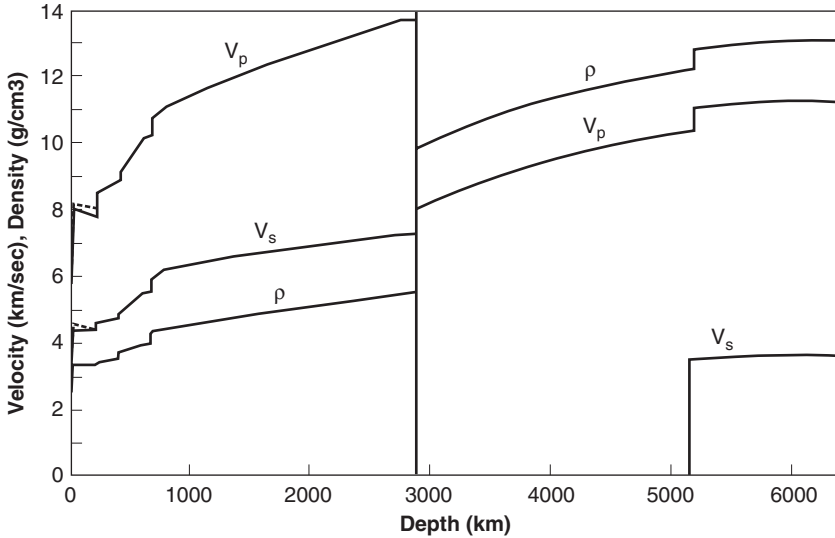
Fig. 1-2. A seismological model of Earth's interior (PREM) (after Dziewonski and Anderson 1981). Vp is the compressional wave velocity; Vs, shear wave velocity; and ρ, density.

(by seismic tomography) has revolutionized the course of research on the dynamics of Earth's deep interior. The progress in geodynamics in connection to seismic tomography will be discussed in chapters 3 and 4.

To construct a first-order model for Earth, we need to know the distributions of pressure and temperature within Earth. The effect of pressure is especially important when estimating chemical composition from density and elastic constants. The effect of temperature is smaller than that of pressure. To a very good approximation, pressure within Earth is in hydrostatic equilibrium. The reason for this approximate hydrostatic balance is that, as will be described later, the viscosity of Earth materials is so small that Earth materials cannot support a large nonhydrostatic stress. The estimated distribution of pressure is shown in figure 1-3. Pressure at the center of Earth is approximately 360 GPa. Although pressure is determined by the basic principle of hydrostatic equilibrium, the distribution of temperature depends on several uncertain factors. Therefore, the temperature distribution, shown in figure 1-3, has large uncertainties. While these uncertainties do not affect the estimation of chemical compositions very much, temperature distribution is important for the dynamics of Earth's interior. I will discuss it in relation to dynamics later in this chapter.
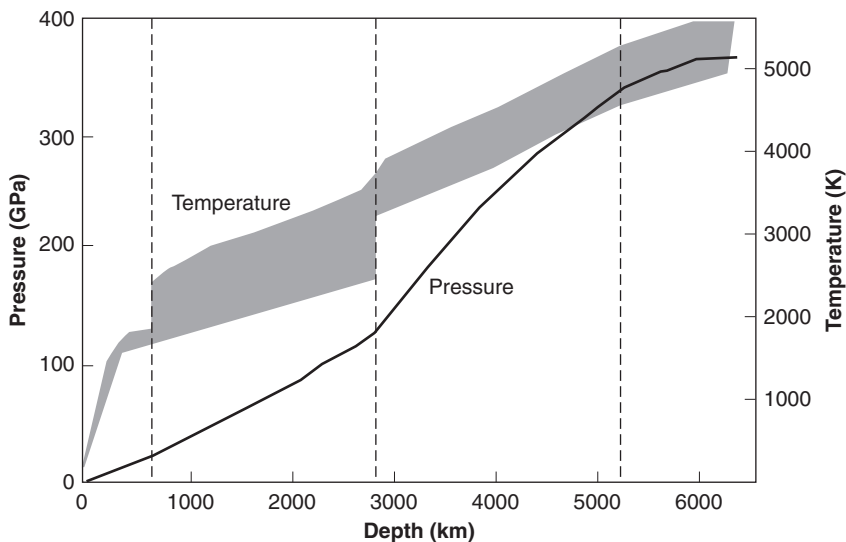
Fig. 1-3. Temperature and pressure distribution in Earth's interior. Temperature distribution greatly depends on the pattern of convection. This figure shows the currently accepted range of temperature distribution.

1-1-3. Earth Models

A. The Layered Structure of Earth: The Zeroth-Order Approximation

There are three chemically distinct layers in Earth that have distinct elastic properties and densities. The crust is a thin and very heterogeneous layer near the surface, which is composed of light silicate minerals such as quartz and feldspar. Crustal thickness varies from place to place; oceanic crust is homogeneous and about 7 km thick, and continental crust is about 30–70 km thick. Beneath the crust, there is a layer composed of denser silicate minerals such as olivine. This region is called the mantle. The mantle continues to the depth of about 2,900 km. Since both compressional and shear waves can propagate through both the crust and the mantle, they must be solid to a large extent. Even if they are molten, the degree of melting should be small in order for shear waves to propagate. At the depth of about 2,900 km, there is a boundary with the largest density jump in Earth. A dense layer beneath this boundary is called the core. Based on its density (and elastic properties), the core is considered to be made mostly of iron.

B. The Layered Structure of Earth: The First-Order Approximation−Phase Transitions

The simplest (zeroth-order) layered structure of Earth described above is due to differences in chemical composition. The chemical composition of each layer is relatively uniform. When we examine each layer closely, however, we can find a variety of layered structures within each layer. The most prominent is the core; while only compressional waves can propagate through the shallower part of the core (~ 2,900 to ~ 5,150 km depth), both compressional and shear waves can propagate through the deeper part (~ 5,150 to ~ 6,370 km). From this, we can conclude that the shallower part of the core (the outer core) is made mostly of molten iron, whereas the deeper part (the inner core) is made mostly of solid iron. The melting point of iron increases toward the center of Earth because of the increase in pressure, and this probably results in the formation of the solid inner core. This is a typical example of layering due to a phase transformation. As described later, the layered structure of the inner and outer core is thought to play an important role in the generation of geomagnetic fields.

Although not as prominent as above example, the mantle also has important layering. In the depth range of ~ 410−660 km, both density and elastic wave velocities increase much faster than in other layers. Francis Birch, at Harvard University, was a father of mineral physics who received his doctorate in physics at Harvard in 1932 under the supervision of Nobel laureate Percy Bridgman, established the first high-pressure mineral physics laboratory at Harvard, in the early 1930s. Birch (1952) showed that it is impossible to explain such a rapid increase in density by the compression of the same material with the same structure alone. He derived this conclusion by comparing the density distribution of Earth estimated from seismic observations with the density distribution at some standard condition. We consider density variation due to the vertical motion of a material with *adiabatic* compression (or expansion) as a standard condition. The adiabatic compression (or expansion) refers to a process in which a material is compressed (or expanded) without an exchange of heat with the surrounding materials. This would occur when a piece of rock is moved vertically in Earth rapidly enough. Such a rapid vertical movement is considered to occur in Earth by vigorous *convection*. The change in density due to adiabatic compression (or expansion) can be inferred from seismic wave velocities. The actual variation of density with depth can also be inferred from seismic wave velocities with the help

of other constraints, including the moment of inertia. Therefore, it is possible to compare the depth variation of actual density in Earth with the depth variation in density corresponding to hypothetical adiabatic compression (or expansion). This ratio is called the *Bullen parameter* after a New Zealander seismologist, Keith Bullen:

$$[\text{Bullen parameter}] \equiv \left(\frac{d\rho}{dz}\right) / \left(\frac{d\rho}{dz}\right)_{ad}, \qquad (1\text{-}2)$$

where $d\rho/dz$ is the density gradient (z is the depth) in a real Earth model and $(d\rho/dz)_{ad}$ is the adiabatic density gradient calculated from seismic wave velocities and gravity (box 1-2). The Bullen parameter is 1 if the density variation in Earth occurs solely due to adiabatic compression (or expansion). A value less than unity indicates that density does not increase much with depth, which may be caused by a large temperature gradient. On the other hand, a value greater than unity corresponds to a density increase larger than expected from the standard condition, which may due to phase transformations. Figure 1-4 shows the distribution of the Bullen parameter. At the shallower part of the upper mantle, it is smaller than unity, suggesting a large temperature gradient (see sec. 1-2). At the deeper part (410–660 km), the parameter significantly exceeds unity. On the basis of this fact, Birth inferred that the mantle minerals undergo phase
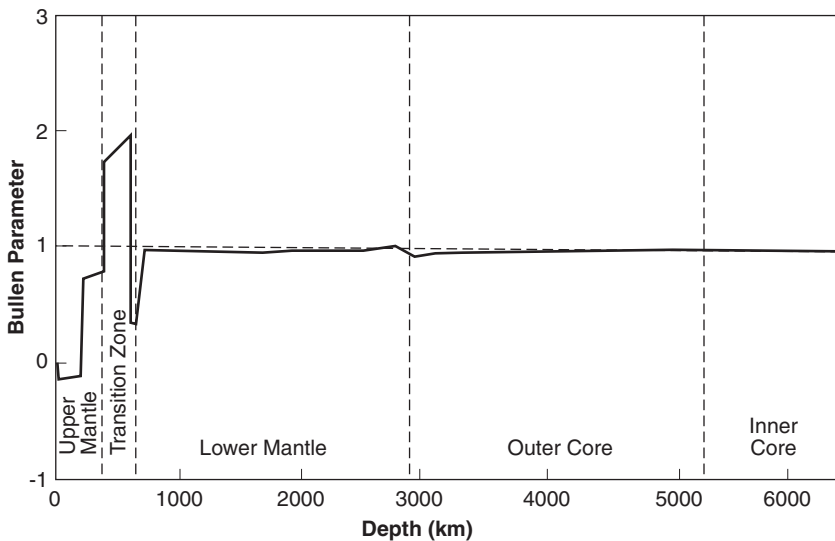

Fig. 1-4. Depth variations of the Bullen parameter (after Dziewonski and Anderson 1981).

Box 1-2. The Bullen Parameter

The Bullen parameter is defined by the ratio of the depth variation of the actual density to the hypothetical depth variation corresponding to adiabatic compression. The adiabatic density gradient, $\left(\frac{d\rho}{dz}\right)_{ad}$, can be determined from seismological observation. Note first that the compression of materials by seismic waves occurs much faster than thermal diffusion. This can be seen by comparing the time-scale for thermal equilibrium with the time-scale of deformation by seismic waves. The time-scale of thermal diffusion is given by $\tau_{th} \sim \lambda^2/\kappa$, where λ is the wave length of seismic waves ($\sim 10-1{,}000$ km) and κ is thermal diffusivity ($\sim 10^{-6}\ m^2 s^{-1}$). The time-scale of deformation by seismic wave is $\tau_{def} \sim 1/\omega$, where ω is the frequency of seismic waves. Therefore, $\frac{\tau_{th}}{\tau_{def}} \sim 10^{14-21}$, indicating that the deformation associated with seismic wave propagation is adiabatic. The bulk modulus for adiabatic deformation is defined by

$$K \equiv \left(\frac{dP}{d\log\rho}\right)_{ad}. \tag{B1-2-1}$$

Therefore one gets

$$\left(\frac{d\rho}{dP}\right)_{ad} = \frac{1}{V_\phi^2}, \tag{B1-2-2}$$

where we used the definition $V_\phi^2 \equiv \frac{K_S}{\rho} = V_p^2 - \frac{4}{3}V_s^2$. Now in Earth, pressure is determined by hydrostatic equilibrium, $dP = \rho g dz$, thus,

$$\left(\frac{d\rho}{dz}\right)_{ad} = \frac{\rho g}{V_\phi^2}. \tag{B1-2-3}$$

Therefore,

$$(\text{Bullen parameter}) = \frac{\left(\frac{d\rho}{dz}\right)\left(V_p^2 - \frac{4}{3}V_s^2\right)}{\rho g}. \tag{B1-2-4}$$

All of these quantities can be obtained from an Earth model based on seismology.

transformations in the transition zone. This prediction was later confirmed by the experimental work of Ted Ringwood and Lin-Gun Liu in Australia and Syun-iti Akimoto, Naoto Kawai, Mineo Kumazawa, and Eiji Ito in Japan. In addition to an initial quick identification of new
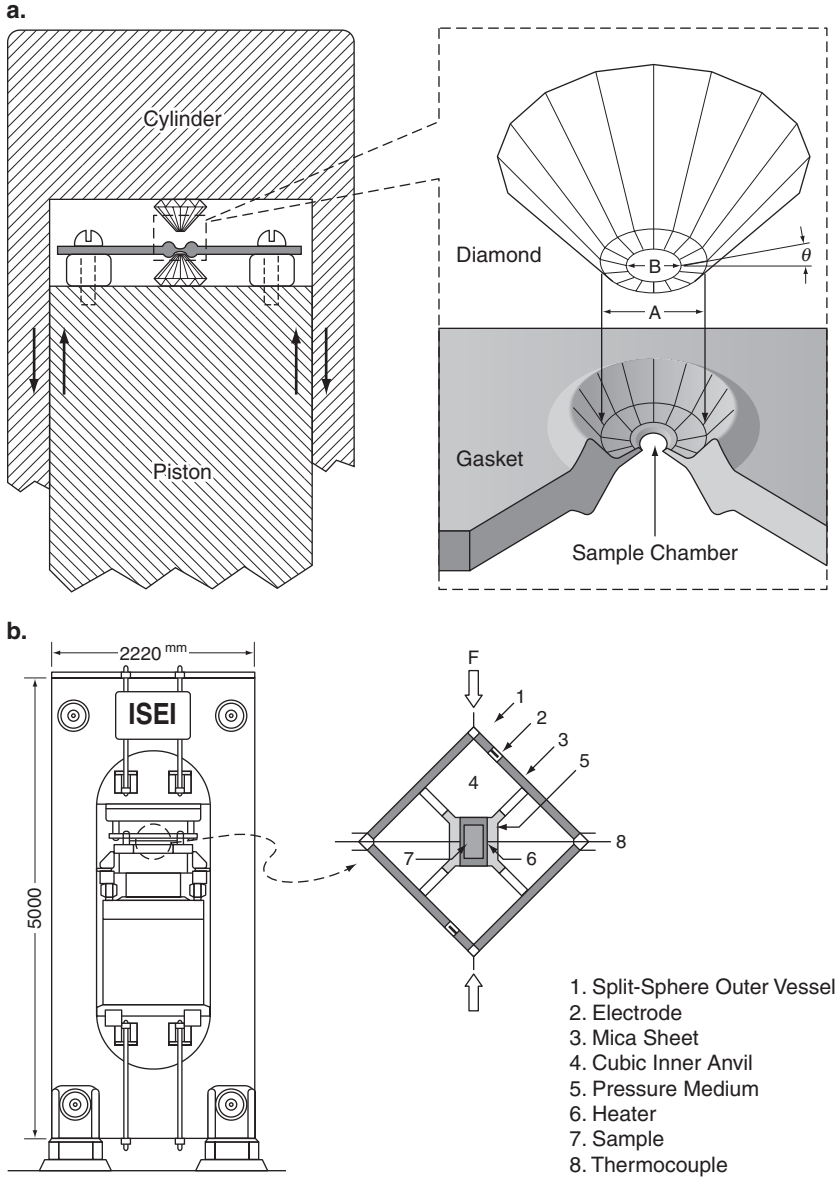
a.

Cylinder

Piston

Diamond

B

A

θ

Gasket

Sample Chamber

b.

2220 mm

ISEI

5000

F

1
2
3
5
4
7 — 6
8

1. Split-Sphere Outer Vessel
2. Electrode
3. Mica Sheet
4. Cubic Inner Anvil
5. Pressure Medium
6. Heater
7. Sample
8. Thermocouple

Fig. 1-5. (a) A schematic drawing of a diamond anvil cell (DAC). A small sample ($< 10^{-4}$ mm^3) is squeezed between two single crystals of a diamond. A high temperature can also be generated by laser-beam heating. The pressure and temperature that can be generated by a DAC exceed those at the center of Earth. (b) A multi-anvil apparatus designed originally by Naoto Kawai. A relatively large volume of samples ($1-10$ mm^3) can be squeezed. The maximum pressure and temperature conditions achievable with this apparatus are limited to those in the shallow lower mantle. However, since the sample volume is large and the temperature distribution is nearly homogeneous, this type of device is useful for studies of chemical re-actions (phase transformations) and plastic properties (from Mao and Hemley 1998).

phases by a laser-heated diamond-anvil cell (DAC; fig. 1-5a), quantitative analyses of phase diagrams using a multianvil apparatus (fig. 1-5b), developed by Kawai and others in Japan, played a major role.

Phase transformations in the mantle transition zone affect the pattern of mantle convection. Recent seismic tomography has suggested that mantle convection seems to change its pattern around these depths (chaps. 3, 4). One of the important themes in the study of mantle dynamics is to understand how convection is affected by phase transformations. At greater depths in the lower mantle and the core, the Bullen parameter is close to 1, suggesting a nearly adiabatic temperature gradient. However, one needs to be careful about temperature gradients. Even if a temperature gradient is twice as large as the adiabatic value, the Bullen parameter varies from 1 to 0.92 at the utmost, which is still consistent with seismological observations (see sec. 1-2).

The phase diagram of $(Mg,Fe)_2SiO_4$, which is a representative mantle mineral, is shown in figure 1-6a. Minerals with this composition have the olivine crystal structure at low pressures, and they transform to wadsleyite (modified spinel), then to ringwoodite (spinel) at high pressures. At the pressure of about 24 GPa, they finally decompose into perovskite and magnesiowüstite. The gross picture of these phase transitions had been established almost completely by the mid-1980s.

While the phase diagram of iron, the main constituent of the core, has long been known under low-pressure conditions, it has not yet been completely known under the high-pressure and high-temperature field corresponding to the actual core conditions. On the basis of currently available theoretical and experimental grounds, iron with the hexagonal close-packed (hcp) structure seems to be the most likely material for the inner core (fig. 1-6b).

Once the phase diagram of a material is determined, the next step is to measure the density and elastic properties of each phase. This type of research has made rapid progress since the pioneering work by Birth (1952), and the measurements for core materials as well as mantle materials have been conducted. Particularly important in these studies are the development of new techniques of the generation of high pressures and temperatures and the development of techniques of measurements of properties under extreme conditions (from small samples, usually less than 1 mm^3). Earth scientists have taken the lead in these areas, and these technologies often contribute to materials science and engineering, such as the development of new hard materials. The results of experimental studies have demonstrated that the depth variations of density and elastic properties
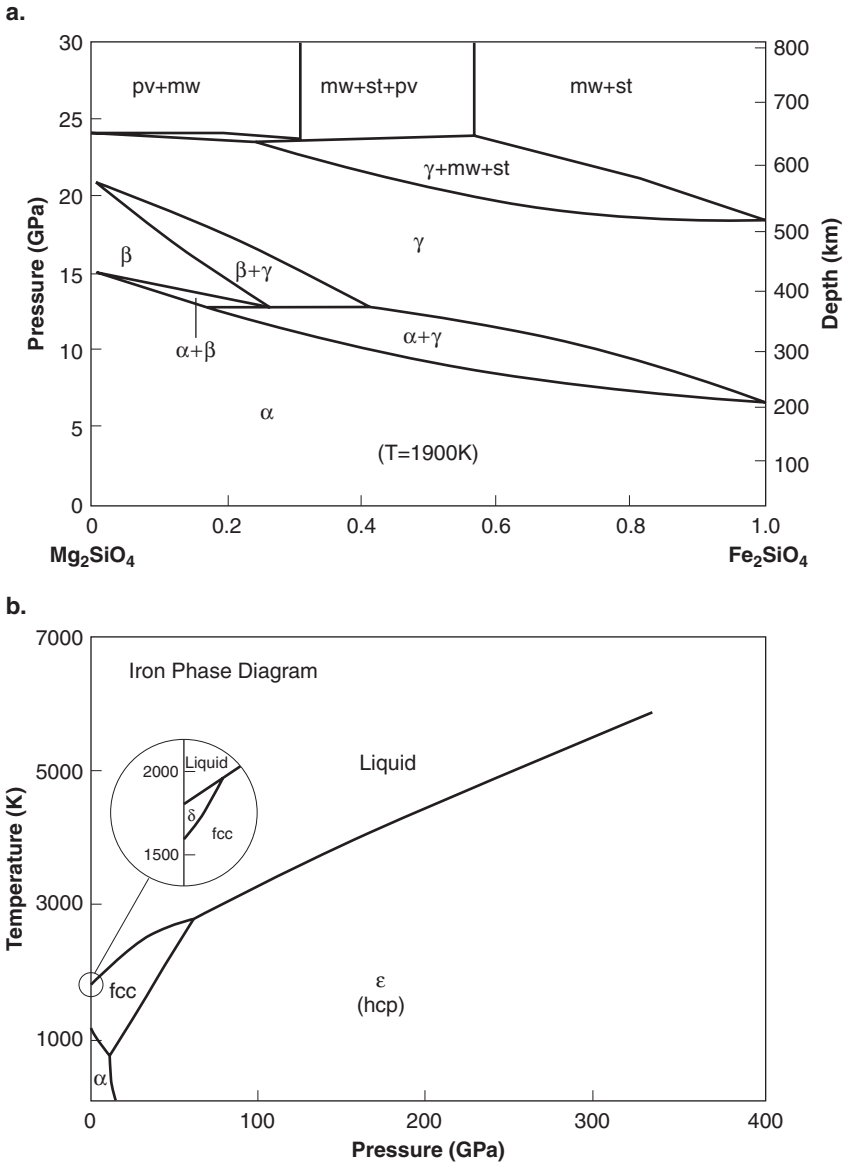
a.

b.

Fig. 1-6. Representative phase diagrams for Earth materials. (a) A phase diagram of $(Mg,Fe)_2SiO_4$ (Ringwood 1991) and (b) a phase diagram of Fe (Anderson 2002). A phase diagram for inner-core pressures and temperatures is still incomplete.
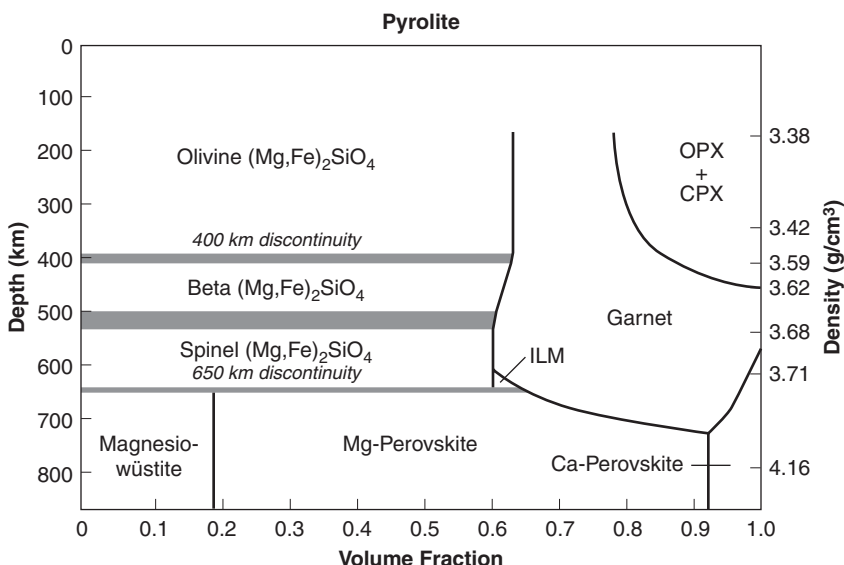
Fig. 1-7. A model for Earth constituent materials (after Ringwood 1991).

within Earth can be explained mostly as the effect of compression and abrupt phase transformations at some particular depths. The depth variation of major constituents for the mantle and the core, as estimated from these experimental results and seismological observations, is shown in figure 1-7.

There are two important points regarding phase transformations. The first is that the pressure at which a phase transformation takes place varies with temperature (temperature dependency), and the second is the possibility of non-equilibrium phase transformation.

a. Temperature Dependence of Phase Transformation

Many of phase transformations that take place within Earth are caused by the increase in pressure, so pressure is the most important variable. The effect of temperature, however, cannot be ignored. Thus, the pressure at which a given phase transformation occurs depends on the temperature. Therefore, when one plots the stable phases on a pressure-temperature plane, the boundary between the stability fields of two phases has a slope. This slope, $\left(\frac{dP}{dT}\right)_{eq}$, is called the Clapeyron slope, after the nineteenth-century French physicist who first constructed such a diagram (box 1-3).

Box 1-3. Phase Transformations and the Clapeyron Slope

A given substance, say H_2O, assumes various structures (phases) dependent on thermodynamic conditions (pressure, temperature, etc.). At room pressure, H_2O will be liquid water if the temperature is between 273 K and 373 K; below 273 K, it will be solid ice (ice I). This transition between liquid water and solid ice (ice I) will occur at different temperatures at different pressures. Similarly, carbon will assume graphite structure at room temperature and pressure, whereas it transforms to a diamond structure at higher pressures. The stability of a material under different conditions is determined by the Gibbs free energy, G,

$$G = U + PV - TS \tag{B1-3-1}$$

where U is internal energy, V is volume, and S is entropy (P is pressure and T is temperature). A given material will assume a structure with the lowest Gibbs free energy. Therefore a material assumes a structure with a smaller volume (i.e., higher density) at higher pressures, and a structure with higher entropy at higher temperatures. French physicist Clapeyron invented a way to show the stability of various phases, called a phase diagram. If we choose temperature (T) and pressure (P) as independent variables, then a P-T plane will be divided into several regions, each region corresponding to a stability field of a given phase. The slope of the boundary between the stability fields of two phases is called the Clapeyron slope.

Consider two adjacent points across a phase boundary corresponding to a temperature and pressure of (T, P). Because the two phases are in equilibrium at this boundary (T, P), the Gibbs free energy of the two phases (1 and 2) must be equal. Thus,

$$U_1 + PV_1 - TS_1 = U_2 + PV_2 - TS_2. \tag{B1-3-2}$$

At a nearby point ($T + dT$, $P + dP$), we have a similar relationship:

$$U_1 + (P + dP)V_1 - (T + dT)S_1 = \\ U_2 + (P + dP)V_2 - (T + dT)S_2. \tag{B1-3-3}$$

From equations (B1-3-2) and (B1-3-3), we have

$$\left(\frac{dP}{dT}\right)_{eq} = \frac{S_1 - S_2}{V_1 - V_2} \tag{B1-2-4}$$

where *eq* is used to clearly indicate that this slope is for equilibrium between the two phases. This is the *Clapeyron slope*. The Clapeyron slope shows how temperature affects the pressure at which a given phase transformation occurs. If phase 1 is a high-pressure phase and the phase 2 is a low-pressure phase, then $V_1 < V_2$. But the relative magnitudes of entropy of the two phases can vary from one case to another. Usually, a high-pressure phase has a stiffer structure and has lower entropy than a low-pressure phase ($S_1 < S_2$). In such a case, the Clapeyron slope has a positive value. However, a phase transformation to a high-pressure phase can reduce the strength of the chemical bond in some unusual cases. In these cases, the Clapeyron slope has a negative value.

The Clapeyron slope is related to the change in volume and entropy associated with the phase transformation—namely,

$$\left(\frac{dP}{dT}\right)_{eq} = \frac{S_1 - S_2}{V_1 - V_2}, \tag{1-3}$$

where S_1 is the molar entropy of phase 1 (the high-pressure phase), S_2 is the molar entropy of phase 2 (the low-pressure phase), V_1 is the molar volume of phase 1 (the high-pressure phase), V_2 is the molar volume of phase 2 (the low-pressure phase), respectively. While $V_2 - V_1 > 0$ because the high-pressure phase always has a smaller volume than the low-pressure phase, the difference in entropy depends on the nature of each phase transition. Entropy is a key concept in thermodynamics that represents the degree of *disorder* of a system. The lattice vibration of atoms has the greatest influence on the entropy of solids. A higher frequency of lattice vibration corresponds to lower entropy. The high-pressure phase usually has stronger chemical bonding, which results in a higher frequency of atomic vibration, so it has less entropy than the low-pressure phase ($S_2 - S_1 > 0$). Thus $(dP/dT)_{eq} > 0$. When the coordination number (the number of atoms adjacent to a particular atom) is changed significantly by a phase transformation, however, the Clapeyron slope can be negative—for example, the phase transformation from ringwoodite to perovskite and magnesiowüstite (fig. 1-8). In this case, a silicon atom is surrounded by four oxygen atoms in ringwoodite but by six oxygen atoms in perovskite. Thus, the Si-O bonding in perovskite is relatively weak, and its entropy becomes greater, resulting in negative ΔS,—that is, $(dP/dT)_{eq} < 0$. Re-
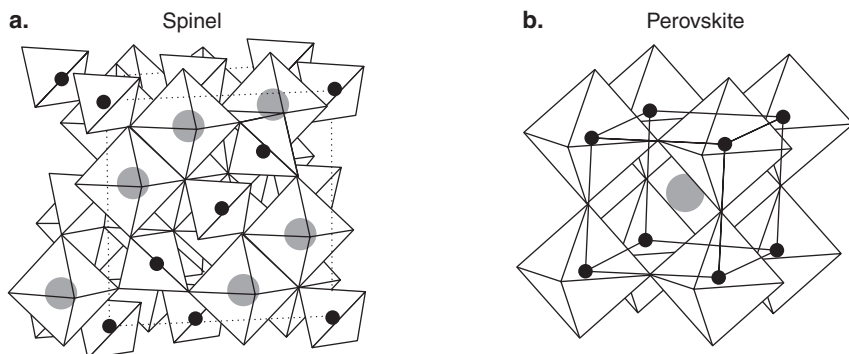
Fig. 1-8. The crystal structures of (a) spinel and (b) perovskite. Small dark circles represent Si, and large hatched circles are Mg (or Fe). (The perovskite structure shown here is the ideal cubic structure.) Note that Si atoms in spinel are located in the tetrahedra surrounded by four oxygen atoms, whereas Si atoms in perovskite are located in the octahedra surrounded by six oxygen atoms.

markably, this negative Clapeyron slope associated with the transformation to perovskite was predicted by the chemist Alexandra Navrotsky (then at University of Arizona, now at the University of California at Davis) (1980) and later experimentally demonstrated by Ito and Yamada (1982). Because of this temperature dependence of phase-transformation pressure, the depth of a phase transformation can vary from place to place if there are regions with anomalous temperatures.

This temperature dependence of phase transformation leads to an important effect. Consider, for example, the depth of the 660-km boundary. At around this depth, ringwoodite transforms to perovskite and magnesiowüstite. Because this phase transformation has a negative Clapeyron slope, it takes place at depths deeper than 660 km in regions with low temperature anomalies. This means that a less-dense phase (ringwoodite) extends into greater depths there than in other normal places, and this results in additional buoyancy, which tends to prevent the colder materials from sinking. The topography of this phase boundary caused by the lateral variation in temperature and therefore acts as a resisting force for convection (chap. 4). Although we usually use terms like the 410-km boundary and the 660-km boundary, the depths of these boundaries are actually different at different places. The fine structure of these phase boundaries has become observable in the 1990s (Shearer and Masters 1992).

b. Non-equilibrium Phase Transformation

In most of Earth's deep interior, the temperature is so high that phase transformations take place near chemical equilibrium. Inside a cold subducting plate, however, the temperature is much lower than the average mantle temperature. Phase transformations take place very slowly in these cold regions, so non-equilibrium phase transformation is possible. This type of non-equilibrium phase transition is frequently seen in our daily life. The best example is the diamond. Diamonds, which are one form of carbon, are unstable at room temperature and pressure. Though they should transform to graphite, this does not happen because the rate of phase transition at room temperature is extremely slow. A similar case can be found for silicates. For example, the depth of the olivine-wadsleyite transformation, which has a positive Clapeyron slope, should be shallower than average (410 km) for regions with low-temperature anomalies. If the temperature is very low, however, phase transformation may not take place at equilibrium, and olivine may be brought much deeper, until the phase transformation finally occurs. Although this possibility was already pointed out in a pioneering paper by Sung and Burns (1976), we were not able to discuss it quantitatively until Dave Rubie, now at Bayreuth in Germany, and his colleagues conducted detailed experimental research using a synchrotron radiation facility (box 1-4) in Japan (Rubie et al. 1990). Based on these results, some people consider that non-equilibrium transformation may occur in a subducting slab and that the non-equilibrium phase transformation is a cause for deep earthquakes (see chap. 5).

C. The Layered Structure of Earth: The Second-Order Approximation−Chemical Structure of the Mantle and the Core

In the previous, first-order approximation, the crust, the mantle, and the core are assumed to be chemically homogeneous. This assumption does not hold at a higher-order approximation. A prominent example is the core. The density of the outer core is lower than that of pure iron (or iron-nickel alloy) by about 10%, suggesting that it contains impurities in large quantities. On the other hand, the density of the inner core is close to that of pure iron (or iron-nickel alloy). Thus, the inner-outer core boundary is thought to be due not only to the change in phase (liquid and solid), but also to the change in composition. A plausible model to explain these phase and compositional differences is that the inner core has grown from materials in the outer core as a result of cooling. Because solubility of im-

Box 1-4. Synchrotron Radiation and New Mineral Physics Studies

Synchrotron is one of the particle accelerators designed for high-energy physics research (fig. B1-4-1). Charged particles (such as electrons) are accelerated by a huge ring of magnets and when they collide, strong X rays are emitted. The strength of these X rays is many orders of magnitude higher than the X rays generated by a conventional device, making them highly useful for research in many areas, including high-pressure mineral physics and medical science. In high-pressure mineral physics research, small samples
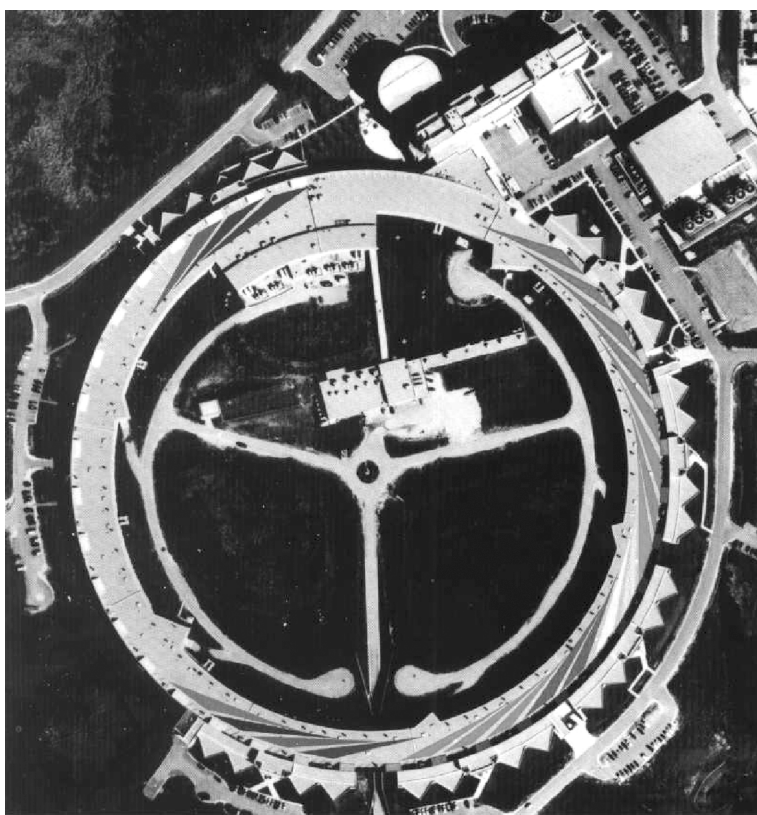

Fig. B1-4-1. The synchrotron facility at Argonne National Laboratory. Along a large circular ring (~ 150 m radius), charged particles are accelerated and upon collisions they emit high-energy X-rays. These X-rays can be used to investigate the properties of materials under high pressures.

surrounded by other materials that often absorb X rays must be investigated. Therefore, strong X rays are essential for these studies. A measurement that would take several days with a conventional X-ray source can be made within a few seconds with a synchrotron radiation facility. As a result, a number of experimental studies which were impossible or difficult are now possible with this powerful new facility. Many physical properties can be investigated through the use of the synchrotron facility, including phase relationships, elastic constants, densities (equation of state), viscosity of melt, plastic (rheological) properties, and the kinetics of phase transformations. These facilities are located in national laboratories such as Brookhaven National Laboratory and the Argonne National Laboratory in the United States, KEK (Ko-Energy-Ken) and Spring8 in Japan, and Grenoble in France.

purities is lower for solids than for liquids, impurities are accumulated in the outer core. According to this model, the growth of the inner core releases latent heat by solidification as well as gravitational potential energy by the removal of impurities, which provides significant energy for convection in the outer core (chap. 6).

How about the mantle? Is the mantle chemically homogeneous? As I explained before, most of the mantle structure, especially the structure of the transition zone, can be attributed to a first order, to the phase transformations of constituent minerals. This does not mean, however, that all of the mantle structure can be explained by changes in the physical properties of an isochemical material. Similarly to the core, chemical stratification is possible in the mantle due to the melting process. Differences between continental mantle and oceanic mantle are generally considered to be due to chemical heterogeneity. By combining seismic and gravity observations, Tom Jordan, then at Scripps Oceanographic Institution in California, concluded that the continental upper mantle is colder than the oceanic upper mantle, but that the density of the continental upper mantle is about the same as that of the oceanic upper mantle (1975). He suggested that this is due to the difference in the chemical composition between the continental upper mantle and the oceanic upper mantle: because the continental mantle, compared to the oceanic mantle, experienced more extensive partial melting, it is relatively depleted in dense minerals like garnet (see chap. 2).

Whether there is compositional heterogeneity in the deep mantle or not is an important problem related to the chemical evolution of the mantle, but it is still highly controversial. From geochemical observations, we know that there are at least two or more regions with different chemical compositions, and that these regions have not been mixed very much for more than one billion years (e.g., Hofmann 1997) (chap. 4). One hypothesis is that these different regions are the upper and lower mantle. In this case, a part of the difference in physical properties between the upper and lower mantle is due to compositional difference. In fact, some scientists have suggested that the density difference between the upper and lower mantle is due to the difference in the concentration of iron and other elements (Jeanloz and Knittle 1989). Others have proposed that the difference lies in silica concentration (e.g., Stixrude et al. 1992). Some consider that 410 km, not 660 km, is the chemical boundary. Don Anderson (1989) and Carl Agee (1993), argue that seismological data are consistent with the idea that the region between 410 and 660 km has a garnet-rich chemical composition. In principle, these hypotheses can be tested if we conduct accurate measurements of physical properties and compare them to seismological data. The problem is that differences in physical properties among different models are small, and the differences among the various models themselves are within the uncertainties in experimental and seismological observations. Furthermore, these properties (e.g., density) depend on temperature, which itself has some uncertainty, so it is difficult to arrive at a definitive conclusion.

In this type of argument, the main issue has usually been whether the upper and lower mantle have the same chemical composition. Recently, Rob van der Hilst and H. Karason (1999), at the Massachusetts Institute of Technology (MIT), suggested that a chemical boundary may exist at ~ 1,600 km depth, not at the upper-lower mantle boundary (660 km). Their argument is based on the anticorrelation, in their seismic tomography, between the bulk modulus and the shear modulus observed deeper than ~ 1,600 km, which cannot be explained as the effect of temperature alone (I will explain this in more detail in chap. 3).

Though small in quantity, water distribution in the mantle is also heterogeneous. The distribution of water can be estimated by chemical analyses of igneous rocks. Water is abundant at island arcs such as Japan and less abundant at mid-ocean ridges. Water is also redistributed by the partial melting of rocks; a large volume of water can be dissolved into melt (molten rocks), whereas very little water can be dissolved in solid minerals. Thus, partial melting results in the drying out of minerals because

water in minerals is absorbed into melt. Water can greatly modify the viscosity of rocks, so the distribution of water is important for geodynamics. It is also considered to have a large effect on the layering of the lithosphere and the asthenosphere. This issue will be discussed in more detail in chap. 2.

Though the origin of chemical heterogeneity of the mantle is yet uncertain, its presence is strongly supported by geochemical observations. The question is how chemically different materials are distributed spatially and temporally. We have just started to work on this problem on the basis of observation. It is a fundamental problem related to the evolution of Earth and its dynamics, and significant progress in the future is expected. Chapters 3 and 4 of this book deal with this problem in some detail.

1-2. THE THERMAL STRUCTURE OF EARTH'S INTERIOR

In the previous discussion on the composition of Earth's interior, the issue of temperature was not so important because density and elastic properties vary little by temperature. However, some properties such as viscosity can strongly depend on temperature. Because viscosity is one measure of the mobility of materials, temperature is expected to have a great effect on flow patterns in Earth's interior. Conversely, the temperature distribution also depends on the pattern of convection and the distribution of the heat source. If we can estimate the temperature distribution, the results can provide useful constraints on geodynamics such as the convective pattern.

The temperature gradient near the surface can be measured by measuring the temperatures in a deep well. These measurements show that the temperature gradients are about 10–50 K/km, although they can differ from one region to another. Heat transfer near the surface occurs mostly by thermal conduction, so we may assume a roughly constant temperature gradient to some depth. If the temperature in the deep portion is estimated this way, the temperature at the depth of 200 km is estimated to be 2,000–10,000 K. Since these temperatures exceed the melting temperatures of rocks, rocks must be largely molten at these greater depths. This inference is, however, inconsistent with the observation that shear waves propagate through the mantle. We must conclude that our assumption of (nearly) constant temperature gradient is wrong; a temperature gradient in the deep interior should be much smaller than that near the surface.

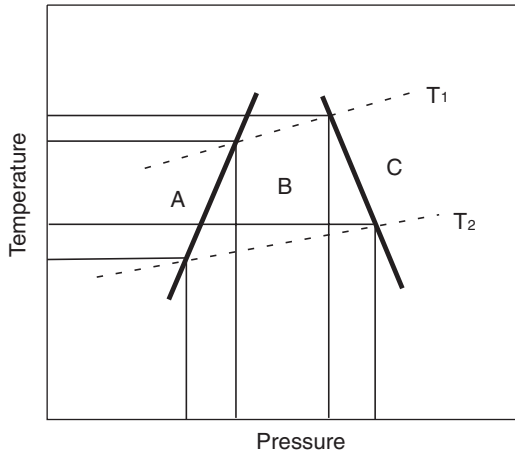This conclusion is also supported by the following argument. As ex-

Fig. 1-9. The Clapeyron slope and the depth and temperature of phase transformation. Phase transformation A-B has a positive Clapeyron slope, and B-C has a negative one. By observing phase transformation depth by seismology, temperature at that depth can be estimated. T_1 and T_2 are hypothetical geotherms.

plained before, the pressure at which a phase-transformation pressure occurs depends on temperature. Seismic observations tell us that there are seismic discontinuities at the depths of 410 km and 660 km. If these discontinuities are caused by phase transformations, these depths can be used to estimate the temperatures at the discontinuities (fig. 1-9). By this approach, Eiji Ito and Tomoo Katsura, at Misasa in Japan (1989) estimated the temperature of the mantle transition zone and showed that the temperature gradient there (≈ 0.5 K/km) is much smaller than the surface value. This variation in temperature gradient can also be supported by an argument using the Bullen parameter, shown in equation (1-2).

Though we do not know for certain the temperature below 660 km, we can at least place several important constraints. For example, at the depth of 5,150 km (the inner-outer core boundary) the temperature must be the melting temperature of iron. So we can infer the temperature there by measuring the melting temperature of iron. This type of measurement is, however, very difficult, and the best estimate is ~ 5,000 K with a large margin of error, of 500 K or more. This estimate is still useful, and we can draw an important conclusion on the temperature distribution and heat-transfer mechanism in Earth's interior. In the outer core, where the magnetic field is believed to be generated by a dynamo, we can assume vigorous convection; its temperature gradient is close to the adiabatic thermal

gradient (around 0.6–0.8 K/km). This gradient gives 3,600 ± 500K for the uppermost part of the core (i.e., the core-mantle boundary). This is significantly higher than 2,500 K, which is calculated from the estimated temperature at 660 km (1,900 K), assuming the adiabatic thermal gradient. From this discrepancy, we can conclude that there must be a high temperature gradient at the lowermost mantle (called the D″ layer).

This is the current standard model for temperature distribution in the mantle (fig. 1-8a). In this standard model, thermal boundary layers exist only at the uppermost and lowermost parts of the mantle. This type of temperature distribution is expected when not only the mantle but also the core act as heat sources and when mantle convection is of whole-mantle scale. In this case, there is a large temperature gradient at the core-mantle boundary, so plumes are likely to form there.

To understand the thermal structure in terms of the dynamics of Earth's interior, we need to know some fundamentals of heat transfer. Thermal conduction and thermal convection are two important heat-transfer mechanisms in Earth's interior (heat transfer by radiation may also be important in the very hot regions). Heat transfer by convection occurs when a material is carried to a position where the temperature is different from that of a material being transported. Vertical convective heat transfer is not expected in the lithosphere (a shallow cold region), so vertical heat transfer occurs mostly by conduction within the lithosphere. Conduction is not an efficient mechanism of heat transfer, so its thermal gradient is large. At greater depths, viscosity becomes low due to the higher temperature, and (vertical) heat transfer by convection becomes more efficient. The temperature gradient for convection is much smaller than that for conduction and is close to the adiabatic thermal gradient (0.3–0.4 K/km). The basic physics for this change in heat-transfer mechanism, from thermal conduction to convection, can be understood as follows: Convection takes place by temperature difference in either a horizontal or a vertical direction. Let us focus here on the vertical temperature difference because we are concerned with heat transfer from the deep mantle to the surface. Consider a layer of viscous fluid heated from below. The deeper part of this layer becomes lighter by heating and starts to rise. As it rises, it loses heat and hence buoyancy through conduction of heat to the surrounding colder fluid. Continuous vertical fluid motion—that is, thermal convection—is possible only when the time-scale for thermal conduction is larger than that for vertical advection, so that fluid temperature does not change so much during its vertical motion. Lord Rayleigh, a British physicist, analyzed this problem in the early-twentieth century and found a con-

dition under which thermal convection occurs. The condition for thermal convection is given by a nondimensional number, the Rayleigh number, defined by

$$Ra \equiv \frac{\rho g h^3 \alpha \Delta T}{\eta \kappa}, \tag{1-4}$$

where α is thermal expansion, ρ is density, ΔT is the temperature difference between top and bottom boundaries, g is acceleration due to gravity, h is layer thickness, k is thermal diffusivity, and η is viscosity. The Rayleigh number denotes the ratio of the thermal diffusion time-scale to the vertical advection time-scale. Convection takes place when the Rayleigh number exceeds a certain value (approximately 1,000). Viscosity is the most uncertain quantity when evaluating the Rayleigh number for the mantle, but it is generally thought to be about 10^{20}–10^{22} Pa·s on the basis of various kinds of geophysical inference, as I will discuss later. With this range of viscosity, the Rayleigh number is 10^5–10^7 for the mantle. Therefore we conclude that the mantle is vigorously convecting. The thermal structure and flow pattern of an intensely convecting fluid layer can be well explained by the *boundary layer theory*. At high Rayleigh numbers, the buoyancy forces driving convection are concentrated in the top and bottom boundary layers, and fluids slowly move vertically in between. Within the boundary layers, fluid motion is almost horizontal, and a vertical temperature gradient there is controlled by thermal conduction. The boundary layers have a large temperature gradient, which is determined by the layer thickness, thermal conductivity, and heat flux through the layer. In most of the fluid layer, on the other hand, vertical fluid motion is significant and heat is transported by advection, and hence the temperature gradient is adiabatic (0.3–0.4 K/km). Thus, typical temperature distribution in a fluid layer becomes the one shown in figure 1-10a. Note that the top boundary layer is well known as the lithospheric plate, but the bottom boundary layer at the base of the mantle is still elusive. The temperature gradient at the bottom layer is determined by the heat flux from the core, which is quite uncertain.

There are a few uncertainties about this standard model. The first one is whether a boundary layer exists between the upper and lower mantle. As I will explain later in detail, there are a growing number of observations that cast doubt about a simple model of whole-mantle convection. If convection is separated into the upper and lower mantle, there should be a thermal boundary layer between them, and the thermal gradient there should be larger than the adiabatic value (fig. 1-10c). The second uncer-
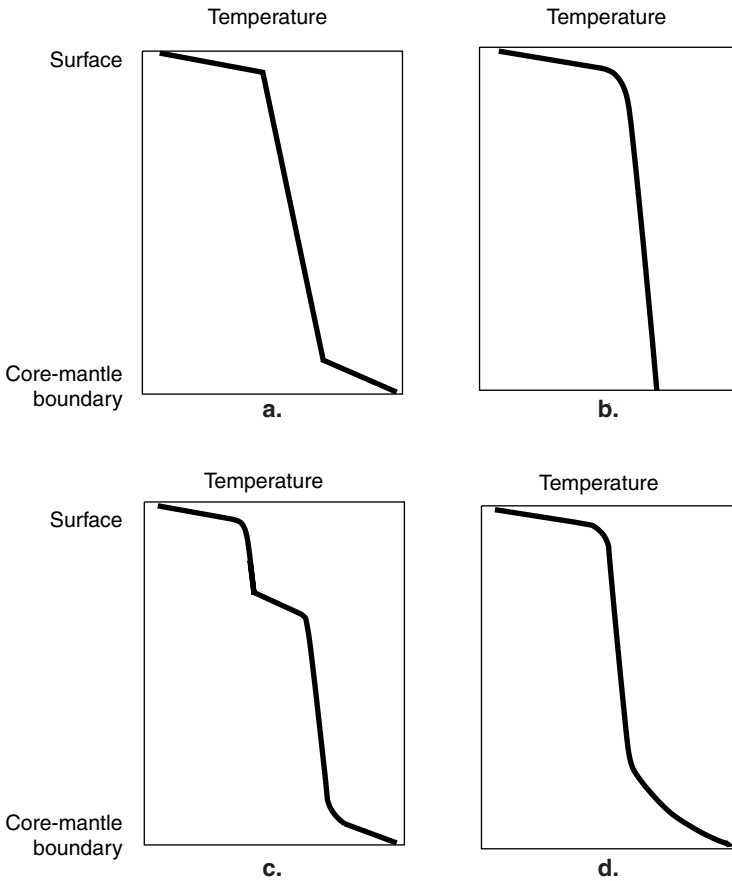
Fig. 1-10. Temperature distribution in the mantle and the mode of convection. (a) Whole-mantle convection: the case of a large heat flux from the core. Thermal boundary layers are formed at both top and bottom boundaries. (b) Whole-mantle convection: insignificant heat flux from the core. The thermal boundary layer is formed only at the top boundary. (c) Layered convection: large core heat flux is assumed. The thermal boundary is formed also at the boundary of two layers. (d) Whole-mantle convection with less vigorous convection at the deep mantle. Thermal gradient at the weak convection region is higher than the adiabatic thermal gradient.

tainty is related to the temperature gradient in the lower mantle. Recent studies seem to indicate that the thermal gradient in a significant portion of the lower mantle is larger than the adiabatic value. On the basis of their seismic tomography, van der Hilst and Karason (1999) suggested that there is a chemically distinct layer below ~ 1,600 km, which does not mix

with the shallower region. In this case, because there is no vertical fluid motion below ~ 1,600 km, the temperature gradient there should be larger than the adiabatic value (fig. 1-10b). This possibility of a superadiabatic thermal gradient in the lower mantle is also supported by the radial distribution of elastic constants and density as estimated by seismological studies and by the radial variation of viscosity (which I will explain later).

In the above arguments, we have considered only the radial structure of temperature. Of course, temperature in a convecting fluid varies also in a horizontal direction. Temperature beneath mid-ocean ridges where hot-mantle materials rise is much higher than that beneath old cratonic continents. Similarly, there seems to be a large lateral thermal gradient at the core-mantle boundary.

1-3. RHEOLOGICAL STRUCTURE: SEISMIC WAVE ATTENUATION AND VISCOSITY

In addition to elastic properties, anelastic properties and viscosity are also important for geodynamics. In this section, I will summarize the distribution of seismic wave attenuation and viscosity. Both seismic wave attenuation and viscous deformation occur through non-elastic deformation and they are often collectively called *rheological properties* (*rheo* means "flow" in Greek) (box 1-5). A seismic wave is an elastic wave, but its amplitude decreases as it propagates because elastic energy is converted into heat. This conversion occurs because Earth materials are not a perfect elastic body and have some viscous character. This behavior of material is called *anelasticity*. Briefly, anelasticity is a type of mechanical behavior of a material that is between elastic and viscous behavior. If Earth were a perfect elastic body, it would continue to oscillate forever after an earthquake, and this would give us a lot of trouble. Fortunately, seismic oscillation eventually stops because Earth is not a perfect elastic body.

Because viscous behavior in solids occurs only at high temperatures, the anelastic effect in solid rocks is particularly prominent at high temperatures and low frequencies. Therefore, the effects of anelasticity are negligible for the high-frequency (100–1,000 MHz) elastic waves usually used in laboratory experiments. However, seismic waves have low frequencies (1–0.001 Hz), so this anelastic effect cannot be neglected. The anelastic effect is expressed by Q, where Q^{-1} denotes the fraction of elastic vibration energy dissipated as heat. A smaller Q means low attenuation. In the mantle, Q is on the order of 50–500. That is, about 1–10% of energy is lost during one cycle of vibration. Q has a close relation to viscosity, and

Box 1-5. The Atomistic Basis of Elastic and Plastic (Viscous) Deformation

In solid Earth geophysics, we often deal with the deformation of materials by an external force. When a small force is applied at a relatively low temperature and for a short time, deformation occurs instantaneously. As soon as the force is removed, the material goes back to its original undeformed state. This style of deformation is called *elastic* deformation. Deformation associated with the propagation of seismic waves is nearly elastic. In contrast, deformation associated with long-term gravitational forces (caused by buoyancy forces) in the hot mantle occurs *viscously*. Deformation in this case occurs slowly, and even after the removal of force, permanent deformation (strain) remains.

The difference between the two modes of deformation can be understood at the atomic scale. Recall that each solid is made of a periodic array of atoms. Each atom sits at each position where the interaction energy between the atoms is at the minimum. When a small force is applied, atoms move slightly out of their stable (equilibrium) position, which creates a restoring force. When the displacement is small, the restoring force is proportional to the distance of movement and to the *curvature* of the interatomic potential. When the force is removed, this restoring force brings an atom back to its original position. The curvature of the interatomic potential changes slightly with temperature (fig. B1-5-1). At high temperatures, the curvature becomes small and the restoring force is weak (effects caused by the change in the curvature of the interatomic potential are referred to as *anharmonic effects*). Therefore, the elastic constants decrease slightly with temperature, and the dependence is usually nearly linear. The degree of atomic motion can be higher when a large force is applied or a force is applied for a long time. In these cases, atoms can move over the potential maximum to neighboring positions. Once atoms move to neighboring positions, then even after the removal of the force, they will not automatically move back to the original positions: the deformation is permanent (plastic deformation). Consider a case where a small force is applied for a long time at a high temperature. At high temperatures, all atoms vibrate—that is, they do not remain at their equilibrium positions—but by statistical fluctuation, their positions move around their equi-

<div align="right">(continued)</div>

librium positions. Consequently, there is a finite probability that an atom can go over the potential hill to the neighboring position. According to the analysis by Austrian physicist Ludwig Boltzmann, this probability is dependent upon the amplitude of atomic vibration; hence, temperature as ~*exp[−H*/RT]*, where T is temperature, R is the gas constant, and H^* is the activation enthalpy (the height of the potential hill). Thus, the rate at which plastic deformation occurs is proportional to ~*exp[−H*/RT]* and is highly sensitive to temperature. For a typical H^* of ~ 500 kJ/mol, a 100-degree increase in temperature (at $T = 1,600$ K) causes an increase of the rate of deformation by a factor of ~ 10.

Actually, the atomic jump that causes plastic deformation is facilitated by the presence of *lattice defects* such as crystal dislocations and vacant lattice sites (vacancies). Some details of mechanisms of plastic deformation are discussed in chapter 2.
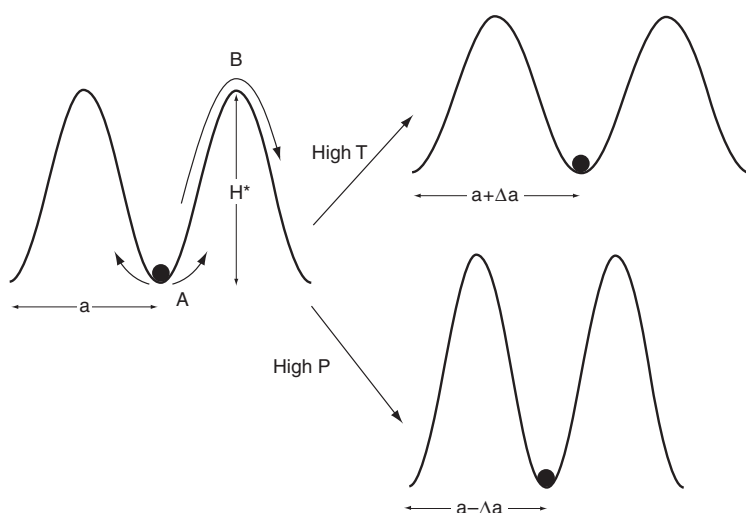

Fig. B1-5-1. A change in interatomic potential with temperature (T) and pressure (P). The curvature of the potential determines the magnitude of the restoring force for small displacement (A) and, hence, elastic constants. With higher temperatures (pressures), the mean atomic spacing, a, increases (decreases) due to thermal expansion (compression), and the curvature of the potential becomes smaller (larger). With a certain probability, atoms can also move into the next stable position (B). This causes permanent (plastic) deformation. The probability of this atomic jump is determined by the height of potential barrier (H^*) and the magnitude of atomic vibration (i.e., temperature), as well as stress magnitude.
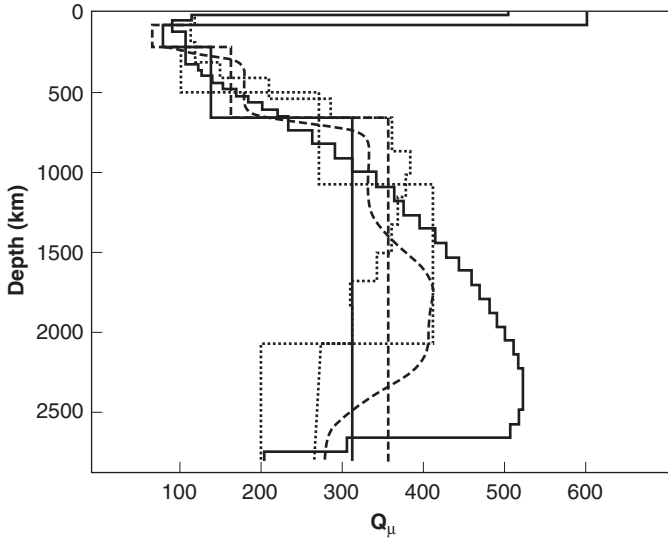
Fig. 1-11. Depth variation of Q_μ (seismic attenuation parameter for shear waves) (after Romanowicz and Durek 2000).

it depends on temperature, pressure, and frequency. Because Q is related to viscosity, it also depends on impurities such as water.

Seismic wave attenuation can be measured by the amplitudes of waves and by the width of peaks in the spectrum of Earth's free oscillation. Figure 1-11 shows a model in which Q is assumed, for simplicity, to be a function of depth only. The Q is large and attenuation is low near the surface. Attenuation is high at depths of 100–300 km. At greater depths in the mantle, attenuation becomes lower again.

Thus, the viscous character of Earth is already seen at the frequency range for seismic waves. At much lower frequencies (or a longer time-scale), the viscous character becomes much more pronounced. By analyzing such long time-scale phenomena, one can estimate the viscosity variation within Earth. About 10,000 years ago, thick ice sheets covered regions in the northern hemisphere. These ice sheets quicky melted about 6,000 years ago, and the regions that had been covered with ice started to rise slowly. By theoretically analyzing this crustal uplift, mantle viscosity can be calculated. This type of research was first conducted in 1935 by Norman Haskell, at MIT, and others. In the latter decades of the twentieth century Dick Peltier, Jerry Mitrovica, and Alessandro Forte (now at the University of Western Ontario), at the University of

Box 1-6. Dimensional Analysis for Relaxation Time for Post-Glacial Rebound

The postglacial rebound (slow vertical crustal movement) after the melting of glacier occurs by gravity force and is controlled by the viscosity of Earth materials. Therefore, its time-scale (relaxation time) (τ) is controlled by gravity (g), density (ρ) and viscosity (η), as well as the size (λ) of the regions of melting (because the gravity force depends on the size). Thus, one can write,

$$\tau \propto g^{\alpha} \eta^{\beta} \lambda^{\gamma} \rho^{\delta}, \tag{B1-6-1}$$

where α, β, γ, and δ are constants. The quantities of both sides of this equation must have the same unit. Therefore,

$$[s] = [m \cdot s^{-2}]^{\alpha} [kg \cdot m^{-1} \cdot s^{-1}]^{\beta} [m]^{\gamma} [kg \cdot m^{-3}]^{\delta}, \tag{B1-6-2}$$

where [] stand for the unit, and hence

$$\alpha - \beta + \gamma - 3\delta = 0,$$
$$-2\alpha - \beta = 1, \text{ and} \tag{B1-6-3}$$
$$\beta + \delta = 0.$$

Solving these equations with a condition $\alpha = \delta$ (because the gravitational force is always in the form ρg), one gets $\alpha = \delta = -1$ $\beta = 1$ and $\gamma = -1$. Thus,

$$\tau = (\rho g \eta / \lambda) \cdot F, \tag{B1-6-4}$$

where F is a nondimensional constant.

Toronto in Canada; Masao Nakada (now at Kyushu University, Japan); and Kurt Lambeck, at ANU, made important contributions to this approach.

Suppose that a surface load is changed at some time. Because hydrostatic balance is then lost, a medium starts to flow. In the case where a load is suddenly removed, a medium flows into a place where a compressive load had formerly been applied, leading to crustal uplift around the region. The time-scale of this uplift depends on the viscosity of the

medium. Thus, by measuring the uplift as a function of time, viscosity can be estimated. A dimensional analysis shows that this time-scale τ depends on viscosity as (box 1-6),

$$\tau = \frac{\eta}{\rho g \lambda} F, \tag{1-5}$$

where λ is the horizontal length scale of the load, ρ is density, g is acceleration due to gravity, and F is a nondimensional number. F is a function of λ, when viscosity is depth-dependent (e.g., $F \approx [\lambda/H]^3$) in the presence of a low-viscosity layer whose thickness is H). If viscosity is assumed to be independent of depth, $F = 4\pi$, and the average mantle viscosity is estimated to be 3×10^{21} Pa·s using the observed values of τ and λ. If viscosity can vary as a function of depth, however, the expression for F becomes complicated, and it is difficult to determine with precision the depth variation of viscosity. In addition, because this approach is based on deformation due to a surface load, it is difficult to determine viscosity in the deep mantle. Consequently, the viscosity in the deep regions of Earth's mantle had been poorly determined and controversial. In the 1970s until the late 1980s, a majority of scientists argued, based on the analysis of postglacial rebound, that the viscosity of the mantle is nearly independent of depth (Cathles 1975; Peltier 1989), until Jerry Mitrovica and Dick Peltier showed that the viscosity of the mantle deeper than ~ 1,200 km cannot be estimated by this method (1991). The work by Masao Nakada and Kurt Lambeck was an exception. Through a thorough analysis of the postglacial rebound incorporating the important effects of coastline geometry, which had been ignored by most of the previous works, Nakada and Lambeck (1989) showed a significant increase in viscosity with depth.

A completely different method for determining the depth variation of viscosity uses gravity observation to investigate how density anomalies in the mantle are maintained by the viscous flow. In the dynamic Earth's interior, mass anomaly is maintained not by static elastic deformation but by dynamic viscous flow. Because of this, even when there is a dense material beneath the surface, we do not always observe a positive gravity anomaly (stronger gravitational field) at the surface. Viscous flow causes surface deformation, and the resultant *dynamic topography* (topography caused by viscous flow) has a large effect on gravity. Consequently, when mass anomalies are maintained by viscous flow, surface gravity anomalies are sensitive to the depth variation of viscosity as well as the mass distribution within Earth. By comparing observed gravity anomalies (more precisely, anomalies in the equi-potential surface of gravity field, or *geoid*)

with those calculated from theoretical models of flow corresponding to a particular viscosity-depth profile, we can estimate the depth profile of viscosity. This method was developed in the mid-1980s by Brad Hager and Mark Richards, then at Caltech (now at MIT and the University of California at Berkeley, respectively) (Hager 1984; Richards and Hager 1984). This important pioneering work has been followed by a number of later studies. Their work gave the first physically sensible explanation for the observed large positive geoid anomalies in the western Pacific. In that region, cold (heavy) materials are sinking into the mantle beneath ocean trenches. Therefore, it might appear obvious that positive geoid anomalies occur there. However, Hager showed that if the mantle viscosity is independent of depth, a popular idea at that time, then one should expect negative geoid anomalies at the surface due to the strong effect of the depression of surface topography due to viscous flow (dynamic topography). The observed positive geoid anomalies in the western Pacific imply that the effects of dynamic topography are much less than expected from the fluid motion with a constant viscosity. In other words, the vertical fluid motion is more sluggish due, presumably, to the rapid increase in viscosity with depth. Using this approach, Hager (1984) was the first to clearly demonstrate that the viscosity of the lower mantle is much higher than that of the upper mantle. To use this method, density anomalies in the mantle must be known. Hager considered only the contribution of subducted oceanic slabs to density anomalies, and he calculated the density of subducted slabs from the estimated temperature difference. Later, as seismic tomography developed, it became popular to use seismic velocity anomalies to estimate density anomalies, and then to estimate viscosity variations. As I will explain in chapter 3, to estimate density anomalies from velocity anomaliesit is not a straightforward process; one of the limitations of this method lies in the uncertainty of estimated density anomalies. As a common issue for gravity-based inferences, there is also a non-uniqueness problem. More than one model can explain the same gravity anomalies.

Figure 1-12 shows the estimated viscosity variations by these methods. Although the estimation of viscosity is not quite as accurate as that of seismic velocity, we may summarize as follows:

1. The average mantle viscosity of the mantle is well constrained; it is around 3×10^{21} Pa·s.
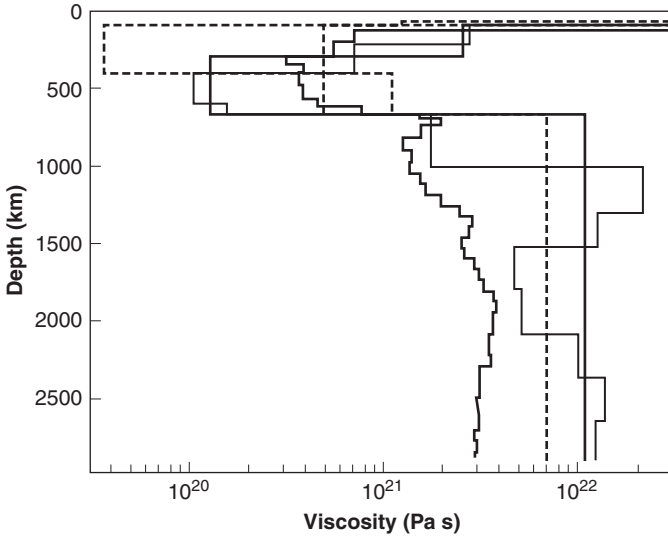2. At shallow mantle (100–300 km), there is a low viscosity layer, with

Fig. 1-12. Depth variation of viscosity (after Peltier 1998). All models consistently show that the upper mantle (asthenosphere) has low viscosity and the lower mantle has high viscosity. Transition zone viscosity as well as viscosity variation in the lower mantle are not very well determined.

viscosity of 10^{19}–10^{20} Pa·s, though it seems to vary from place to place.

3. The deep mantle (deeper than ~ 700 km) has a higher viscosity (10^{21}–10^{22} Pa·s) than the shallower mantle.

4. Some models suggest the existence of a low-viscosity layer in the mid-mantle. Without this layer, mantle convection seems to produce too large a surface deformation (dynamic topography). The details of this low-viscosity layer, such as its exact depth, are, however, not well resolved.

This type of viscosity structure roughly corresponds to the depth variation of seismic wave attenuation shown in figure 1-11. Both results show that the shallow mantle (~ 100–300 km) is weak (low viscosity) and the deep mantle is strong (high viscosity). Attenuation and viscosity are closely related because both of them are caused by similar microscopic processes (the motion of lattice defects) (see box 1-7 and chap. 2). With a reasonable range of parameters determined by laboratory studies, a fac-

Box 1-7. Seismic Wave Attenuation and Viscosity

Both seismic wave attenuation and the viscosity of rocks are caused by the slow motion of atoms over the potential hill (box 1-5) and are highly sensitive to temperature. Seismic wave attenuation is often measured by a Q-factor, which is defined by $Q^{-1} \equiv$ *[dissipated energy]/[stored energy],* That is, a low Q means high attenuation. Laboratory experiments show that in most cases, Q depends on temperature and frequency, as

$$Q^{-1} \propto \omega^{-\alpha} exp[-\frac{\alpha H_Q^*}{RT}], \qquad (B1\text{-}7\text{-}1)$$

where ω is the frequency, R is the gas constant, T is the temperature, α is a constant (~ 0.2–0.3) and H_Q^* is the activation enthalpy for attenuation. The rate of deformation (strain rate) in viscous deformation changes with temperature and stress (σ) as,

$$\dot{\varepsilon} \propto \sigma^n exp[-\frac{H_\eta^*}{RT}]. \qquad (B1\text{-}7\text{-}2)$$

Hence the viscosity, η ($= \sigma/\dot{\varepsilon}$), depends on temperature as

$$\eta \propto exp[\frac{H_\eta^*}{RT}] \qquad or \qquad \eta \propto exp[\frac{H_\eta^*}{nRT}] \qquad (B1\text{-}7\text{-}3)$$

for constant stress or a constant strain-rate, respectively, where H_η^* is the activation enthalpy for viscosity. From these equation, one has

$$Q \propto \eta^\beta, \qquad (B1\text{-}7\text{-}4)$$

with $\beta = \alpha H_Q^*/H_\eta^*$ for constant stress and $\beta = \alpha n H_Q^*/H_\eta^*$ for a constant strain rate. In most mantle materials, $H_Q^* \sim H_\eta^*$ and $\beta = 0.2$–0.9.

tor of 3 increase in Q corresponds to a $\sim 4 - \sim 200$-fold increase in viscosity. This roughly explains the correlation between the depth variation of attenuation shown in figure 1-11 and that of viscosity shown in figure 1-12.

Both attenuation and viscosity change with depth corresponding to the change in pressure and temperature (and crystal structure and possibly chemical composition) with depth. The pressure dependence of seismic

wave attenuation and viscosity can be expressed in terms of the pressure dependence of activation enthalpy (box 1-5),

$$H_{Q,\eta}^* = E_{Q,\eta}^* + PV_{Q,\eta}^*, \tag{1-6}$$

where $E_{Q,\eta}^*$ is activation energy, $VE_{Q,\eta}^*$ is activation volume (for attenuation and viscosity respectively), and P is pressure. Activation volume represents the degree to which the potential barrier for atomic motion increases with pressure, and it is usually positive. Thus an increase in pressure decreases attenuation and increases viscosity. The overall depth variation of rheological properties shown in figures 1-11 and 1-12 can be explained by the effects of temperature and pressure. At shallower parts, temperature rapidly increases, so viscosity decreases with depth. Going even deeper, the pressure effect becomes more significant, so viscosity starts to increase with depth. In this argument, however, the effect of phase transformation is not considered. Phase transformations may significantly modify rheological properties in some cases.

In this type of global-scale estimate of viscosity, smaller-scale variations of viscosity are ignored. However, small-scale viscosity variations may be important in some situations. An example may be found in the dynamics of the subduction process. The subducted oceanic plate probably has high viscosity because it is much colder than the surrounding mantle. In the subducted slab, however, viscosity may change in a complex manner because of successive phase transformations and the variations in temperature. This issue is closely related to the fate of subducted slabs and whole-mantle circulation, and I will discuss it in detail in chapter 4. Another example is mid-ocean ridge dynamics. Partial melting redistributes water in the upper mantle, and this may cause a change in viscosity (chap. 2).