**9.01 Earth Formation and Evolution**

D. J. Stevenson, California Institute of Technology, Pasadena, CA, USA

© 2007 Elsevier B.V. All rights reserved.

9.01.1 Introduction

9.01.1.1 How Should We Think of Earth and Earth Evolution?

Evolutionary science is for the most part based on observation and indirect inference. It is not experimental science, even though experiments can certainly play a role in our understanding of processes. We can never hope to have the resources to build our own planet and observe how it evolves; we cannot even hope (at least in the foreseeable future) to observe an ensemble of Earth-like planets elsewhere in the universe and at diverse stages of their evolution (though there is certainly much discussion about detection of such planets; e.g., Seager (2003)). There are two central ideas that govern our thinking about Earth and its history. One is 'provenance': the nature and origin of the material that went into making Earth. This is our cosmic heritage, one that we presumably share with neighboring terrestrial planets, and (to some uncertain extent) we share with the meteorites and the abundances of elements in the Sun. The other is 'process': Earth is an engine and its current structure is a consequence of those ongoing processes, expressed in the form it takes now. The most obvious and important of these processes is plate tectonics and the inextricably entwined process of mantle convection. However, this central evolutionary process cannot be separated from the nature of the atmosphere and ocean, the geochemical evolution of various parts of Earth expressed in the rock record, and life.

Figure 1 shows conceptually the ideas of Earth evolution, expressed as a curve in some multidimensional space that is here simplified by focusing on two variables ('this' and 'that'), the identities of which are not important. They could be physical variables such as temperature, or chemical variables (composition of a particular reservoir) or isotopic tracers. The figure intends to convey the idea that we have an initial condition, an evolutionary path, and a present state. The initial condition is dictated not only by provenance but also by the physics of the formation process. By analogy, we would say that the apples from an apple tree owe much of their nature not only to the genetics of apples (the process of their formation) but also, to some extent, the soil and climate in which the tree grew. We are informed of this initial condition by astronomy, which tells us about how planets form in other solar systems, by geochemistry (a memory within Earth of the materials and conditions of Earth formation), and by physical modeling: simulations and analysis of what may have occurred. Notably, we do not get information on the initial condition from geology since there are no rocks or landforms that date back to the earliest history of Earth. Geology, aided by geochemistry and geobiology, plays a central role informing us about Earth history. Though some geophysicists study evolution, nearly all geophysical techniques are directed toward understanding a snapshot of present Earth, or a very short period prior to present Earth, and it is only through modeling (e.g., of geological data) that the physical aspects of evolution are illuminated.

In Figure 2, another important idea is conveyed: for many purposes, we should think of time logarithmically. This is in striking contrast to the way many geoscientists think of time, because they focus (naturally enough) on where the record is best. As a result, far more geological investigations are carried out for
the Phanerozoic (10% of Earth history) than the entire period before this. More importantly, the processes that govern early history are very energetic and fast. As a consequence, more could have happened in the first millions to hundreds of millions of years than throughout all of subsequent geologic time.

Table 1 develops this idea further by identifying some of the important timescales of relevance to Earth history and prehistory (here taken to mean the important events that took place even before Earth formed). From this emerges the subdivision of geologic time into the accretion phase (the aggregation of bodies to make Earth), lasting a hundred million years at most, an early evolution in which the high-energy consequences of the accretion (the stored heat) and possibly later impacts still play a role, perhaps lasting as long as half a billion years, and the rest of geologic time in which the energetics of Earth is strongly affected by the long-lived radioactive heat sources.

In this overview chapter, an attempt is made to identify the main themes of Earth history, viewed geophysically, and to provide a context for appreciating the more detailed following chapters. At the end, some of the outstanding issues are revisited, reminding us that this is very much a living science in which there are many things not known or understood.

### 9.01.1.2 History and Themes

Hopkins (1839) in his “Preliminary observations on the refrigeration of the globe” illustrates well the prevailing

---

**Table 1** Some important timescales

<table>
<thead>
<tr>
<th>Process</th>
<th>Timescale</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formation of Earth</td>
<td>$10^7$–$10^8$ years</td>
<td>Infrequent large impacts; background flux of small impacts</td>
</tr>
<tr>
<td>Cooling of Earth after a giant impact</td>
<td>1000 years (deepest part of magma ocean)</td>
<td>A wide range of timescales, some of which are very fast but nonetheless important. Formation of a water ocean can be fast</td>
</tr>
<tr>
<td></td>
<td>100–1000 years (condensation of a silicate vapor atmosphere)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sim 10^6$ years (condensation of steam atmosphere, assuming no additional energy input)</td>
<td></td>
</tr>
<tr>
<td>Elimination of heat in excess of the thermal state that allows convective heat transport in equilibrium with radioactive heat production</td>
<td>$\sim 10^9$ years</td>
<td>Earth loses most of the thermal memory of a possible very hot beginning</td>
</tr>
<tr>
<td>Decline of impact flux</td>
<td>$(1–7) \times 10^8$ years</td>
<td>The late heavy bombardment at 3.8 Ga may have been a spike rather than part of a tail in the impact flux</td>
</tr>
<tr>
<td>Current timescale to cool mantle by 500 K</td>
<td>$5 \times 10^9$–$10^{10}$ years</td>
<td>Very slow because of the high mantle viscosity</td>
</tr>
</tbody>
</table>

---

**Figure 1** Conceptual view of Earth evolution, identifying the three crucial elements (the initial condition, the evolution path, and the present state) and the sciences that contribute to their understanding. The axes are unimportant, since the diagram is merely a 2-D slice of a multidimensional phase space. They might represent temperature or composition, for example.

**Figure 2** The logarithmic representation of geologic time. The energy budget and rapidity of processes at early times motivates this perspective. A similar view is often taken of cosmology.
view of that time when Earth started hot and was cooling over time. This hot beginning now seems natural to us as a consequence of the gravitational energy of Earth formation, and it has been a consistently popular view even when the justifications for its advocacy were imperfectly developed. Famously, Lord Kelvin (Figure 3) took the hot initial Earth and applied conduction theory to the outermost region to estimate the age of Earth at 100 My or less. Burchfield (1975) in his book, *Lord Kelvin and the Age of the Earth*, documents Kelvin’s various estimates and the conflict with the Victorian geologists of the time who believed that Kelvin’s estimates could not be sufficient to explain the features we see. Kelvin’s confidence was bolstered by the similar estimate he obtained for the age of the Sun. Indeed, astrophysicists refer to the ‘Kelvin time’ as a characteristic cooling timescale for a body, defined as the heat content divided by luminosity. We now know that Kelvin was wrong about the Sun because he was unaware of the additional (and dominant) energy source provided by fusion of hydrogen to helium. Ironically, Kelvin could have obtained a correct order of magnitude estimate for Earth’s age had he evaluated Earth’s Kelvin time. For a plausible estimate of mean internal temperature of \( \sim 2000 \text{ K} \) (a number that would have seemed perfectly reasonable to Kelvin), a heat capacity of 700 J kg\(^{-1}\), a mass of \( 6 \times 10^{24} \text{ kg} \) and an energy output of \( 4 \times 10^{13} \text{ W} \), he could have obtained

\[
\tau_{\text{Kelvin}} \approx \frac{(6 \times 10^{24}) \times (700) \times (2000)}{4 \times 10^{13}} \approx 2 \times 10^{17} \text{ s}
\]

It should not have been unreasonable for him to suppose that this was physically sensible since at that time the fluidity of Earth’s interior was still in doubt and the concept of efficient convective transport already existed. The subsequent discovery of fission and long-lived radioactive heat sources was ‘not’ the reason he got the wrong answer. Indeed, we now think that Earth could have a significant part of its heat outflow and dynamics even if those radioactive heat sources did not exist. Davies discusses this at far greater length in his chapter.

Plate tectonics and mantle convection is a central theme as the primary controlling principle of most of Earth evolution. Mobility of Earth was proposed by Wegener (1912) and the connection to deep-seated motions was also suggested long ago, for example, Bull (1921). The acceptance of these ideas was delayed, especially in the geophysical community, by the perceived absence of compelling evidence together with doubts about physical process: Could rocks flow in the way that was needed? So much has been written on this that any attempt to summarize briefly here would be superfluous. However, one aspect deserves comment – our current view of Earth evolution is not merely the physical picture of how Earth loses heat and thereby drives the plates; it is the profound interconnection of this to all the other aspects of Earth: (1) the nature of the atmosphere, (2) the existence and persistence of the hydrosphere, (3) the maintenance of the magnetic field, (4) the evolution of the continents, and (5) the evolution of life. Most importantly, these are coupled systems, not one where the mantle is dictating all else. For example, life influences the sediments on Earth, which in turn influence what is cycled back into the mantle, which in turn influences volcanism and plate motions. These interconnections are illustrated in Figure 4.

There are some other themes that are important and yet sometimes escape critical attention. The first is the central role of ‘common processes’ rather than ‘special processes’. In the early, perhaps more speculative, days...
of scientific theory-building, it was acceptable to invoke a special process for planet formation, including Earth, one that is rare for stars in general. Beginning in the 1950s, the Soviet school of cosmogony developed the view that planet formation is a natural process, building on much older Laplacian ideas of a nebula around the Sun, but placing that in the context of a universal concept of a disk around a forming star. Safronov (1972) played a central role in this development. Contemporaneously, the astronomical community developed a star formation scenario that naturally developed an attendant disk designed to handle the angular momentum budget of the originating interstellar cloud of gas and dust. Support for this picture grew from the 1980s onward because of astronomical observations. We now have abundant evidence for planets around Sun-like stars (of order 200 examples), though it is not yet clear how many of these systems possess terrestrial planets. This absence of evidence is not surprising given the insensitivity of the Doppler technique used to find most planets. Still, the prevailing view now is that there is nothing very special about how Earth came to be. Similarly, we are reluctant to suggest something very special about how Earth evolved, even though there is clearly a major difference between the nature of Earth and the nature of Venus, the most Earth-like of our neighbors in size, though not the most Earth-like in habitability.

Another important theme comes from ‘meteoritics’. It is widely accepted, yet not entirely obvious, that meteorites inform us about terrestrial planets and about the building blocks for Earth. Certainly we have a remarkable amount of information about meteorites: (1) their composition, (2) the timing of their formation, and (3) the conditions that they encountered (both physical and chemical). Chapter 9.02 provides much information on this. We must ask, however, whether the validity of this approach is as self-evident as it is sometimes portrayed or whether it is more an example of using what we have, as in the classic story of the drunk looking for his lost car keys under the lamppost because it is the only place where he can see. The relationship of materials in the Earth-forming zone to those that formed asteroids (meteorite bodies) remains poorly understood.

Also central to current ideas is the role of large impacts rather than dust or merely small bodies in the accumulation of Earth. Unlike the other themes listed here, this one is more theoretical, though it is consistent with the ‘clearing’ of dusty disks around stars, taken to infer the formation of planetesimals (though not necessarily requiring the formation of mostly large bodies). This theme originated in work in the 1970s (Hartmann and Davis, 1975; Cameron and Ward, 1976) motivated in large part by ideas of lunar formation, but also consistent with the Safronov model of planet formation as developed further by Wetherill (1976) in particular. Our current ideas of planet formation (e.g., Chambers, 2004) retain the feature, first noted by Wetherill, that the material that goes into making Earth comes from a wide range of heliocentric distances, but that there is nonetheless some expected variation in final composition simply because of the small number of large bodies that participate. This important idea is illustrated in Figure 5. It is not known whether this is in fact true; the test lies in a better understanding of the compositions (including isotopic make-up) of the other terrestrial planets.

9.01.2 Physical and Chemical Constraints

9.01.2.1 Important Ideas

We cannot figure out origin and early evolution except to the extent that Earth has a ‘memory’. The most obvious memory of a planet is its total mass: we do not know of any way of significantly modifying this after the planet forms. In respect of major elements, the planet is a closed system. The application of cosmochemistry and meteoritics to planets is heavily dependent on this idea of closed systems, especially when we speak of ‘provenance’ (the reservoir of material that was available for forming a planet). It is less obvious but true that a planet will not change its orbital radius significantly. Some other physical attributes such as spin or obliquity or orbital eccentricity can vary.
The results of four outcomes for the final stages of accretion of the terrestrial planets. The shaded bar graph represents an initial distribution of some tracer (e.g., oxygen isotopes or some compositional variable). The circles and their relative sizes give the final outcome in number spacing and relative masses of the resulting planets. The pie diagrams within each circle represent the relative amounts of material arising from each part of the initial distribution. This shows how substantial mixing is typical, that is, the material that ends up in Earth is not typically the material that was originally at 1 AU from the Sun. Reproduced from Chambers JE (2004) Planetary accretion in the inner solar system. Earth and Planetary Science Letters 223: 241–252, with permission from Elsevier.

Figure 5 The results of four outcomes for the final stages of accretion of the terrestrial planets. The shaded bar graph represents an initial distribution of some tracer (e.g., oxygen isotopes or some compositional variable). The circles and their relative sizes give the final outcome in number spacing and relative masses of the resulting planets. The pie diagrams within each circle represent the relative amounts of material arising from each part of the initial distribution. This shows how substantial mixing is typical, that is, the material that ends up in Earth is not typically the material that was originally at 1 AU from the Sun. Reproduced from Chambers JE (2004) Planetary accretion in the inner solar system. Earth and Planetary Science Letters 223: 241–252, with permission from Elsevier.

significantly (they are not conserved quantities). Aspects of the history of Earth rotation are in Chapter 9.10.

The retention of a memory is not guaranteed, especially for physical attributes, even when conservation principles apply. For example, the properties of an ice cube are independent of whether the water molecules were once in the vapor phase or once in a liquid form, at least if the material is in strict thermodynamic equilibrium. The water molecules have no memory. More subtle analysis would however discern whether the isotopic mix of the oxygen on the ice cube were similar or the same as some other ice cube. One could in this way decide whether the ice cube came from Mars or Earth.

Thermal memories can be elusive. For example, consider a body that cools according to this equation

$$\frac{dT}{dt} = -kT^{a+1}$$

where $T$ is temperature, $t$ is time, and $k$ is a constant. The exponent $a$ is assumed to be greater than unity, and in many realistic cases (e.g., mantle convection) it may be substantially larger than unity. The solution

$$T(t) = [a kt + T(0)]^{-1/a} \approx (akt)^{-1/a}$$

retains poor memory of its initial condition if $a > 1$ and substantial cooling has taken place, that is, $T(t)$ is substantially smaller than $T(0)$. This is a common situation in planets. The loss of memory can be even more striking if there is a long-lived heat source (radioactivity). Examples of this are displayed in the thermal histories discussed in Chapter 9.08. By contrast, a planet that differentiates (e.g., into core and mantle) may retain some memory of initial thermal state should the deeper reservoir (e.g., the core) be considerably hotter than the near-surface reservoir. This would appear to be the case for Earth’s core.

Plants are not in thermodynamical equilibrium; that is, one part of the planet is often prevented (by finite diffusivity) from equilibrating with other parts, especially if at least one of the reservoirs is solid. This is a central tenet of many aspects of geochemistry. For this reason, chemical reservoirs are a very important source of memory.

### 9.01.2.2 Some Useful Estimates

It is useful to have an appreciation of various energy budgets, timescales, and dimensionless numbers. Among terrestrial planets, solar energy is large compared with the energy normally available from the interior of a planet. For example, the incident sunlight on Earth exceeds the current heat flow from Earth’s interior by a factor of 5000. This means that the thermal state of a planet surface is determined by the Sun. However, nonsolar energy (e.g., accretion) can be comparable or more important for short periods of time during the planet formation phase.

The total time of planet accumulation can be long compared to free fall times because the starting material is so widely dispersed. This initial dilution is due in turn to the requirements of angular momentum conservation. For example, suppose we allowed the mass of Earth to be initially dispersed in a volume that is of order one-tenth of a cubic astronomical unit, that is, 0.1 AU$^3$ (where 1 AU is the Earth–Sun separation). Then if an embryo Earth were plowing through this material with a relative velocity of 1 km s$^{-1}$ (a significant fraction of orbital velocity), it would grow at a rate given by

$$4\pi \rho_o R^2 \frac{dR}{dt} = \pi R^2 v \rho$$

where $\rho_o$ is the density of solid matter, $\rho \sim 10 M_\oplus/(1 \text{ AU})^3$ is 12 orders of magnitude smaller, $v \sim 1$ km s$^{-1}$, $R$ is the radius, and a geometrical cross section is assumed. This predicts $dR/dt$ of a few centimeters...
per year and a planet formation time of order 100 Ma. Notice that this is independent of whether the accreting material is in the form of dust or larger bodies.

The formation of a planet takes dispersed matter and aggregates it into a mass $M$ of radius $R$. As a consequence, the gravitational energy changes from a small value to about $-GM^2/R$. By the first law of thermodynamics, this energy release must go somewhere. A small part at most is consumed breaking up material but most of this energy is converted into heat. If all that energy were stored internally then the temperature rise is

$$
\Delta T \sim \frac{GM}{RC_p} \sim 40000 \left( \frac{R}{R_\odot} \right)^2
$$

where $C_p$ is the specific heat of terrestrial materials and $R_\odot$ is Earth radius. For the larger bodies, this is so large that some of the rock and iron is converted into vapor; indeed, $GM/RL_v \sim 1$ at Earth mass, where $L_v$ is the latent heat of vaporization.

If a planet is initially undifferentiated and then separates into core and mantle, then the resulting energy released as heat (from the resulting larger but more negative total gravitational energy) is evidently some fraction of this amount. For a core that is one-third of the mass but twice the density of rock and settles an average of one-half the planetary radius, the resulting temperature increase is reduced from eqn [5] by a factor of roughly 10, implying a large effect (thousands of degrees potentially) for Earth. As with accretional heating, this energy may not be uniformly distributed internally, but unlike that in eqn [5] it is not readily radiated to space.

Suppose we were to radiate away all of the energy of accretion in time $t$ at a temperature of $T_e$ (ignoring the effect of the Sun or local nebula environment) as infrared (IR) black body radiation. Accordingly,

$$
4\pi R^2 \sigma T_e^4 t \sim \frac{GM^2}{R}
$$

$$
\Rightarrow \quad T_e \sim 350 \left( \frac{10^9 \text{ yr}}{t} \right)^{1/4} \left( \frac{R}{R_\odot} \right)^{3/4}
$$

This equation requires careful interpretation. If it were indeed true that Earth formed from small particles (e.g., dust) over 10 or 100 My, then it would not get hot – the energy of accretion is similar to delivered sunlight (at current solar luminosity) over that period. However, the delivery of mass is highly nonuniform and much of the mass is delivered in giant impacts. This same equation would say that if you had to radiate away one-tenth of Earth’s energy of formation (the energy arising from impact with a Mars-sized impacting body) at $T \sim 2000 \text{ K}$ appropriate to a silicate vapor atmosphere, then it would take $\sim 1000$ years. This is far longer than the delivery time of the energy in a giant impact ($\sim$ a day or less) implying that very high temperatures are unavoidable.

The early state of the planet depends on the partitioning of energy input between surficial (available for prompt radiative loss) and deep seated (available for storage and only eliminated if there is efficient heat transfer from the interior). A crude but useful way to estimate this relies on the introduction of a parameter $f$, the fraction of input energy that is delivered to the deep interior. In the simple case of little energy transport from interior to surface, eqn [5] might then be replaced by

$$
\Delta T \approx \frac{fGM(t)}{R(t)}
$$

where $\Delta T$ is the temperature difference between the near-surface interior and surface at the time when the planet has mass $M(t)$ and radius $R(t)$. It is common practice to think of $f \sim 0.1$ or even less, but the physical basis for this choice is unclear since the heat loss will be high if eqn [7] predicts a high temperature. In other words, $f$ is a function, not a number. However, the prevailing view of the initial state of Earth is that it is set by a giant impact, presumably the impact that made our Moon. For this view, eqn [7] is not useful and one must instead appeal to the outcome of impact simulations, for example, Canup (2004). These show that temperature increases of many thousands of degrees are likely, though the temperature distribution is very heterogeneous. This is discussed further in Chapter 9.03 and also figures prominently in the initial condition for the discussion of the resulting magma ocean in Chapter 9.04.

An enormous range of heat fluxes $F$ from Earth’s interior is possible, both in the accretion epoch and subsequently: in fully molten medium, convective transport can be enormous if the medium is even only slightly superadiabatic. For a superadiabatic temperature difference $\delta T$, convective length scale $L$, and coefficient of thermal expansion $\alpha$, mixing length theory (see Chapter 9.04) predicts $F \sim \rho C_v \delta T \left(g \alpha \delta TL\right)^{1/2}$, assuming viscosity is small enough to be unimportant. For layer of thickness $L$, the time to cool by $\Delta T$ is accordingly

$$
\tau_{\text{cool}} \sim 10 \left( \frac{1000}{\Delta T} \right)^{1/2} \left( \frac{\Delta T}{\delta T} \right)^{3/2} \left( \frac{L}{g} \right)^{1/2}
$$
Since \((L/g)^{1/2}\) is a short timescale (e.g., hours), we see immediately that the cooling time is short, even for very small temperature fluctuations driving the convection (e.g., of order one degree).

By contrast, cooling times in a system that is mostly solid (and therefore very viscous) can be enormously longer. The corresponding formula for heat flux in this case is \(F \sim 0.1 \rho c_0 \beta T (g \alpha \delta T e^{2/\nu})^{1/3}\), where \(\kappa\) is the thermal diffusivity and \(\nu\) is the kinematic viscosity (see Chapter 9.08). For the usually realistic choice of \(\kappa \sim 0.01 \text{ cm}^2 \text{s}^{-1}\), this predicts a cooling time,

\[
\tau_{\text{cool}} \sim 10^{12} \text{yr} \left( \frac{L}{1000 \text{ km}} \right) \left( \frac{\Delta T}{1000} \right)^{4/3} \left( \frac{100}{\delta T} \right)^{1/3} \left( \frac{\nu}{10^{20} \text{ cm}^2 \text{s}^{-1}} \right)
\]

where the result has been scaled to a choice of viscosity that is roughly like that expected for silicate material near its melting point. The precise value is not important because the real significance of this result lies not in the specific numbers predicted but in the profound difference between cooling of a liquid and cooling of a solid. A planet that melts can lose heat prodigiously, but when Earth is mostly solid, the cooling rate is far slower. One consequence of this is that the time that elapses between the very hot Earth and the formation of a water ocean can be small. Evidence of an early ocean is discussed in Chapter 9.05 and also figures in the discussion by Stein and Ben-Avraham on the origin and evolution of continents. It may also be relevant to the initiation of plate tectonics (see Chapter 9.06).

We also note that

\[
\frac{GM^2}{R} \int_0^\infty Q_{\text{radio}}(r) \, dr \sim 10 \left( \frac{R}{R_e} \right)^2
\]

where \(Q_{\text{radio}}\) is the chondritic radiogenic heat production, excluding short-lived sources \((\alpha\text{Al}, \alpha\text{Fe})\). This emphasizes the importance of gravity as setting the stage for subsequent evolution on Earth.

Work done breaking up materials is small because the strength of materials is small compared to \(GM^2/R^4\), the typical energy per unit volume associated with gravity. On the other hand, \(GM^2/R^4 K \sim (R/R_e)^2\) where \(K\) is the bulk modulus. This means that for Earth mass planets, gravitational self-compression leads to a higher density (smaller radius) than a body comprising the same constituents but zero internal pressure. This effect is enough to affect significantly the mineral assemblage within larger terrestrial planets, thereby affecting melting behavior, differentiation, core properties, and possible mantle layering. It also means that Earth’s interior can be heated by adiabatic compression alone.

The biggest effect of adiabatic compression is in a possible massive atmosphere, since gas is highly compressible. As a consequence, the radiative temperature of a planet can be much less than the surface temperature of the planet even when the atmospheric mass is a small fraction of the total mass. This assumes that it is opaque (e.g., as in the greenhouse effect, but the argument provided here is not limited to that effect). For example, an adiabatic atmosphere that radiates at a pressure \(P_e\) at temperature \(T_e\) and has a basal pressure \(P_\ast \gg P_e\) will have a surface temperature \(T_s\) given by

\[
T_s \sim T_e \left[ 10^{12} \left( \frac{R}{R_e} \right)^2 \left( \frac{1 \text{ bar}}{P_e} \right) \left( \frac{M_{\text{atm}}}{M} \right) \right]^{2/3}
\]

where \(T \propto P^\gamma\) is the adiabatic relationship and \(M_{\text{atm}}\) is the atmospheric mass. The factor of \(10^{12}\) demonstrates the remarkable blanketing effect that is possible even for an atmospheric mass that is less than the planet mass by a factor of a million.

Planetary embryos form early enough that they are imbedded in the solar nebula. Gravitational attraction increases the nebula gas density near the embryo surface. For an isothermal atmosphere of negligible mass,

\[
\rho(r = R) / \rho(r \to \infty) = \exp \left[ \frac{GM}{R^2} \right]
\]

where \(c\) is the speed of sound for the nebula (primarily hydrogen). Since the nebula is very low density, this is only of interest for surface conditions (e.g., ingassing of nebula at the surface of a magma ocean) if \(GM/Rc^2\) is larger than \(\sim 5\) or \(10\). For a Mars mass and \(T \sim 300\text{ K}, GM/Rc^2 \sim 12\). For such an embryo, the nebula might be \(\rho \sim 10^{-9} \text{ g cm}^{-3}\) and the surface density could then be \(\rho \sim 10^{-4} \text{ g cm}^{-3}\), potentially optically thick, and with a surface pressure (\(~1\text{ bar with a large uncertainty}) that allows modest ingassing. Thus, embryos larger than Mars, including the growing Earth, could have had a massive atmosphere of near-solar composition. However, the planetary evidence suggests that this effect is small. For example, the neon-to-argon ratio for Earth is much lesser than the solar nebula ratio even though the ingassing ability (predicted by Henry’s law) is similar. This is discussed further in Chapter 9.05. Presumably, the nebula was eliminated early in the period of growth of large embryos.
9.01.3 Commentary on Formation Models

Astronomical observations of newly forming stars indicate the presence of disks of gas and dust. The material in the disk is typically a few percent of a solar mass, more than sufficient to explain the observed planetary mass of our system or other systems discovered thus far. If the disk has solar composition, then the amount of condensable material (as rock or iron) internal to a few astronomical units is sufficient to explain the masses of the terrestrial planets. However, the disks extend to tens of astronomical units or more, and this is enforced by the angular momentum budget of the originating interstellar cloud. This angular momentum is responsible for the possibility of planet formation but also guarantees that the terrestrial zone is a small part of a much bigger picture. This means that we probably cannot understand formation of terrestrial planets without some understanding of the formation of the gas giants, especially Jupiter. The disks have a radial variation in temperature, wholly or partly because of the energy release of the central body (the forming star). As a consequence, the region within a few astronomical units of the star is hundreds of degrees kelvin or more, sufficient to avoid the condensation of water ice. The absence of large amounts of water in the terrestrial zone in our solar system is interpreted as evidence for terrestrial planet formation internal to the 'snow line' (the place outward of which water ice can condense). Typically, this temperature is of order 160 K (low compared to 270 K because of the very low vapor pressures characteristic of such ice). Chapter 9.05 discusses the source of water on Earth, which is presumably external to the terrestrial planetary zone.

The central ideas in current models of terrestrial planet accretion are three: (1) planetesimal formation, (2) runaway and oligarchic growth, and (3) late stage accumulation. Planetesimal formation refers to the accumulation of bodies of order kilometers in size. Runaway and oligarchic growth are two dynamical stages of a process that we can (for our purpose) lump into the rapid-accumulation planetary embryos of up to order Mars in mass but in closely spaced low eccentricity and inclination orbits. The last and slowest stage proceeds for hundreds of Moon- and Mars-sized objects to the terrestrial planets that we see. According to astronomical observations, this last stage is probably occurring after the solar nebula has been removed (by accretion onto the Sun or expulsion to the interstellar medium). Chapter 9.02 connects this physical picture to the cosmochemical evidence.

Planetesimal formation remains mysterious even though there is no doubt that it occurred, since its consequences are expressed among the meteorites arising from differentiated asteroids that must have formed in the first million years after the formation of the solar nebula. They may have formed by the poorly understood process of sticking of dust particles during very slow velocity collisions in the dusty gas. They may alternatively have formed by a gravitationally mediated process as the dust settled toward the mid-plane of the disk. Gravitational instabilities are an attractive mechanism, especially given the uncertainties of the physics of sticking, but the relevant fluid dynamics is still debated (Weidenschilling, 2006). This remains one of the major puzzles of planet formation.

The next stage, from planetesimals to Moon and Mars mass embryos, is perhaps better understood theoretically but less well founded in direct observation. This stage, lasting $10^5$–$10^6$ years, is rapid because of gravitational focusing. Bodies encounter each other at velocities considerably less than the escape velocity from the larger body, and, as a consequence, the cross section for collision is far larger than the geometric cross section, perhaps by as much as a factor of a thousand. The planetesimal swarm is 'cold' (i.e., the random velocities of the bodies are very small compared to Keplerian orbital velocities.) This process of embryo growth is terminated by isolation: when the orbits are nearly circular, the bodies reach a stage where there are no crossing orbits (i.e., no overlap of their gravitational spheres of influence). From the point of view of understanding the bodies that we see, this stage is of great interest in three ways. (1) These embryos form quickly and therefore may be hot, both because of possible short-lived radioactive elements and also because of accretional heating (cf. the scaling discussed in Section 9.01.2). They could therefore form cores and primordial crusts. In this sense, the primordial crusts and current cores of the terrestrial planets have the potential to predate the formation of the planets. (2) The largest of these embryos could conceivably be a surviving planet, most plausibly Mars. In this picture, Mars is special as an isolated outlier in the formation process. (3) Embryos form while the solar nebula is still present and could therefore have
component of solar composition gas. Equation \[12\] suggests that this, however, is small.

The late-stage aggregation is conceptually like that envisaged long ago in the work of Safronov (1972), whereby the scattering of bodies causes growth of eccentricity and inclination allowing the orbits to cross. It has received a lot of attention in recent times because of the development of \(N\)-body codes that can handle the outcome of a population of bodies scattering from one another gravitationally and occasionally colliding (Chambers, 2004). The main shortcoming of these calculations is the failure to take full account of what actually happens when two bodies have a very close encounter (a quite common occurrence) or collide. Tidal disruption is possible in a close encounter (i.e., the outcome is not necessarily the two intact bodies that existed before encounter) and collisions do not always lead to a clean merging of the two bodies. Close encounters can also create additional bodies through tidal disruption.

It is easy to see by order of magnitude (a slightly more sophisticated version of eqn \[4\]) that this late-stage process can take as long as 100 Ma. However, it is stochastic and it involves large bodies. As a consequence, it is not possible (and may perhaps never be possible) to say exactly what sequence of major events took place in the formation of a particular planet.

**9.01.4 Commentary on Early Evolution Models**

The high-energy events of late-stage accumulation are expected to play a central role in setting the stage for Earth structure and evolution. The building blocks are a set of planetary embryos that almost certainly form iron-rich cores because of the combined effects of gravitational energy release in a relatively short time and the possible effects of \(^{26}\)Al heating. This early but important differentiation event is characterized by relatively modest temperatures (\(\sim 2000 \text{ K}\) or less) and pressures (20 GPa or less, perhaps a lot less) appropriate to bodies in this size range. It is likely and perhaps significant for some constituents (e.g., noble gases) that this phase occurs in the presence of the solar nebula. In highly idealized numerical models (Chambers, 2004), it is usually assumed that nearly all of Earth’s mass accumulates in the subsequent merging of these massive embryos. In reality, it is not known how much material arrives in small bodies (planeteimals), whose origin could be either the bodies that were not swept up during runaway, or debris created during frequent close encounters in the final orbit-crossing phase. Giant impacts will certainly create extensive melting and at least a transient magma ocean, especially if one assumes that the interiors of the embryos prior to impact are at or even above the solidus of mantle minerals. This is a reasonable assumption because of the limited time available for elimination of earlier heating, together with the higher radioactive heating of that epoch. It is also possible that there is a persistent magma ocean, even in the long intervals of time (\(\sim 10^7\) years) between the giant impacts, sheltered by a steam atmosphere greenhouse that is sustained by delivery of a ‘rain’ of small bodies that heat its base. This kind of magma ocean is less certain than the transient high-energy ocean that is present immediately after a giant impact. These magma ocean scenarios are analyzed in detail in Chapter 9.04 and also figure prominently in the discussion of core formation in Chapter 9.03.

Of course, short-lived high-energy events could have a bigger role than sustained lower-energy events in determining the composition of current core and mantle because they can create extensive melting in which core–mantle separation is especially efficient. There are two central questions that arise in this picture:

1. To what extent is Earth’s core formation accomplished by merging the cores of embryos rather than by separation of iron alloy from the mantle within the Earth?
2. To what extent is the separation of core from mantle and possible internal differentiation of mantle accomplished in a magma ocean environment?

These issues figure prominently in Chapter 9.03.

In simulations designed to understand the origin of Earth’s Moon, smoothed particle hydrodynamics (SPH) is most commonly used to describe the outcome of an impact involving a Mars-sized projectile and a mostly formed Earth (Canup, 2004). This represents perhaps the last (and perhaps most important) giant impact. Core and mantle are tracked and one can follow the extent to which these constituent parts of projectile and target ‘mix’. The outcome is essentially stochastic; that is, it varies considerably as a function of details, such as impact parameter and mass ratio of projectile to target, parameters that we can never hope to establish deterministically. However, in typical simulations, only around 10% of the projectile (and thus at most a few percent of an
Earth mass) is placed in orbit and available for making the Moon. Very little of Earth is placed into orbit. Nearly all of the rest of the projectile either merges immediately with Earth or crashes back into Earth within several hours of the initial impact. The iron core of the projectile may be stretched into filaments or broken into blobs. These simulations are low resolution, so even the finest scale features of SPH are hundreds of kilometers in size. It seems likely that much of the iron is emulsified to small scales and equilibrates with the mantle, but this is not certain.

The transitional phase between a mostly molten mantle and an almost entirely solid mantle is still poorly understood. In respect of thermal history, it could be argued that the details of this phase are unimportant (cf. Chapter 9.08). In respect of the geochemical evolution, it is probably very important for setting up the conditions for formation or preservation of the earliest crust (see Chapter 9.02).

The initially hot core will be allowed to cool by the overlying mantle and will therefore convect, allowing for the possibility of magnetic field generation. Nimmo deals with core evolution in Chapter 9.09, particularly the puzzle of whether there is enough energy in core cooling and inner growth to explain the energy budget demanded by the persistence of a dynamo throughout nearly all of geologic time. The answer to this puzzle remains uncertain.

What about the origin of plate tectonics? On the one hand, this could be viewed as a geologic question, provided there was agreement about the distinctive signatures of plate tectonics in the rock record. On the other hand, this could be viewed as a fluid dynamical question (albeit one that is unavoidably coupled to the crustal evolution and rheological puzzle of a fragmenting lithosphere). Sleep discusses both aspects of this unresolved question in Chapter 9.06.

Finally we come to life. The origin of life on Earth remains one of the central scientific problems of our time, even though many in the scientific community have adopted the view that this origin was ‘easy’, by which they mean that the physical conditions, amounts of material, energy budget, and abundance of liquid water and surfaces allows for conditions that are suitable. Less controversial but perhaps as important, it is abundantly clear that biological evolution has affected the physical aspects of Earth evolution profoundly, just as the physical conditions have mediated that biological evolution. In many ways this geobiological question is the most exciting frontier of Earth science, along with the more immediate issues of climate change (Earth evolution on a much smaller timescale).

### 9.01.5 Outstanding Questions

What are the main issues that arise in our understanding of Earth origin and evolution? We end by providing below a list of 10 questions, linked to the subsequent chapters in this volume. In some cases, these emerge naturally from the perspectives of the various authors while in other cases they show the interconnections that might be less evident from a single chapter.

1. **What is the legacy of Earth formation in the planet as we see it?** What range of accretion scenarios is consistent with Earth? What is the role of chance? What is the role of distance from the Sun? How does geochronology compare with physical estimates of formation times? (see Chapter 9.02)

2. **How is core formation in Earth expressed in the composition of Earth now and the thermal structure and evolution?** Can we figure out the core formation conditions from the abundances of siderophiles in the mantle? Or will we find that the pattern is a mix of low-temperature, low-pressure and high-temperature, high-pressure events? (see Chapter 9.03)

3. **What was the duration of the early magma ocean phase of Earth and what memory (if any) is there of this phase in the early mantle differentiation or crust formation?** Is there an opportunity for melt–solid separation in the early mantle and can this be reconciled with current models of mantle structure and composition? What is the role of an early dense atmosphere? How long did it take for a magma layer within Earth (e.g. in the transition zone) to completely freeze? (see Chapter 9.04)

4. **Where did Earth’s water come from?** How much do we have and how much of it has cycled between mantle and hydrosphere over geologic time? Did water play an essential role in the onset of plate tectonics? (see Chapter 9.05)

5. **Why does Earth have plate tectonics?** Mantle convection is easily understood; plate tectonics still resists understanding because of our imperfect knowledge of how the lithosphere fails. The perspective of geologic time helps us solve this puzzle. (see Chapter 9.06)

6. **Why continents? When continents?** Are continents a sideshow in the overall dynamics of Earth evolution (as passengers on plates) or do they play a central role in how the system evolves? (see Chapter 9.07)
7. How did Earth’s mantle temperature evolve through time? Can we reconcile simple convective models with the radiogenic heat supply, the geologic evidence, and our understanding of plate tectonics? The mismatch between known heat generation from long-lived radioactivity and the actual heat flow out of Earth persists and is incompletely resolved in simple cooling models, suggesting the evolution is more complex. (see Chapter 9.08)

8. Can we reconcile Earth thermal evolution with the persistence of the geomagnetic field throughout geologic time? When did the inner core form? The mantle governs the rate at which the core cools and models acceptable on the mantle side suggest a possible problem for core heat flows, including the possibility that the inner core was absent for a major part of early Earth history. But if so, can we supply enough energy to run Earth’s magnetic field? (see Chapter 9.09)

9. What do variations in Earth rotation tell us about the nature of Earth and its evolution? Have we made sufficient use of rotation history, including true polar wander, in our attempts to reconstruct how Earth evolved? (see Chapter 9.10)

10. How are the geophysical and biological evolutions of Earth interrelated? What does the history of life tell us about plate tectonics and the history of water on Earth? To what extent is the co-evolution of life and Earth’s physical attributes so strongly coupled that we connect major events in biological evolution to major events in geological history? (Chapter 9.11)

References