UREY RATIO AND THE STRUCTURE AND EVOLUTION OF EARTH’S MANTLE

Jun Korenaga

Received 15 August 2007; accepted 8 January 2008; published 12 June 2008.

[1] The Urey ratio describes the contribution of internal heat production to planetary-scale energy balance, and knowing this thermal budget for Earth is essential to understand its long-term evolution. Internal heat production is provided by the decay of radiogenic elements, whose budget is constrained by geochemical models of Earth. Understanding the thermal budget thus requires contributions from both geophysics and geochemistry. The purpose of this review is to elucidate various geochemical and geophysical arguments and to delineate the most likely thermal budget of Earth. While the bulk Earth Urey ratio is probably ~0.35, the convective Urey ratio is estimated to be ~0.2. That is, only ~20% of convective heat flux seems to originate from radiogenic elements at present, so the rest should be supported by secular cooling. A likely scenario for Earth’s thermal history indicates that the peculiarity of today’s significant imbalance between heat production and heat loss may result from the initiation of plate tectonics in the early Earth.


1. INTRODUCTION

[2] The interior of Earth is hotter than its surface. This temperature difference results in the heat flux from Earth’s surface to space, which has been estimated to be about 40 TW at the present day [Sclater et al., 1980; Pollack et al., 1993; Jaupart et al., 2007]. Loss of internal energy is balanced primarily by (1) heat production from radiogenic elements and (2) a decrease in the primordial heat content of Earth (i.e., secular cooling). (Italicized terms are defined in the glossary, after the main text.) Other energy sources such as tidal dissipation are negligible compared to these two processes [e.g., Verhoogen, 1980]. To be significant for Earth’s thermal history, radiogenic elements must have long half-lives as well as sufficient abundance or large decay energy. Short-lived radiogenic nuclides such as $^{26}$Al are important only for the very early Earth (within the period of the first 10 Ma or so) [e.g., Urey, 1955], and we can limit ourselves to the following four isotopes: $^{238}$U, $^{235}$U, $^{232}$Th, and $^{40}$K, which decay as

$$^{238}\text{U} \rightarrow 206 \text{Pb} + 8^4\text{He} + 6e^- + 6\bar{\nu}_e + 51.7 \text{MeV,} \quad (1)$$

$$^{235}\text{U} \rightarrow 207 \text{Pb} + 7^4\text{He} + 4e^- + 4\bar{\nu}_e + 46.0 \text{MeV,} \quad (2)$$

$$^{232}\text{Th} \rightarrow 208 \text{Pb} + 6^4\text{He} + 4e^- + 4\bar{\nu}_e + 42.7 \text{MeV,} \quad (3)$$

$$^{40}\text{K} \rightarrow 40 \text{Ca} + e^- + \bar{\nu}_e + 1.31 \text{MeV,} \quad (4)$$

$$^{40}\text{K} + e^- \rightarrow 40 \text{Ar} + \nu_e + 1.51 \text{MeV,} \quad (5)$$

where $e^-$, $\nu_e$, and $\bar{\nu}_e$ denote an electron, an electron neutrino, and an electron antineutrino, respectively. At present, $^{238}$U/U = 0.9928, $^{235}$U/U = 0.0072, all thorium is $^{232}$Th, and only 0.0128% of potassium is $^{40}$K. The half-lives of $^{238}$U, $^{235}$U, $^{232}$Th, and $^{40}$K are $4.468 \times 10^9$ years, $7.038 \times 10^8$ years, $1.405 \times 10^{10}$ years, and $1.277 \times 10^9$ years, respectively; 89.3% of the $^{40}$K decay is by beta decay (equation (4)) with the rest by electron capture (equation (5)). Geochemical models for Earth’s composition suggest that the present-day internal heat production is about 20 TW [e.g., McDonough and Sun, 1995; Allègre et al., 2001; Lyubetskaya and Korenaga, 2007a], so the rest of the surface heat flux must be from secular cooling. At least at the present day therefore internal heat production and secular cooling must be equally important, and Earth’s interior must have been hotter in the past.

[3] This is a sketch of the present-day thermal budget of Earth, but going beyond this rough description is not a trivial task because the thermal budget is inseparably
intevelops the structure and evolution of Earth's mantle, which has been perhaps the most controversial subject in solid Earth sciences for the last few decades [e.g., Ringwood, 1975; Jeanloz and Richter, 1979; Davies, 1981; Spohn and Schubert, 1982; Hager et al., 1985; Christensen, 1985a; Silver et al., 1988; Anderson, 1989; Fukao et al., 1992; Tackley et al., 1993; Allègre et al., 1996; van der Hilst et al., 1997; Hofmann, 1997; Davaille, 1999; Kellogg et al., 1999; Tackley, 2000a; Helffrich and Wood, 2001; Korenaga, 2003; Boyet and Carlson, 2005; Lyubetskaya and Korenaga, 2007b]. The problem is that geophysics and geochemistry appear to indicate conflicting pictures for the structure of mantle convection, and as a result, a variety of convection models have been proposed to reconcile both geophysical and geochemical observations [e.g., Kellogg et al., 1999; Becker et al., 1999; Albarède and van der Hilst, 2002; Bercovici and Karato, 2003]. Before attempting to reconcile different types of observations, however, it is important to evaluate the strength of each observational constraint and the validity of any underlying assumption; otherwise, a resulting model might become unnecessarily convoluted. Unfortunately, this critical step seems to have been overlooked in the past, and conflicts between geophysics and geochemistry tend to be taken at face value. Moreover, there have been some conceptual misunderstanding and semantic confusion regarding the thermal budget, which may have aggravated such interdisciplinary conflicts. The purposes of this paper are therefore to review both geochemical and geophysical observations pertinent to the thermal budget of Earth, with an emphasis on their uncertainty and limitation, and to discuss an emerging model of mantle convection along with its implications for the thermal and chemical evolution of Earth.

[4] The Urey ratio is the ratio of internal heat production to surface heat flux. It is a key parameter for a planetary thermal budget, but because geophysicists and geochemists have defined this ratio in different ways and are often unaware that they have different definitions, this nondimensional number has also been a good source of confusion. In order to facilitate discussion, it is convenient to define two different Urey ratios: the bulk Earth (BE) Urey ratio and the convective Urey ratio (Figure 1). The BE Urey ratio denotes the contribution of internal heat generation in the entire Earth to the total surface heat flux. The Urey ratio in the geochemical literature is usually this BE Urey ratio. Conversely, the convective Urey ratio is the ratio of internal heat generation in the mantle over the mantle heat flux (Figure 1, see also equation (8) in section 3.2). The contribution of continental crust is subtracted from both the internal heating budget and the surface heat flux because the continental crust does not directly participate in convective heat transfer. This is the Urey ratio commonly implied by geophysicists when they discuss the thermal history of Earth. Actually, the convective Urey ratio is the original Urey ratio, because the term “Urey ratio” was first introduced by Christensen [1985a] for his parameterized convection models. Note that some authors used a reciprocal definition (i.e., the ratio of heat flux over internal heat generation) [e.g., Hart and Zindler, 1986; Richter, 1984; Davies, 1993]. Also note that there is a similar term called “the internal heating ratio,” which has frequently been used in the study of mantle convection, but this ratio is used in a very different context and should not be confused with the Urey ratio (see section 5.1).

[5] The convective Urey ratio as defined above implicitly assumes whole mantle convection; that is, the entire mantle convects as a single layer (or one could say that the physical interpretation of the Urey ratio is straightforward for whole mantle convection). Earth’s mantle, however, could be internally layered (Figure 2a), and if this is the case, describing the thermal budget would be more involved. There have been a variety of geochemical and geophysical arguments for and against layered-mantle convection, and even after a few decades of debates, the style of mantle convection has not yet been settled. In fact, given the likely resolution and nonuniqueness of our observational constraints, we may never be able to settle this issue completely. Therefore, this review article chooses to adopt Occam’s razor to organize discussions. Whole mantle convection is regarded as the simplest hypothesis, which would be abandoned only if there is a strong argument against it. This is indeed the way many geochemists have argued for layered-mantle convection, i.e., by presenting paradoxes caused by assuming whole mantle convection. A variety of geochemical arguments are thus reviewed first in section 2. Geophysical arguments for layered convection are then reviewed in section 3. If we appreciate what assumptions are involved and how uncertain our observational constraints and theoretical predictions are, these geochemical and geophysical arguments for layered-mantle convection do not seem to be as strong as commonly believed. Note that the choice of whole mantle convection as the preferred working hypothesis is not arbitrary; it is required to estimate the thermal budget in a self-consistent manner (section 2.1.3). Finally, a summary on the present-day thermal budget is given in section 4, and a few remaining issues relevant to the thermal budget are discussed in section 5.

2. GEOCHEMICAL PERSPECTIVES: GLOBAL MASS BALANCE AND ISOLATED RESERVOIRS

[6] There are mainly two different kinds of geochemical arguments against whole mantle convection. One is based on the claimed inability to make a self-consistent global mass balance with the assumption of whole mantle convection, and the other is the presence of distinct geochemical reservoirs, which seem to have been isolated for the timescale of a few billion years. The former calls for a hidden reservoir (i.e., the lower convecting layer in Figure 2a) by a proof by contradiction, and the size of the hidden reservoir may be estimated by global mass balance (though this estimate may be problematic; see section 2.1.3). The latter argument does not yield a tight size constraint and involves more assumptions such as the origin of hot spots and the rate of mantle mixing. The isolated reservoir argument is thus usually presented in conjunction with the
These geochemical arguments have always imposed the need of chemical stratification on mantle convection models, and because of this, unified convection models, i.e., models that aim to satisfy both geophysical and geochemical observations, have become more and more complex. The classical model of layered-mantle convection was simple and well defined. Earth’s mantle has two major seismological discontinuities at the depths of 410 and 660 km, and the 660-km discontinuity was thought to represent an impermeable barrier. Although both discontinuities can be explained reasonably well by phase changes (in a chemically homogeneous mantle) [e.g., Ringwood, 1975; Ita and Stixrude, 1992] and do not require compositional stratification, the mass of a missing reservoir inferred from global mass balance happened to roughly correspond to that of the lower mantle (i.e., the mantle below the 660-km discontinuity) [e.g., Jacobsen and Wasserburg, 1979; O’Nions et al., 1979]. Thus, it was considered that the chemical stratification inferred from geochemistry also corresponded to a seismic discontinuity. As seismological evidence for slab penetration into the lower mantle accumulated [e.g., van der Hilst et al., 1997], however, it became difficult to postulate the coincidence of the impermeable boundary with the 660-km discontinuity, and accordingly, the chemical stratification has been decoupled from the seismological structure. Recent variants of layered-mantle convection thus need to push down the boundary below the discontinuity and assume a chemically distinct layer occupying a deeper portion of the lower mantle [e.g., Kellogg et al., 1999]. A problem is that there is no global seismic discontinuity known in the middle of the lower mantle. While it may be possible to explain the invisibility of this presumed deep layer by, for example, an irregular boundary topography, the lack of a geophysical constraint on mantle layering makes these recent models rather arbitrary. The size and shape of the deep layer are now free

---

**Figure 1.** Schematic drawing to clarify the definitions of the bulk Earth Urey ratio (Ur), the convective Urey ratio, and the internal heating ratio (IHR). $Q_{cc}$ is surface heat flux from continental regions, $Q_{om}$ is surface heat flux from oceanic regions, $Q_{cm}$ is heat flux from subcontinental mantle, $H_{cc}$ is the heat production of continental crust, $H_m$ is the heat production of the mantle, $S_m$ is heat flux due to the secular cooling of the mantle, $Q_c$ is core heat flux, and $S_c$ is heat flux due to the secular cooling of the core (including the effect of inner core growth). The presence of heat-producing elements in the core is neglected here for simplicity.

**Figure 2.** (a) Large-scale versus (b) small-scale geochemical reservoirs and their implications for the structure of mantle convection. In order to sequester a large-scale reservoir from convective mixing, some kind of layered-mantle convection would be necessary. The existence of small-scale reservoirs does not conflict with whole mantle convection and is actually expected from plate tectonic convection.
parameters to be adjusted, and these parameters may even be time-dependent if there is finite mass transport between the upper and lower layers. Such temporal variation provides vast degrees of freedom to a layered-mantle model because seismology yields only the present-day snapshot. In this line of effort, making a unified mantle model becomes a severely underdetermined problem. Though increasing model complexity is certainly one way to cope with the apparent conflicts between geophysics and geochemistry, pushing down the boundary interface below 660 km and making it invisible as well as time varying has resulted in an enormous jump in the number of model parameters, with little prospect of observationally constraining them. Thus some of us may be tempted to ask ourselves, have we come to a dead end?

[8] Alternatively, it may be the time for us to step back and see whether geochemical arguments for mantle layering are so robust to force us to consider something we cannot physically observe. In this section, various arguments involving global mass balance are examined first, followed by discussion on isolated reservoirs and mantle mixing.

2.1. Global Mass Balance and Its Uncertainty

2.1.1. Bulk Silicate Earth and Its Derivatives

[9] The central concept in any type of global mass balance argument is the bulk silicate Earth (BSE), which is also known as the primitive mantle (PM). BSE refers to the solid Earth excluding the core, and it may be regarded as the primordial mantle formed after core formation but before the extraction of continental crust. By definition, all of present-day silicate reservoirs have been derived from BSE, so they must add up to BSE. We can readily recognize two major silicate reservoirs: the continental crust (CC) and the depleted mantle (DM), which have long been considered to be complementary to each other [e.g., Hofmann, 1988, 2003]. That is, the continental crust is the product of mantle melting, whereas the depleted mantle is the solid residue. A number of trace elements are incompatible with solid phases, so they tend to be highly concentrated in a melt phase, which is the continental crust in the context here. The mass of continental crust is only ~0.6% of that of the mantle, but the continental crust is highly enriched in incompatible elements, so its contribution to mass balance is substantial. The depleted mantle is depleted in those incompatible elements with respect to BSE but not depleted in compatible elements such as major elements.

[10] Note that although the oceanic crust (OC) is also generated by mantle melting, it is usually not considered as an independent chemical reservoir in terms of global mass balance because it is quickly recycled back to the mantle by subduction. At a normal condition, suboceanic mantle beneath mid-ocean ridges has a potential temperature (hypothetical temperature of mantle adiabatically brought to the surface without melting) of ~1350°C and starts to melt at the depth of ~80 km, resulting in an ~6-km-thick oceanic crust and an ~70-km-thick depleted mantle lithosphere (DML) [e.g., McKenzie and Bickle, 1988; Langmuir et al., 1992]. DM in global mass balance refers to the source mantle before this melting takes place, and DML is even more depleted than DM. By mass conservation, the sum of OC and DML should be equivalent to DM if the effects of degassing and seawater alteration are taken into account.

2.1.2. Missing Heat Source Paradox and Its Solutions

[11] Among a few global mass balance arguments, the missing heat source paradox is most relevant to the thermal budget, so it is explained in some detail here. The continental crust is estimated to have 1.3 ppm U, 5.6 ppm Th, and 1.5% K [Rudnick and Gao, 2003], the heat production of which is ~7.5 TW (assuming the continental mass of $2.3 \times 10^{22}$ kg). According to the model of Salters and Stracke [2004], the abundance of those elements in the depleted mantle is 4.7 ppb U, 13.7 ppb Th, and 60 ppm K, and the corresponding heat production is ~4.2 TW. Combined, these two reservoirs yield ~12 TW. However, the bulk silicate Earth is commonly assumed to have 20 ppb U, 80 ppb Th, and 240 ppm K [e.g., McDonough and Sun, 1995], which give rise to the heat generation of ~20 TW. Assuming that these values are reasonably accurate, the global mass balance does not seem to hold for these heat-producing elements. This is called the missing heat source paradox, and a key question is the following: where did we make a mistake?

[12] This failure of mass balance has been used to argue against whole mantle convection and to support the presence of a hidden reservoir in the deep mantle. This line of logic is understandable because the composition of the depleted mantle is based largely on the chemistry of mid-ocean ridge basalts (MORB). As mentioned in section 2.1.2, mid-ocean ridge magmatism is the result of mantle melting at shallow depths (<100 km), so a composition model for the depleted mantle is fundamentally biased to the upper portion of the mantle. The above mass balance thus implicitly assumes that the mantle is well mixed by whole mantle convection so the shallow mantle sampling is not an issue. As we just saw, however, this assumption does not lead to a self-consistent mass balance, so it seems that we should abandon the notion of well-mixed mantle and call for a deep reservoir that is more enriched in the heat-producing elements. As the classical layered-mantle model gradually lost its popularity, various alternatives have been proposed for this “hidden” reservoir. The reservoir may occupy a substantial portion of the lower mantle, and its interface with the upper layer is somehow seismologically invisible [e.g., Kellogg et al., 1999]. Or the reservoir may be distributed as a number of blobs [e.g., Manga, 1996; Becker et al., 1999; Helffrich and Wood, 2001], though it is uncertain why such blobs do not contribute much to mid-ocean ridge magmatism (to an extent that enriched MORB becomes much more common than observed). Or it may correspond to the thin, D” layer at the base of the mantle [e.g., Coltice and Ricard, 2002; Tolstikhin and Hofmann, 2005], but the layer might be too enriched in heat-producing elements to have been gravitationally stable in the past.

[13] The very first step we should have taken instead may be to examine the accuracy of the mass balance calculation that caused the paradox. As discussed in the following,
every uncertainty. First of all, the most recent compilation for continental crust composition by Rudnick and Gao [2003] assigns 30% uncertainty for trace elements, so the heat production in the continental crust is in the range of 5 to 10 TW. If one looks more closely at how the average composition of continental crust is estimated, even greater uncertainty may be more realistic. The continental crust is typically divided into three layers such as upper, middle, and lower crust, and all layers exhibit considerable regional variability. Global averaging is thus a difficult task. The 30% uncertainty adopted by Rudnick and Gao [2003] is for the upper crustal composition, which may be estimated most reliably because a large number of data are available. The deeper layers are expected to have larger uncertainties, which are left unquantified because of the paucity of data. One may argue that such uncertainty may not concern us because the deeper layers are often considered to be depleted in heat-producing elements. The conventional wisdom of heat production exponentially decreasing with depth [Lachenbruch, 1970], however, is just a model and does not seem to be supported by observations [e.g., Jaupart and Mareschal, 2003]. Also, intracrustal differentiation, which is likely to have occurred in the evolution of continental crust [e.g., Rudnick, 1995], may deplete the lower crust of K but probably not of U and Th, because these two elements are highly concentrated in accessory minerals such as zircon, which tend to be unaffected by crustal melting [e.g., Watson and Harrison, 1984]. Existing estimates for the lower crustal composition vary by an order of magnitude in terms of trace elements, and the “best estimate” according to Rudnick and Gao [2003] is based on crustal xenolith data, though they also acknowledge that it is unclear how well xenolith compositions could represent average lower crust and also that xenolith data are spatially limited. Given the potential of the continental lower crust therefore, the total heat production of continental crust may even be higher than 10 TW. Mantle geotherms based on xenolith data [e.g., Finnerty and Boyd, 1987; Russell and Kopylova, 1999] may be used to bound crustal heat production, but such constraints are not available globally. Though the heat production should not exceed the observed continental heat flow, this upper bound itself has been revised from 12 TW [Pollack et al., 1993] to 14 TW [Jaupart et al., 2007] on the basis of more recent heat flow data. The distribution of existing heat flow measurements is highly heterogeneous both on continents and on ocean basins [e.g., Pollack et al., 1993, Figure 1], so the recent upward correction is not surprising.

[14] Estimating the composition of the depleted mantle suffers from more compounded difficulties. At shallow depths, the depleted mantle usually differentiates into oceanic crust and depleted mantle lithosphere by melting, so the composition of the (unmelted) depleted mantle must be indirectly estimated based on these chemically differentiated phases. The oceanic crust itself is layered because of chemical fractionation, and only the very top portion (i.e., MORB) is usually accessible for sampling. So the estimate based on MORB has to take into account the effects of chemical fractionation as well as mantle melting. This indirectness may be compensated, however, by the fact that the oceanic crust is the product of mantle melting over the depths of several tens of kilometers. MORB samples thus have a potential to effectively probe the shallow upper mantle, with its small-scale heterogeneities automatically homogenized by melting. Perhaps a more serious issue is the considerable regional variations observed in the trace element chemistry of MORB, which tends to show the so-called “enriched” signature (higher concentrations of more incompatible elements) near hot spots such as Iceland [e.g., Schilling et al., 1983]. Samples from these anomalous regions have traditionally been disregarded when discussing the depleted mantle, because hot spots have often been considered to originate in a deep geochemical reservoir that is different from the depleted mantle. One may also be tempted to apply this type of screening to limit data variability and to estimate an average with small uncertainty. The origin of hot spot magmatism is, however, still debated (section 2.2.1), and this screening may not be warranted. The concentration of heat-producing elements in the depleted mantle has been estimated to be 8 ppb U, 16 ppb Th, and 100 ppm K by Jochum et al. [1983] and 4.7 ± 1.4 ppb U, 13.7 ± 4.1 ppb Th, and 60 ± 17 ppm K by Salters and Stracke [2004], but these estimates are biased to the depleted end-member of MORB samples because of this screening. An alternative approach by Workman and Hart [2005] based on the trace element systematics of abyssal peridotites (exhumed pieces of DML) may be able to avoid this type of bias, but it suffers from different kinds of indirectness; their approach depends on the BSE composition as well as assumed isotopic evolution due to continental growth. The 30% error assigned by Salters and Stracke [2004] translates to the heat generation of 2.9–5.5 TW, but it would be corrected upward if data screening is loosened [Langmuir et al., 2005]. At the same time, expanded data would certainly exhibit much more variability. A careful uncertainty analysis has not been published, but it would be vital to derive realistic bounds on heat production. An estimate without uncertainty is of little value when scrutinizing global mass balance.

[15] Finally, the composition of the bulk silicate Earth represents what we expect for the bulk silicate Earth. Compared to our estimates for the continental crust and the depleted mantle, therefore, it is most theoretical, depending critically on the validity of theoretical assumptions employed. A number of geochemists have estimated the BSE composition with different sets of assumptions (see a review by Lyubetskaya and Korenaga [2007a]), and the so-called pyrolite approach [e.g., McDonough and Sun, 1995] appears to be most robust as it requires only the following two assumptions: (1) the BSE should be found somewhere along the compositional trend displayed by mantle rocks, and (2) the ratios of refractory lithophile elements (RLE) in the BSE are chondritic. If either of these assumptions is invalid, it is impossible to estimate the BSE composition (not only by the pyrolite approach but also by
other existing methods), so its uncertainty would be essentially unbounded. The first assumption implicitly requires whole mantle convection. Our mantle samples such as mantle xenoliths and massif peridotites are all from the upper mantle, so theoretically the composition of the lower mantle could be different from that of the upper mantle, and if so, the RLE ratios could not be imposed on the composition trend of the upper mantle samples because these cosmochemical constraints are valid only when considering BSE as a whole. For the estimated BSE composition to be justified, therefore, global mass balance must be self-consistent. This issue is explained in more detail in section 2.1.3. The second assumption is reasonable because refractory elements have high condensation temperatures, so they are unlikely to have been fractionated from each other during planetary formation processes, and lithophile elements do not enter the core. The RLE ratios are the most stable ratios among various classes of chondrites, and, indeed, the isotope data of terrestrial samples require that at least Sm/Nd, Lu/Hf, and Th/U (these are all RLEs) should be within a few percent of the chondritic value [e.g., Lyubetskaya and Korenaga, 2007a]. With these assumptions, we can quantify the uncertainty of the BSE composition by taking into account the uncertainty of observed compositional trend as well as that of chondritic RLE ratios, and Lyubetskaya and Korenaga [2007a] obtained ~20% error for the BSE abundance of refractory lithophile elements. This uncertainty represents the influence of scatter in the upper mantle compositional trend under the imposed chondritic constraints. Lyubetskaya and Korenaga also found that the previous estimate of 20 ppb U and 80 ppb Th should be corrected slightly downward, and the new estimate with one standard deviation is 17 ± 3 ppb U and 63 ± 11 ppb Th. As K is not a RLE, it must be derived with additional constraints such as bulk Earth K/U and K/La, which further amplify uncertainty. The revised estimate is 190 ± 40 ppb K. The likely range of BSE heat production is then 13–19 TW.

[16] Note that these uncertainties in heat production do not correlate to each other because the compositions of these silicate reservoirs are estimated independently of each other. Using one standard deviation therefore, the range of combined CC and DM heat production is 7.9–15.5 TW, which overlaps with the range of BSE heat production (note also that the upper bound of 15.5 TW is likely to increase if the DM composition is estimated with more even data sampling). The missing heat source paradox is thus unproven and probably does not exist; it could well be an artifact caused by neglecting unavoidable uncertainties in geochemical inference. Given the uncertainties, it is still possible to postulate a hidden reservoir (e.g., by assuming 5 TW for CC, 3 TW for DM, and 19 TW for BSE), but it is no longer required.

2.1.3. Importance of Self-Consistent Mass Balance

[17] The self-consistent mass balance is vital to justify the estimated BSE composition because, as noted in section 2.1.2, whole mantle convection has to be assumed when deriving the composition. In other words, if mass balance is not self-consistent, the BSE composition is undefined. Interestingly, however, the notion of layered-mantle convection has coexisted with the BSE composition estimated by assuming whole mantle convection. A common layered-mantle model advocated by geochemists is illustrated in Figure 3b; the upper layer is the depleted mantle as observed through mid-ocean ridge magmatism, while the lower layer is identified as the primitive mantle [e.g., Jacobsen and Wasserburg, 1979; DePaolo, 1980; Allegre, 2002]. Note that the PM composition is identical to the BSE composition, both in major and trace elements, and is very similar to the DM composition in terms of major elements.
The volume of the upper layer is determined so that the average composition of these three reservoirs (CC, DM, and PM) is equal to the BSE composition. This is, of course, one possibility, but the average composition can be (vastly) different from the estimated BSE composition both in major and trace elements (Figure 3c). In fact, if a layered-mantle model has to be true, the major element composition of the lower layer may be required to be different from the PM composition. In the classical layered-mantle model, the internal boundary coincides with the 660-km discontinuity, and it was speculated that the endothermic phase change expected at this discontinuity may be able to sustain the layering [e.g., Christensen and Yuen, 1984]. Later studies on mantle convection, however, suggested that it was difficult to maintain the layered state with the phase change alone [e.g., Tackley et al., 1993; Zhong and Gurnis, 1994], and this idea became obsolete particularly after the internal boundary was pushed down the 660-km discontinuity. To achieve a layered mantle in a physically plausible manner, therefore, a different major element composition (e.g., enriched in iron) seems necessary to make the lower layer intrinsically denser than the upper layer [e.g., Kellogg et al., 1999; Davaille, 1999].

A failure to close global mass balance therefore presents a serious challenge to our understanding of the bulk Earth chemistry. It does not only indicate the presence of a hidden reservoir, but it also makes the BSE composition (and thus heat production) unconstrained (Figure 3c). The BSE composition is meaningful only with whole mantle convection (Figure 3a), and it cannot be used to estimate the composition of a hidden reservoir.

2.1.4. Missing Argon Paradox and More

As discussed in section 2.1.2, the missing heat source paradox does not seem to be as compelling as commonly described. Given the uncertainty involved, however, the argument could still be used to support layered-mantle convection if any of the other mass balance arguments is more robust, and it would be difficult for geochemistry to constrain the amount of internal heat production in the bulk of the mantle. Other arguments involve the mass balance of noble gases such as argon and xenon [e.g., Allègre et al., 1996; Porcelli and Turekian, 2003] and lithophile elements [e.g., Helffrich and Wood, 2001; Albarede and van der Hilst, 2002], but as recently reviewed by Lyubetskaya and Korenaga [2007b], on the basis of the revised BSE composition, these mass balances do not contradict whole mantle convection if uncertainty is properly taken into account.

2.1.5. Isotopic Mass Balance and Two Versions of Chondrite Assumptions

The final kinds of mass balance to be discussed are the ones that involve isotopic ratios such as $^{143}\text{Nd}/^{144}\text{Nd}$, $^{176}\text{Hf}/^{177}\text{Hf}$, and, more recently, $^{142}\text{Nd}/^{144}\text{Nd}$. Though similar to other mass balances discussed so far, isotopic mass balance arguments involve a much more strict definition for the chondrite assumption, warranting a separate treatment.

The $^{143}\text{Nd}/^{144}\text{Nd}$ ratio, for example, reflects the long-term evolution of $^{148}\text{Sm}/^{144}\text{Nd}$ (or Sm/Nd) as $^{143}\text{Nd}$ is a decay product of $^{147}\text{Sm}$. Chondritic meteorites exhibit a tight distribution of $^{147}\text{Sm}/^{144}\text{Nd} (0.196 \pm 0.002)$, which corresponds to $^{143}\text{Nd}/^{144}\text{Nd} (0.51263 \pm 0.00005)$ [Patchett et al., 2004, Figure 1] (Note that the $\sim 1\%$ deviation in $^{147}\text{Sm}/^{144}\text{Nd}$ reduces to only $\sim 0.01\%$ deviation in $^{143}\text{Nd}/^{144}\text{Nd}$ because of the long half-life involved.) The $^{143}\text{Nd}/^{144}\text{Nd}$ ratio of terrestrial samples is usually compared to this well-defined (present day) chondritic ratio or its value in the past, and the difference is expressed in the $\varepsilon^{143}\text{Nd}$ notation defined as

$$
\varepsilon^{143}\text{Nd}(t) = \left[ \frac{\left( \frac{^{143}\text{Nd}}{^{144}\text{Nd}} \right)_{\text{sample}} (t)}{\left( \frac{^{143}\text{Nd}}{^{144}\text{Nd}} \right)_{\text{CHUR}} (t)} - 1 \right] \times 10^4, \tag{6}
$$

where $t$ is the age of a sample and CHUR refers to a chondritic uniform reservoir. A common argument is that the BSE $\varepsilon^{143}\text{Nd}$ would be $\sim 5$ if CC contains 20 ppm Nd with $^{143}\text{Nd}$ of $-16$ and DM contains 0.7 ppm Nd with $^{143}\text{Nd}$ of 9 [e.g., Lassiter, 2004]. A major source of uncertainty is the depleted mantle, which stores more than $80\%$ of total Nd in this mass balance. Though both $\varepsilon^{143}\text{Nd}$ and Nd content of MORB vary from sample to sample by $\sim 30\%$, they tend to be negatively correlated (i.e., a more enriched sample has a lower $\varepsilon^{143}\text{Nd}$) [e.g., Hofmann, 2003, Figure 14], so it is difficult to lower the BSE $\varepsilon^{143}\text{Nd}$ below $\sim 3$ even if we take into account all of the likely uncertainties (though a more complete statistical study of MORB chemistry may indicate otherwise in the future). Because chondritic samples show $\varepsilon^{143}\text{Nd}$ of $\pm 1$ as mentioned above, this relatively high BSE value seems to require a hidden reservoir with low Sm/Nd (i.e., enriched in incompatible elements because Nd is more incompatible than Sm).

Note that the chondrite assumption for BSE is more strict here than in the case of estimating the BSE composition. Requiring the BSE $\varepsilon^{143}\text{Nd}$ to be within $\pm 1$ is equivalent to assuming that the ratio Sm/Nd is within $1\%$ deviation from the chondrite average, but the Sm-Nd isotopic evolution recorded by terrestrial samples does not place such a tight constraint [Lyubetskaya and Korenaga, 2007a, 2007b]. Up to a few percent deviation is acceptable because the earliest record for the depleted mantle at $\sim 4$ Ga ago shows $\varepsilon^{143}\text{Nd}$ of as high as $\sim 4$ [e.g., Bennett, 2003]. Thus, requiring the RLE ratios such as Sm/Nd to be within a few percent deviation from the chondritic value is justified by existing data, whereas imposing the $\pm 1\%$ deviation is not. It is important to distinguish between these weak and strong versions of the chondrite assumptions. When the strong version is not satisfied by a certain mass balance, it is often called “nonchondritic,” and this might imply that refusing the presence of a hidden reservoir violates cosmochemical considerations. We do not understand well, however, how efficiently elements and isotopes were homogenized in the presolar nebula and how much they were later modified by planetary accretion processes [e.g., Halliday and Porcelli, 2001].

As already mentioned, even $^{147}\text{Sm}/^{144}\text{Nd}$ in chondrites has as much as $\pm 1\%$ variation (and $^{169}\text{Lu}/^{177}\text{Hf}$ has more than $\pm 5\%$ variation [Patchett et al., 2004]). Further-
more, chondrites are believed to originate mostly from the asteroid belt [e.g., Scott and Krot, 2003], so even though the number of chondrite samples is large, we are essentially looking at the highly restricted portion of the solar system; the mass of the asteroid belt is less than 0.1% of that of Earth. The stable statistics of chondrite compositions may be just due to more or less repetitive sampling. Earth is the best sampled planet, and yet we cannot measure its bulk isotope characteristics. We have only a few dozen meteorites from Mars, and no sample for Venus and Mercury, so their bulk isotopic compositions are even more uncertain than Earth’s. Note that the spectroscopic measurement of the photospheric composition of the Sun may have ∼25% uncertainty regarding Sm/Nd [Grevesse and Sauval, 1998] (assuming that reported errors are not correlated). A remarkable correlation between the photospheric composition and the chondrite composition exists for a number of elements, but this correlation in the logarithmic space does not impose a tight constraint on the ratios of trace elements. Some authors are well aware of this possibility of limited sampling by chondrites [e.g., Wasson and Kallemeyn, 1988; Drake and Righter, 2002], but the strong version of the chondrite assumption has long been a part of central dogma in geochemistry.

[24] Recently, Boyet and Carlson [2005] proposed a new kind of argument for a hidden reservoir based on the now extinct isotope $^{146}$Sm, which decayed into $^{142}$Nd. So far, measured terrestrial samples show systematically higher $^{142}$Nd/$^{144}$Nd than chondrites by $\sim 0.2 \pm 0.14 \epsilon$ unit. To have the BSE $^{142}$Nd be (strictly) chondritic, therefore, there must be an unobserved reservoir, whose $^{142}$Nd is more negative than chondritic. Also, because of the short half-life of $^{146}$Sm, this reservoir has to have formed within the first 100 Ma of Earth’s history. This presumed reservoir does not seem to be sampled even by hot spot magmatism, so if its volume is substantial, it must somehow be able to avoid entrainment by the convection mantle. Alternatively, the reservoir could be trivially small or even nonexistent if the BSE $^{142}$Nd (or Sm/Nd) does not have to exactly equal the chondritic average. With igneous differentiation, Sm/Nd in the outer portions of planetesimal bodies would be low, and impact erosion could preferentially remove major portions of the outer silicate portions of accreting planetesimals, resulting in a slightly elevated Sm/Nd ratio in the growing Earth [Halliday, 2003]. As Lyubetskaya and Korenaga [2007b] argued, both $^{142}$Nd and $^{142}$Nd arguments rely heavily on the strong version of the chondrite assumption. Andreasen and Sharma [2006] showed that ordinary chondrites have systematically higher $^{142}$Nd than carbonaceous chondrites, and the bulk Earth may be located somewhere along this trend. To summarize, given the resolution of available observational constraints on the BSE composition, existing mass balance arguments, whether elemental or isotopic, do not necessarily contradict the notion of whole mantle convection.

2.2. Isolated Reservoirs and Mantle Mixing

[25] Among different types of terrestrial magmatism, mid-ocean ridge magmatism is volumetrically the most significant, with the emplacement rate of $\sim 21 \text{ km}^3 \text{ a}^{-1}$ [Crisp, 1984]. Oceanic island magmatism (or hot spot magmatism) is 1 order of magnitude smaller, but understanding its origins has been an important subject both in geophysics and geochemistry because explaining hot spot magmatism requires some special dynamical mechanisms that are different from large-scale mantle circulation associated with plate tectonics [e.g., Morgan, 1971; Davies, 1992] and because ocean island basalts (OIB) exhibit distinct isotopic heterogeneities in comparison to MORB [e.g., Zindler and Hart, 1986; Hofmann, 1997]. Any successful model for hot spot magmatism must be able to address these geophysical and geochemical issues and also to be consistent with the bigger picture of Earth’s mantle. The most popular hypothesis has been the mantle plume hypothesis originally proposed by Morgan [1971]; hot upwellings are generated by the instability of a thermal boundary layer located either at the core-mantle boundary (in the case of whole mantle convection) or at the 660-km discontinuity or somewhere in the midmantle (in the case of layered-mantle convection). Excess mantle temperature expected for hot spot magmatism (i.e., mantle melting beneath oceanic lithosphere) is readily explained because the bottom thermal boundary layer is hotter than the convecting interior, and the unique isotopic characteristics can be explained by assuming long-lived chemical heterogeneities at the base of the mantle or in the lower mantle.

[26] In both versions of the plume hypothesis, the geochemistry of OIB is used to support the presence of a distinct reservoir that must have been isolated from the other, well-mixed parts of the mantle. However, note that this argument does not involve an explicit constraint on the size of such reservoir. Because the scale of hot spot magmatism is small compared to that of mid-ocean ridge magmatism, the volume of OIB source mantle does not have to be large (based on thermal swells, the total volume flux of mantle plumes is estimated to be $\sim 85 \text{ km}^3 \text{ a}^{-1}$, assuming excess temperature of 200 K [Sleep, 1990; Schubert et al., 2001]) but does not have to be small either; it depends on the assumed style of mantle convection. This nonuniqueness is further aggravated by the fact that the plume hypothesis is not the only theory proposed to explain hot spot magmatism. Also, postulating an isolated reservoir at or below the thermal boundary layer implicitly assumes that mantle convection is a relatively efficient mixing process that cannot maintain long-lived chemical heterogeneities within, but the rate of mantle mixing is a strong function of the (time dependent) viscosity structure of the mantle. These issues will be discussed in sections 2.2.1 and 2.2.2.

2.2.1. Origins of Hot Spots

[27] Large-scale mantle circulation associated with plate tectonics does not predict the existence of hot spots, so some kind of small-scale anomalies are necessary to induce mantle melting. The plume hypothesis provides thermal anomalies in the form of convective instabilities rising from the bottom boundary layer. Though such instabilities are likely to exist in the cooling Earth, whether plumes can
explain all hot spots is another issue. At least for some hot spots, in fact, the chemistry of hot spot magma is clearly inconsistent with the melting of hot mantle, suggesting that chemically anomalous mantle may play an important role [e.g., Korenaga and Kelemen, 2000]. In recent years, debates over the existence of mantle plumes have been particularly intense [e.g., Anderson, 2000; DePaolo and Manga, 2003; Foulger and Natland, 2003; Sleep, 2003; Campbell, 2005; Foulger et al., 2005b; McNutt, 2006] but sometimes appear to be largely philosophical and semantic.

What is frustrating is that observations tend to be ambiguous enough to allow conflicting interpretations to coexist and that our understanding of mantle dynamics is too incomplete in a number of aspects to allow robust predictions. It is perhaps true that the notion of mantle plumes has been overused in the past, but it would also be extreme to reject it entirely. Some hot spots such as Iceland and Galapagos may require some special mechanism other than a mantle plume [e.g., Korenaga, 2004; Sallares et al., 2005], but others such as Hawaii may simply be caused by plumes. It might be aesthetically pleasing to have just one theory to explain all hot spots, but existing observations indicate that the reality is unfortunately rather complex [e.g., Courtillot et al., 2003].

The Iceland hot spot, for example, is a classical example of a ridge-centered hot spot, but it does not fit well into the standard theory of a mantle plume. A number of seismological studies show that Iceland is underlain by a low-velocity anomaly at least in the shallow upper mantle [e.g., Wolfé et al., 1997; Foulger et al., 2001], but the thermal boundary layer at the core-mantle boundary is therefore a corollary of the present-day heat budget, unless all of core heat flux is consumed to just warm up cold slabs arriving at the core-mantle boundary. The convective instability of this boundary layer is only one step further in this line of logic. With temperature-dependent viscosity, the instability should evolve into a large plume head followed by a narrow conduit [e.g., Whitehead and Luther, 1975; Griffiths and Campbell, 1990].

The form of convective instability, however, depends strongly on the rheology of the lowermost mantle and may be different from what the standard plume theory assumes [Korenaga, 2005a; Cadek and Fleitout, 2006]. Whereas the rheology of the upper mantle is now understood reasonably well [e.g., Karato and Wu, 1993; Hirth and Kohlstedt, 2003; Korenaga and Karato, 2008], that of the lower mantle is still largely unknown because of experimental difficulties. Only a broad agreement has been achieved for spatially averaged viscosity on the basis of geodynamical estimates [e.g., King, 1995; Forte and Mitrovica, 2001], and any other details such as temperature and pressure dependencies are yet to be determined. The dynamics of the core-mantle boundary region may be further complicated by the likely presence of chemical heterogeneities [e.g., Farquharson, 1997; Montague and Kellogg, 2000; Jellinek and Manga, 2002], the postperovskite phase transition [e.g., Murakami et al., 2004; Nakagawa and Tackley, 2004], and drastic changes in thermal properties [e.g., Badro et al., 2004; Lin et al., 2005; Matyska and Yuen, 2005], so even if we limit ourselves to the dynamics of thermal plumes to a static background, our understanding is still immature. Note that this directly translates to the uncertainty of core-mantle interaction, i.e., how efficiently the mantle could cool the core (section 3.5).

2.2.2. Mantle Viscosity, Mantle Mixing, and Sampling by Magmatism

Numerical studies on mantle mixing indicate that it is difficult to preserve a large-scale geochemical reservoir over Earth’s history in the convecting mantle by anything other than density stratification, such as phase transitions,
higher viscosity in the deep mantle, and/or depth-dependent thermal properties [e.g., van Keken and Ballentine, 1998, 1999; Hunt and Kellogg, 2001; Naliboff and Kellogg, 2007]. Though the presence of multiple geochemical reservoirs with different isotopic characteristics is a well-established observational fact, most notably supported by the studies of noble gas isotopes [e.g., Kurz et al., 1982; M. Honda et al., 1993; Tolstikhin and Hofmann, 2005], this observation by itself does not necessarily demand a large-scale reservoir, as noted in section 2.2, and the reevaluation of global mass balance arguments casts a doubt on the existence of a massive reservoir to begin with (section 2.1). The difficulty in preserving a large-scale reservoir thus does not present any challenge to reconciling mantle dynamics and geochemistry.

32 In contrast, small-scale reservoirs, such as distributed blobs in the so-called “marble cake mantle” or “mantle plum pudding” [e.g., Allègre and Turcotte, 1986; Phipps Morgan and Morgan, 1999], are expected to be better preserved [e.g., Christensen and Hofmann, 1994; Davies, 2002; Xie and Tackley, 2004], and thus whole mantle convection with small-scale chemical heterogeneities (Figure 2b) may be a promising concept both geophysically and geochemically. The fate of small-scale heterogeneities in the convecting mantle and their relation to geochemical observations are, however, difficult issues to be fully explored in future. A few key issues are highlighted in the following.

33 First of all, “small-scale” heterogeneities refer here to anything but a massive reservoir that occupies a significant portion of the mantle, so their spatial scales are only vaguely defined. Theoretical studies on mantle mixing suggest that convective mixing, through repeated stretching and folding, could easily homogenize heterogeneities down to submeter scales within a billion years or so [e.g., Olson et al., 1984; Hoffman and McKenzie, 1985; Ferrachat and Ricard, 1998], but note that such studies focus on the destruction of existing heterogeneities by isoviscous convection. Though isotopic heterogeneities are most easily detected by geochemistry and thus most frequently discussed, they are likely to be accompanied by variations in trace element and major element concentrations (e.g., subducted oceanic crust and depleted mantle lithosphere), which, in turn, should accompany viscosity variations because viscosity is a strong function of lithology as well as trace elements such as hydrogen [e.g., Karato, 1986; Hirth and Kohlstedt, 1996; Karato, 1997]. Convective mixing in the presence of such composition-dependent viscosity is expected to be inefficient [Manga, 1996]. Moreover, as noted in section 2.2.1, chemical heterogeneities are continually generated by plate tectonics, which must adversely affect homogenization. Field observations indeed indicate that chemical heterogeneities exist at almost all plausible scales: from centimeters to more than a few hundreds kilometers [e.g., Zindler and Hart, 1986; Carlson, 1994]. A number of seismological studies also support the existence of chemical heterogeneities throughout the mantle [e.g., Hedlin et al., 1997; Helffrich and Wood, 2001]. These observations are not surprising because a chemically heterogeneous mantle is a rule, not an exception, in the presence of plate tectonics, but at the same time they demand that we tackle a highly complex problem. For example, we know that a wide range of spatial scales must be involved, but we do not know the relative significance of different scales [cf. Gurnis, 1986]. The nature of chemical heterogeneities could also vary from place to place [e.g., Hirschmann and Stolper, 1996]. The concept of chemically heterogeneous mantle thus necessarily involves a few poorly constrained parameters, so unfortunately it tends to be regarded as an ad hoc explanation or a “just so” story, especially if applied to regional problems.

34 Another important issue regarding chemical heterogeneities is how they are sampled by geochemistry. Different lithologies are likely to have not only different viscosities but also different densities, and as mentioned in section 2.2.1, more fertile mantle may be preferentially sinking and thus less likely to be sampled by melting, which usually requires upwelling. The upwelling of intrinsically dense heterogeneities requires special conditions such as thermal anomalies and plate tectonic stresses, and the interplay among these competing factors needs to be systematically explored. Also, how a composite mantle with different lithologies would melt is an interesting problem by itself, which is still poorly understood. Enriched crustal fragments embedded in a more depleted mantle matrix, for example, are expected to melt more easily than the matrix, and because of this differential melting behavior, a single composite mantle could potentially yield melts with different isotopic characteristics depending on whether its upwelling (and thus melting) is limited to the base of lithosphere (in case of midplate hot spots) or not (under mid-ocean ridges) [e.g., Sleep, 1984; Meibom and Anderson, 2003; Ito and Mahoney, 2005]. Geochemical evidence for such composite mantle has been slowly accumulated [e.g., Hauri, 1996; Takahashi et al., 1998; Korenaga and Kelemen, 2000; Sobolev et al., 2007], but inferring the composition of a source mantle and its potential temperature from lava samples is still a daunting task because predicting a lava composition from a given composite mantle is much more involved than the case for a homogeneous mantle. In a homogeneous mantle, for example, the melt phase can almost always migrate through the residual mantle because a porous flow network is expected to form at a very low porosity [Kohlstedt, 1992], but it is not clear whether the melting of a composite mantle behaves in a similar manner. Early melts in the enriched fragments may not escape into the surrounding matrix until the matrix itself starts to melt, and because the early melting extracts heat from the matrix [Sleep, 1984; Phipps Morgan, 2001], it could also delay the melting of the matrix. The consequence of polybaric melting such as chemical reaction between the residual mantle and the migrating melt phase is already a complex problem even for a homogeneous mantle [e.g., Kelemen et al., 1995; Spiegelman and Kelemen, 2003], and extending it to the case of a composite mantle may require more than a brute
force approach given compositional diversity and various spatial scales involved.

[35] Finally, duration is an essential factor for mixing. Even in the case of simple isoviscous mixing, for example, homogenization would not be achieved if convection lasts for the period of only 100 Ma. Earth’s mantle has been convecting probably for more than 4 Ga, but given the variable mantle properties, can it be considered sufficiently long? A good starting point is to estimate how quickly the mantle can be processed by mid-ocean ridge magmatism, which may be expressed as

\[ \dot{M} = \rho_mRD, \]

where \( \rho_m \) is the density of shallow upper mantle (3300 kg m\(^{-3}\)), \( R \) is the rate of plate construction at mid-ocean ridges, and \( D \) is the initial depth of mantle melting. With the current plate motion and mantle temperature, the processing rate \( \dot{M} \) is estimated to be \(~6.7 \times 10^{14} \text{ kg a}^{-1} \) [Korenaga, 2006], so it would take \(~6 \text{ Ga} \) to process the entire mantle. It is commonly assumed that mantle convection was more vigorous in the past, so this estimate based on the present-day processing rate may not be useful. In fact, some recent mixing calculations explicitly simulate more rapid convection in the past [e.g., Davies, 2002; Xie and Tackley, 2004].

Greater convective vigor in the past is, however, merely an assumption without geological evidence, and it is also known to cause a serious problem in the thermal history of Earth (section 3). A scaling law for plate tectonic convection suggests that mantle convection may have been less vigorous when the mantle was hotter in the past [Korenaga, 2003, 2006], and if so, the present-day processing rate could be a good approximation for the last 4 Ga or so (Figure 5f). This slow processing rate may be a good proxy for how quickly chemical heterogeneities are created and destroyed by terrestrial magmatism in general. In order to predict how efficiently convection could homogenize the mantle over the geological time, therefore, we need to have a good idea of how mantle convection has evolved over the last 4.6 Ga. Simulating realistic mantle mixing requires successfully simulating the long-term evolution of mantle convection at the same time, but doing so without violating observational constraints on the thermal budget is yet to be attempted.

[36] It should be clear now that we have just begun to understand how geochemical observations and mantle convection may be related. A deep, isolated reservoir has often been called for to preserve isotopic heterogeneities, but it is based mostly on the following two assumptions: (1) hot spot magmatism is caused by deep mantle plumes, and (2) mantle convection can quickly homogenize chemical heterogeneities. Both of them need to be carefully examined in the framework of plate tectonic convection. Chemically heterogeneous mantle, in terms of isotopes as well as trace and major elements, is an unavoidable consequence of mantle convection with plate tectonics, carrying a number of important implications for terrestrial magmatism and mantle dynamics. Whole mantle convection is conceptually a simple mechanism, but it can create complex phenomena, so its potential to reconcile geophysics and geochemistry should not be underestimated.

3. GEOPHYSICAL PERSPECTIVES: THERMAL EVOLUTION OF EARTH

[37] The structure of mantle convection has long been controversial even without considering the geochemical arguments detailed in section 2. One persistent theme is the mineral-physics interpretation of the preliminary reference Earth model (PREM) [Dziewonski and Anderson, 1981]. PREM is a one-dimensional model for the radial structure of seismic velocities, density, and attenuation, constrained by a volume of seismological data as well as geodetic observations. A key question here is whether PREM velocity and density profiles throughout the mantle can be explained by a single composition, and mineral physicists have not yet reached a consensus even after a few decades of research [e.g., Ringwood, 1975; Jeanloz and Knitke, 1989; Stixrude et al., 1992; Jackson, 1998; Lee et al., 2004a; Li and Zhang, 2005; Aizawa and Yoneda, 2006]. This is partly because there is always a trade-off between temperature and composition effects and accurately estimating a temperature profile is not trivial especially for deep mantle. Also, measuring the physical properties of minerals at high-pressure and high-temperature conditions expected in the lower mantle is still experimentally challenging, and extrapolating from laboratory conditions to lower mantle conditions tends to suffer from uncertainties in relevant mineral-physics parameters. First-principle calculations could alleviate the situation somewhat, but they also need experimental verification. A promising attempt to formulate this complex geophysical inference with Bayesian statistics has been evolving in recent years [Deschamps and Trampert, 2004; Mattern et al., 2005; Matas et al., 2007]. Perhaps an equally important task to be done is to evaluate the uncertainty of the reference model; PREM has often been treated by mineral physicists as if it is an axiom, but how much of its detail we should trust is still open to question [e.g., Kennett, 2006].

[38] Compared to this ambiguity in how to interpret the reference radial structure, the high-resolution seismic imaging of three-dimensional heterogeneities seems to have provided more compelling evidence for whole mantle convection in a few different perspectives. First, subducted slabs do seem to penetrate the 660-km discontinuity [e.g., van der Hilst et al., 1997; Grand et al., 1997; Karatasev and van der Hilst, 2001], though the style of penetration varies among different regions [Fukao et al., 1992; van der Hilst, 1995; Fukao et al., 2001] as one might expect from the diversity of slab parameters such as age and subduction rate [e.g., Karato et al., 2001]. Second, whereas slab penetration alone may not be able to negate deep layering [e.g., Kellogg et al., 1999], seismic images for deep mantle plumes under some major hot spots such as Tahiti and Samoa [Montelli et al., 2004, 2006] provide a complementary support for whole mantle-scale material transport (though the depth extent of
imaged plume-like structure is still controversial). Third, seismic heterogeneities are concentrated near the surface and the core-mantle boundary and are much reduced between them [e.g., Masters et al., 2000], questioning the presence of any other boundary layer within the mantle.

[39] Unfortunately, geophysical observations are limited to present-day snapshots by their nature, so their significance as constraints on the long-term evolution of Earth may be questioned [e.g., Allegre, 1997]. That is, whole mantle convection may be in operation only recently, and the mantle may have been layered for a major part of Earth’s history. Geophysics, however, could provide a long-term constraint by modeling the thermal history of Earth. Indeed, a need for layered-mantle convection has repeatedly been raised from such theoretical consideration [e.g., Richter, 1985; Kellogg et al., 1999; Butler and Peltier, 2002]. As will be discussed further in this section, we must combine a range of geophysical, geochemical, and geological constraints in a self-consistent framework when calculating a thermal history. While being a powerful tool to construct a unifying model for Earth’s evolution, this modeling approach can also be tricky. A variety of evolution models with different conclusions have been published, but some models suffer from internal inconsistencies, and others fail to satisfy key observational constraints. Without knowing how to critically evaluate the success of a given model, the existing literature on this important geophysical constraint would be confusing at best. Understanding an acceptable range for the present-day convective Urey ratio is of prime importance, so it is discussed first (section 3.1). Classical evolution models are then briefly described to provide a historical background (section 3.2), followed by a review of more recent models (section 3.3). It will be shown that a successful thermal history may be reconstructed without calling for layered-mantle convection, though it depends on assumptions about the dynamics of plate tectonic convection, which remain to be resolved. A few caveats on the rate of secular cooling are also given (section 3.4), and coupled core-mantle evolution models are discussed last (section 3.5).

3.1. Present-Day Convective Urey Ratio
[40] As defined in section 1, the convective Urey ratio denotes the fraction of total convective heat flux supported by heat production within the convecting mantle (Figure 1). This ratio is most relevant because the thermal history of Earth is modeled by focusing on the convective heat loss from the mantle; a contribution from heat production within the continental crust can easily be added afterward to give the total surface heat flux.

[41] A standard value for the global surface heat flux today has been ~44 TW [Pollack et al., 1993], which is the combination of heat flux from ocean basins (~32 TW) and from continents (~12 TW). The global heat flux has recently been revised to be ~46 TW by Jaupart et al. [2007], but this increase is due to their new estimate on continental heat flux (~14 TW); the oceanic heat flux is still ~32 TW. Present-day heat production within the continental crust is estimated to be ~7.5 ± 2.5 TW [Rudnick and Gao, 2003], but as discussed in section 2.1.2, it could be higher and is bounded only by the continental heat flux itself. Also, the uncertainty of the crustal composition is not the only issue. In addition to continental crust, subcontinental lithospheric mantle is usually considered to be sequestered from mantle convection, so its heat production must also be taken into account, though it is poorly known [Rudnick et al., 1998]. Heat flux into the base of subcontinental lithosphere may be close to zero [e.g., Jordan, 1988], and thus the lower bound for the convective heat flux is the oceanic heat flux itself (~32 TW). The upper bound is ~41 TW based on the latest continental heat flow by Jaupart et al. [2007] and the lowest possible heat production according to Rudnick and Gao [2003]. The present-day convective heat flux is therefore estimated to be ~36.5 ± 4.5 TW.

[42] Note that the oceanic heat flux of ~32 TW depends on more than actual heat flow measurements. Standard methods based on a vertical thermal gradient can measure conductive heat flow only, so it is difficult to obtain an accurate estimate of heat flux from young seafloor where a thin sedimentary cover allows a significant part of heat flux to escape by advection (i.e., hydrothermal circulation). Heat flow data are indeed highly noisy on seafloor younger than ~50 Ma, and simply integrating measured heat flow over ocean basins would yield only ~20 TW [e.g., Pollack et al., 1993]. This is a well-known problem with a well-known solution. Heat loss from seafloor must be reflected in the thermal structure of oceanic lithosphere, and because of thermal contraction, cooling results in denser lithosphere, which subsides in order to maintain isostasy. In other words, seafloor subsidence with increasing age should follow closely the integrated heat content of oceanic lithosphere, and the global oceanic heat flux can be reliably estimated by combining the observed depth-age relationship with heat flow data. This is how the global oceanic heat flux as well as the contribution of hydrothermal circulation have been estimated for decades [e.g., Parsons and Sclater, 1977; Stein and Stein, 1992; Jaupart et al., 2007]. The validity of the depth-age constraint has also been verified directly by heat flow data near the Juan de Fuca Ridge, which provides a rare example of young seafloor covered with a thick sedimentary layer [Davis et al., 1999]. Taking noisy heat flow data at face value and neglecting this important depth-age constraint thus most likely lead to erroneous conclusions on oceanic heat flux (see discussion by Von Herzen et al. [2005]).

[43] Assuming whole mantle convection, the present-day heat production in the convecting mantle would be 2.9–5.5 TW according to the composition model for the depleted mantle by Salters and Stracke [2004], but this estimate is better regarded as the lower bound because of the sampling bias discussed in section 2.1.2. An alternative approach based on global mass balance may be preferred, and subtracting the continental heat production of ~7.5 ± 2.5 TW from the BSE heat production of ~16 ± 3 TW gives ~8.5 ± 5.5 TW for the convecting mantle. The large uncertainty is due to the combination of the uncorrelated uncertainties of BSE and CC heat production. The convec-
tive Urey ratio is then estimated to be $0.23 \pm 0.15$. Note that this uncertainty is derived by considering the partially correlated nature of the uncertainties of the convective heat flux and the DM heat production, both of which are affected by the uncertainty of CC heat production.

[44] Recently, Grigné et al. [2005] offered a different perspective on the estimate of the convective Urey ratio. By noting that surface heat flux may significantly fluctuate with the aspect ratio of convection cells, they proposed that convective heat flux may be fluctuating with the period of a few hundred million years (i.e., the period of the Wilson cycle) and that the present-day heat flux may be at the maximum of such temporal variation. If so, a temporally average heat flux, which would be more relevant to long-term thermal evolution, could be substantially lower than $\sim 36$ TW, resulting in a higher estimate of the convective Urey ratio. This point was further elaborated by Labrosse and Jauwart [2007], who suggested that the amplitude of this temporal fluctuation could be as large as 25%. Their reasoning is the following. The present-day seafloor exhibits a peculiar area-age relationship; the total area of seafloor at a given age becomes smaller as the age increases [e.g., Sclater et al., 1980], and this so-called “triangular” distribution is because seafloor is almost randomly consumed by subduction regardless of its age [Parsons, 1982; Rowley, 2002]. Even with the identical rate of seafloor generation at mid-ocean ridges, different area-age relations are possible depending on how seafloor is consumed by subduction, thereby allowing different predictions for the total oceanic heat flow. Considering an extreme transition from a triangular distribution to a “rectangular” one (i.e., the area of seafloor is constant independent of its age), Labrosse and Jauwart [2007] deduced that the 25% variation would be possible. Such a drastic change in the seafloor age distribution, however, should modify the capacity of global ocean basins and should be reflected in the global sea level. Currently available constraints on the past sea level changes indicate that the possible amplitude of heat flux fluctuation is only a few percent at most and that the present-day heat flux is likely to be at its minimum (not the maximum) [Korenaga, 2007]. A similar conclusion was also reached by Loyd et al. [2007], who directly estimated the temporal variation of oceanic heat flux based on plate reconstruction. Thus, the low Urey ratio summarized above remains a robust observational constraint.

3.2. Thermal Catastrophe and Classical Evolution Models

[45] Before the advent of plate tectonics, calculating the thermal history of Earth was based solely on thermal conduction [e.g., Lord Kelvin, 1863; Jeffreys, 1924; MacDonald, 1959], but as the concept of mantle convection started to be widely accepted in the 1970s, the so-called “parameterized convection” approach emerged, in which thermal evolution is modeled based on convective heat loss [e.g., Sharpe and Peltier, 1978; Schubert et al., 1980; Davies, 1980]. A thermal history may be calculated by considering the following energy conservation equation for the entire planet:

$$\frac{dC}{dt} = H(t) - Q(t),$$  \hspace{1cm} (8)

where $t$ is time, $C$ is the heat capacity of the whole Earth ($\sim 7 \times 10^{27}$ J K$^{-1}$ [Stacey, 1981]), $T_i$ is average internal temperature, $H$ is internal heat generation, and $Q$ is convective heat loss. Imbalance between heat generation and loss leads to a secular change in the internal heat content of Earth. More elaborate formulations, such as considering the mantle and the core separately, are possible (section 3.5), but equation (8) suffices for discussion here. Once the present-day concentration of the heat-producing elements is specified, calculating the heat generation term is straightforward as it involves only radiogenic decay. The heart of parameterized convection lies in how to specify the heat loss term $Q$. Mantle viscosity is known to be strongly temperature-dependent, so hotter mantle is less viscous and flows more easily. As faster convection means higher heat flux, one may expect that the convective heat flux can be parameterized as a function of internal temperature:

$$Q(T_i) \approx a T_i^m.$$  \hspace{1cm} (9)

Here $m$ is a scaling exponent, and the constant $a$ is adjusted so that equation (9) predicts the present-day heat flow with the present-day internal temperature. Activation energy for temperature-dependent viscosity is on the order of a few hundred kilojoules per mole [e.g., Weertman and Weertman, 1975; Karato and Wu, 1993], which corresponds to $m \sim 10$ in (9) [e.g., Davies, 1980]. Thus, convective heat loss is extremely sensitive to a change in internal temperature, and this sensitivity turns out to place a tight constraint on the acceptable range of internal heat production. Consider, for example, integrating equation (8) backward in time starting with the present time. When heat production is less than heat flux, it would predict a hotter mantle in the past, which releases more heat. Note that a concurrent increase in heat production takes place only gradually because of the long half-lives of relevant radiogenic elements. Therefore, if the present-day heat production is too low, the increase in convective heat flux overwhelms the increase in heat production, and this imbalance is quickly aggravated, resulting in an unacceptably hot Earth in the not-so-distant past (e.g., Figure 4a, present-day Urey ratio (Ur(0)) = 0.3). By exploring a range of parameters, Schubert et al. [1980] and Davies [1980] concluded that the present-day convective Urey ratio must be greater than $\sim 0.7$ in order to avoid this “thermal catastrophe.” (Note that Davies actually concluded that the ratio was likely to be $\sim 0.5$ in his paper, but as Richter [1984] later pointed out, the value should be corrected to $\sim 0.9$.) With this high Urey ratio, the predicted rate of secular cooling becomes moderate (i.e., $50-100$ K Ga$^{-1}$). Predicted heat flow in the Archean is $\sim 100$ TW, greater than the present-day value by a factor of 3, which
implies that the mantle was convecting an order of magnitude faster then (Figures 4b and 4e).

[46] Given that the BSE composition was not quite established 3 decades ago, it is understandable that this approach was regarded as a geophysical means to estimate internal heat production. In fact, some cosmochemists then were willing to use the result of parameterized convection to estimate the BSE composition [e.g., Morgan and Anders, 1980]. A conflict with geochemical constraints on heat production was, however, gradually appreciated, and two different resolutions were proposed in the mid-1980s. Richter [1985] suggested that layered-mantle convection could reduce the degree of secular cooling because it is a much less efficient cooling mechanism compared to whole mantle convection. Whereas thermal catastrophe could be avoided in the upper mantle, however, the lower mantle would still suffer from catastrophic melting unless it is very depleted in heat-producing elements [Spohn and Schubert, 1982; Honda, 1995]. Alternatively, Christensen [1985a] argued that the heat flow scaling (equation (9)) may not be applicable to mantle convection. On the basis of a series of numerical studies on convection with variable viscosity [Christensen, 1984a, 1984b, 1985b], he suggested that the exponent $m$ in equation (9) may be substantially smaller than a commonly used value, and if so, this weakly temperature-dependent heat flow could prevent thermal catastrophe even with a low Urey ratio. His scaling argument was, however, later questioned by Gurnis [1989], and a more complete study on convection with temperature-dependent viscosity [Solomatov, 1995] showed that Christensen’s argument is only valid for moderately temperature-dependent viscosity and does not hold for a stronger dependency expected for Earth’s mantle.

Figure 4. Examples of classical thermal evolution models (using equation (9) with $m = 6.5$, see [Korenaga, 2006] for details): (a) internal temperature, (b) convective heat flux, (c) internal heat production, (d) convective Urey ratio, (e) plate velocity, and (f) processed mantle mass normalized by whole mantle mass. Integration of equation (8) is done backward in time for the last 4 Ga, starting with the internal temperature of 1350°C and the heat flow of 36 W. Cases with three convective Urey ratios are shown: 0.3 (solid curve), 0.7 (dashed curve), and 0.75 (dotted curve). Note that model predictions are very sensitive to a chosen value of the convective Urey ratio.
[47] Thus, both attempts to reconcile the conflict with geochemistry were not particularly successful, and at the same time, the whole mantle models with a high Urey ratio somehow gained the status of a standard model [Schubert et al., 2001], being applied to a wide range of problems such as continental growth [e.g., Schubert and Reymer, 1985; Franck and Bouhama, 1997], the global water cycle [e.g., McGovern and Schubert, 1989; Williams and Pan, 1992], and the thermal history of the core [e.g., Stevenson et al., 1983; Stacey and Loper, 1984; Davies, 1993]. However, our understanding of the BSE, CC, and DM compositions has considerably improved in the last decade or so, and especially when whole mantle convection is assumed, the BSE composition is well defined. Whether convection in the past was faster or not, the convective Urey ratio today is likely to be around 0.2 and should not exceed ~0.4 (section 3.1). Even in recent years, this important constraint on the Urey ratio is sometimes ignored [e.g., McNamara and van Keken, 2000; Davies, 2007], but how to construct a reasonable thermal history consistent with geochemistry is a fundamental issue to be resolved. Recent attempts are thus examined next in detail.

3.3. Review of More Recent Evolution Models

3.3.1. Variations on Layered-Mantle Models

[48] As mentioned in section 2.1.3, the endothermic phase change expected at the 660-km discontinuity was once considered a promising mechanism to achieve the particular kind of layered mantle favored by geochemists (Figure 3b), and numerical studies on mantle convection suggested that layering would be possible if the Clapeyron slope is more negative than ~6 MPa K⁻¹ [Christensen and Yuen, 1984, 1985]. Later, experimental studies indicated a less negative (but still significant) Clapeyron slope for this phase change (~−2.8 MPa K⁻¹) [e.g., Ito and Takahashi, 1989], and this reinvigorated an interest in the role of the endothermic phase change in mantle dynamics [e.g., Machel and Weber, 1991; Tackley et al., 1993; S. Honda et al., 1993; Solheim and Peltier, 1994; Yuen et al., 1994]. These studies pointed to the possibility of episodically layered convection, and in particular, Yuen et al. [1994] suggested that the degree of layering may increase with an increasing Rayleigh number, which implies that the mantle could have been layered when the vigor of convection was greater because of hotter and less viscous mantle. Motivated by this, Honda [1995] explored the implications this transition with parameterized convection models, but his conclusion is very similar to that of Christensen [1984a]; even with the transition, an unconventional heat flow scaling must be assumed to satisfy the Urey ratio constraint. Davies [1995] investigated the consequence of episodic layering on thermal history but with the present-day Urey ratio of >0.8. More recently Butler and Peltier [2002] developed a more elaborate parameterization for time-dependent layering, and they found that a gradual (not sudden) transition from layered to whole mantle convection could help to satisfy the Urey ratio constraint while keeping a conventional heat flow scaling.

[49] Recent experimental studies indicate, however, that the Clapeyron slope of the endothermic phase change is likely to be much less negative than previously thought (now estimated to be around ~1.3 MPa K⁻¹) [Katsura et al., 2003; Fei et al., 2004]. Furthermore, the role of the endothermic phase change tends to be overemphasized in a majority of previous numerical studies by the use of unrealistic mantle viscosity (such as constant viscosity [e.g., Yuen et al., 1994]); convection is characterized by short wavelengths, lacking long-wavelength structures associated with plate tectonics. This is important because the degree of layering depends not only on the Clapeyron slope but also on the wavelength of convective flow [Zhong and Gurnis, 1994; Tackley, 1995]; large and stiff plates prefer whole mantle convection. On the basis of the chemical buoyancy of oceanic plates, Korenaga [2006] argued that plate dimensions in the past were unlikely to be very different from the present, implying a diminishing chance for time-dependent layering.

3.3.2. Variations on Whole Mantle Models

[50] As the continental crust today probably contains about 40% or more of total heat-producing elements in the bulk silicate Earth (section 2.1.2), its evolution in the past must have affected the thermal history of Earth to some degree. Also, continents are commonly thought to “insulate” Earth’s mantle [e.g., Anderson, 1982; Gurnis, 1988], so they may reduce the degree of secular cooling. Spohn and Breuer [1993] investigated these two effects by extending the parameterized convection model of Stevenson et al. [1983] and suggested that the influence of continental growth would be minor, reducing the predicted value of the present-day Urey ratio by only ~0.1. Various growth scenarios simulated by Spohn and Breuer [1993] are, however, all similar to the so-called “instantaneous” growth model of Armstrong [1981], in which the present-day continental volume is achieved in the early Archean and is then maintained by a perfect balance between the creation of new crust and the recycling of old crust. Grigné and Labrosse [2001] revisited this issue with a more flexible parameterization of continental growth, but the other aspects of their modeling are rather tightly fixed. In particular, their modeling strategy was built solely on the heat flow scaling law developed for an isoviscous fluid [Sotin and Labrosse, 1999], and other possibilities were not explored. The limited parameter search impairs the robustness of their conclusions about a likely scenario for continental growth. They also modeled continents as perfect insulators, but whether continents actually insulate the mantle in terms of global heat loss has recently been called into question [Lenardic et al., 2005]. Given the uncertainty involved, thermal history calculations are unlikely to place a tight constraint on continental growth [e.g., Lenardic, 2006]. Nonetheless, incorporating a plausible growth history is important to improve the accuracy of a predicted thermal history because it definitely affects the heat production of the convecting mantle. The history of continental growth and recycling is still highly controversial among geochemists [e.g., Collerson and Kamber, 1999; Campbell, 2003;
Harrison et al., 2005; Hawkesworth and Kemp, 2006), but if a consensus is ever achieved, thermal evolution modeling may become useful for inferring Archean and Hadean dynamics [e.g., Korenaga, 2006].

[51] The strongly temperature-dependent nature of the heat flow scaling law (equation (9) with $m \sim 10$) originates in high activation energies measured for the deformation mechanisms of olivine, which is stable only in the upper mantle (i.e., above the 410-km discontinuity). Whereas the average viscosity of deeper mantle under present-day conditions can be inferred from geophysical observations such as the geoid and postglacial rebound [e.g., Hager et al., 1985; Mitrovica and Forte, 1997], its temperature dependency must be estimated through experimental studies, which are still difficult to conduct at lower mantle conditions [e.g., Karato et al., 1995]. Also, we can only speculate at this point how different deformation mechanisms, such as diffusion and dislocation creep, might compete in the lower mantle [Karato, 1998], and van den Berg and Yuen [1998] suggested that, under certain conditions, heat flow scaling for the lower mantle may resemble that of Christensen [1985a]. Alternatively, Solomatov [1996] suggested that if the lower mantle is mainly deformed by diffusion creep, which is known to be very sensitive to grain size, the kinetics of grain growth may reverse the temperature dependency of lower mantle viscosity; that is, hotter mantle may become stiffer because grains would grow faster at higher temperatures. Solomatov [2001] pursued this idea with parameterized convection models, but his conclusions were unavoidably equivocal because the kinetics of grain growth itself is still under debate [e.g., Yamazaki et al., 1996; Solomatov et al., 2002] and how grain growth should interact with convective deformation is yet to be understood [e.g., Hall and Parmentier, 2003; Bercovici and Ricard, 2005]. Recently, however, Korenaga [2005a] pointed out that the present-day plume dynamics may provide an observational support for this grain-size-dependent rheology.

[52] Yet another mechanism to delay secular cooling was proposed by van den Berg and Yuen [2002] based on variable thermal conductivity. As temperature increases, the thermal conductivity of mantle minerals first decreases rapidly and then might increase gradually owing to the competing effect of conduction and radiation processes [Hofmeister, 1999], and this temperature dependency can result in a low-conductivity zone in the shallow upper mantle (though this dependency is controversial [cf. Shankland et al., 1979]). van den Berg et al. [2004] extended their earlier study with constant viscosity to temperature- and pressure-dependent viscosity and simulated the thermal evolution of Earth. In their studies, the case of variable conductivity is compared with that of constant conductivity to demonstrate less efficient cooling for the former, but this is simply because the constant conductivity they used corresponds to the surface value of variable conductivity (i.e., the highest conductivity). Note that the heat flow scaling of the kind of equation (9) does not assume a particular value of thermal conductivity because the absolute value of its prediction is normalized with the present-day heat flux. The relative variation in predicted heat flow (as a function of internal temperature) is what matters most and is governed largely by the temperature dependence of mantle viscosity, compared to which the effect of variable conductivity is negligible. A moderate degree of secular cooling observed in the simulation of van den Berg et al. [2004] is because mantle viscosity is assumed to be weakly temperature-dependent in their calculation.

[53] Recently, by combining several geological and geochemical arguments, Silver and Behn [2008] proposed that the efficiency of plate tectonics may have experienced large fluctuations (as much as an order of magnitude) on a billion-year timescale and suggested that such intermittent plate tectonics may help to avoid thermal catastrophe even with the conventional heat flow scaling. Their heat flow parameterization, however, fails to account for the significance of slow thermal diffusion and the role of convective instability in the limit of stagnant lid convection, and their solution for thermal evolution can be shown to be physically unrealistic. They suggested an order-of-magnitude reduction in heat flux at $\sim 1$ Ga ago, which should be accompanied by an order-of-magnitude increase in the average thickness of the top thermal boundary layer (or oceanic lithosphere); lithosphere would then occupy the entire upper mantle. Given our understanding of upper mantle rheology, convective instability would certainly prevent such thick lithosphere, and even if the mantle is rigid and allows the indefinite growth of the boundary layer, it would take as long as about 2 Ga and not be instantaneous as assumed by Silver and Behn [2008]. Intermittent plate tectonics by itself, even if true, does not help to solve the long-standing paradox in Earth's thermal evolution, and available geological records for the Phanerozoic eustasy [Algeo and Seslavinsky, 1995; Miller et al., 2005] indicate that its influence on surface heat flux should be small [cf. Korenaga, 2007].

3.3.3. Plate Tectonics and Whole Mantle Convection

[54] Efforts to reconstruct thermal evolution by starting from establishing heat flow scaling appropriate for Earth’s mantle may be called a physical approach; its underlying principle is that a better understanding of the physics of mantle convection should also help to build a reasonable thermal history. In contrast, one could also infer an acceptable form of heat flow scaling simply by imposing observational constraints on thermal evolution. For example, because thermal catastrophe results from the strongly temperature-dependent heat flow scaling, it is obvious that such an unrealistic history can be avoided if convective heat flow does not depend much on internal temperature. A similar conclusion can also be obtained using the argon budget [Sleep, 1979; Tajika and Matui, 1993]. Conclusions drawn from this empirical approach are essentially the same as those of Christensen [1985a], and the only difference between these two approaches is whether one strives to understand the physics behind heat flow scaling. Labrosse and Jaupart [2007] have recently revived the empirical approach by noting that there is no theoretical framework currently available to predict an area-age relation for sea-
floor, which is essential to obtain an accurate estimate of convective heat flux. The empirical approach, however, has no predicting power because a physical mechanism is left unresolved. Convection in Earth’s mantle is indeed a complex phenomenon, and it may be tempting to get discouraged, but it would be unwise to settle for an ad hoc solution by bypassing the important physics.

[55] The operation of plate tectonics is one of the unique features of Earth, but it is not fully understood why this special kind of mantle convection can take place to begin with [e.g., Tackley, 2000a; Schubert et al., 2001; Bercovici, 2003]. With the strongly temperature-dependent viscosity of Earth’s mantle, the most likely style of thermal convection should be stagnant lid convection [Solomatov, 1995], in which convection takes place only beneath a rigid spherical shell. On Earth, however, the surface is broken into a number of rigid plates, which can recycle back to deep mantle by subduction. Besides temperature-dependent viscosity, therefore, there must be some additional mechanism to weaken the otherwise stiff surface, and how to simulate plate tectonic convection by introducing some kind of self-sustaining weakening has been intensively studied particularly since the mid-1990s [e.g., Bercovici, 1996; Moresi and Solomatov, 1998; Trompert and Hansen, 1998; Tackley, 2000b; Richards et al., 2001; Auth et al., 2003; Gurnis et al., 2004; Stein et al., 2004; Bercovici and Ricard, 2005].

[56] Along with these studies, our understanding of the likely energy balance in plate tectonic convection has been gradually improved. Solomatov [1995] first noted that the deformation of a cold and strong boundary layer (i.e., the bending of a subducting plate) must accompany a significant fraction (~50%) of total energy dissipation, and this was investigated in further detail by Conrad and Hager [1999a, 2001]. On the basis of the notion that plate bending may be the bottleneck of plate tectonics, Conrad and Hager [1999b] explored its implications for heat flow scaling and suggested that a temperature-independent scaling may be possible. Their scaling, however, depends on assumed plate thickness, to which bending dissipation is very sensitive. The maximum thickness of oceanic plates is usually considered to be ~100 km [e.g., Parsons and Sclater, 1977; Stein and Stein, 1992; McKenzie et al., 2005], but this is merely the present-day value. To what thickness oceanic plates can grow under different conditions was not understood well until recently [e.g., Korenaga and Jordan, 2002, 2003; Huang et al., 2003; Zaranek and Parmentier, 2004]. On the basis of this better understanding of plate thickness variation, Korenaga [2003] revisited the scaling of Conrad and Hager [1999b] and found that bending dissipation alone would not reduce the temperature dependence because plates should get thinner for a hotter mantle. Plate tectonics is, however, accompanied by mantle melting beneath divergent boundaries, and this melting strengthens the depleted mantle lithosphere because of dehydration stiffening [e.g., Hirth and Kohlstedt, 1996]. As a hotter mantle starts to melt at a greater depth, this results in a thicker depleted lithosphere [e.g., McKenzie and Bickle, 1988]; therefore, it is possible to expect a thicker plate and thus lower heat flux for a hotter mantle, reversing the sense of conventional temperature dependency [Korenaga, 2003]. The implications of this scaling for the thermal evolution of Earth have been explored extensively, and it is shown to be compatible with a range of existing geological, geophysical, and geochemical observations [Korenaga, 2006; Lyubetskaya and Korenaga, 2007b].

[57] Reconstructed thermal histories according to the formulation of Korenaga [2006] are shown in Figure 5 for the convective Urey ratio of 0.23 ± 0.15, with and without continental growth. To be consistent with the uncertainty analysis in section 3.1, the convective heat flux and the continental heat production at present are set to 36 and 10 TW, respectively, for the Urey ratio of 0.08 and to 37 and 5 TW for the Urey ratio of 0.38. Linear continental growth starting at 4 Ga ago and continuing to the present is assumed. There is no reason to prefer this particular growth history over other possibilities; it is used here for illustrative purposes only. As in Figure 4, the thermal histories are integrated backward in time. The uncertainty of the Urey ratio (0.08–0.38) results in an ~200 K difference at 3 Ga ago (Figure 5a). Though this temperature ambiguity is certainly nontrivial, one might also see it as surprisingly small given the large uncertainty of the Urey ratio (see/compare Figure 4a). This is because of a negative feedback originating in the heat flow scaling of Korenaga [2003], which also enables stable predictions for heat flow, plate velocity, and mantle processing rate in the past (Figures 5b, 5e, and 5f). Note that the hotter and stiffer lower mantle proposed by Solomatov [1996, 2001] would reinforce this negative feedback. The effects of continental growth can be seen as minor, as already indicated by previous studies [e.g., Spohn and Breuer, 1993].

[58] Recently, Labrosse and Jaupart [2007] suggested that almost all of previous heat flow scaling laws obtained by the physical approach do not apply to Earth’s mantle because they are based on a single convection cell model and thus fail to capture the observed area-age distribution of seafloor. This argument is, however, not consistent with the nature of heat flow “scaling” in parameterized convection. Heat flow scaling is usually normalized such that it correctly yields the present-day heat flow with the present-day internal temperature, and what the scaling tries to predict is how this heat flux would change when the internal temperature is varied. In other words, the present-day area-age distribution is implicitly adopted as a reference convective planform, and its sensitivity to temperature perturbations is assessed by a simplified convection model. Whether this approach is too simplified to be useful is debatable. The conventional scaling of Schubert et al. [1980] and Davies [1980], for example, usually predicts very high heat flux and rapid plate tectonics in the past (Figures 4b and 4e), implying corresponding flow geometry that may be vastly different from today. In this case, deviations from the reference may be too large for a simple scaling to be useful. Conversely, the scaling of Korenaga [2006] indicates that the style of plate tectonics may not have changed much in
the past (Figures 4b and 4e), which provides a posteriori justification for this scaling.

[59] The normalization with the present-day heat flux implies that the derived scaling is useful only for plate tectonic convection, so the operation of plate tectonics has to be assumed when reconstructing a thermal history. Plate tectonics was probably present at least back to the late Archean (~3 Ga ago) [e.g., Hoffman, 1997], but the emergence of plate tectonics in Earth’s history is not understood well. The very early Earth is likely to have been governed by a different mode of mantle convection [e.g., Sleep, 2000; Stevenson, 2003]. It is thus preferred to integrate a thermal history backward from the present. As emphasized by Korenaga [2006], a common forward integration is not recommended because it has to assume the same physics of convection throughout Earth’s history; a more recent part of the history is unavoidably affected by this strong assumption of Hadean and Archean dynamics.

[60] The backward reconstruction of a thermal history, as shown in Figure 5, is in a sense promising because it may provide important boundary conditions on mantle dynamics in the early Archean and the Hadean. The negative feedback suggested by the recent heat flow scaling already makes robust predictions for heat flow and plate velocity in the past. Through its unifying framework, thermal evolution modeling assimilates a variety of geophysical and geochemical observations and converts them to quantitative constraints on the distant past. A better understanding of the

---

**Figure 5.** Thermal evolution modeling with the heat flow scaling of plate tectonic convection [Korenaga, 2006]: (a) internal temperature, (b) convective heat flux, (c) internal heat production, (d) convective Urey ratio, (e) plate velocity, and (f) processed mantle mass normalized by whole mantle mass. As in Figure 4, integration is done backward in time starting with the present-day temperature of 1350°C. The convective heat flux and the (optional) continental heat production at present are set to 36 and 10 TW, respectively, for the case with the Urey ratio of 0.08 (solid curve), and 37 and 5 TW, respectively, for the case of the Urey ratio of 0.38 (dashed curve). Cases with continental growth are shown in gray; linear continental growth starting at 4 Ga and continuing to the present is assumed. Compared to classical models (Figure 4), model predictions are much less sensitive to a chosen value of the convective Urey ratio.
physics of plate tectonics, a more accurate estimate on the heat production of the convecting mantle, and a less ambiguous history of continental growth will all help to sharpen this theoretical inference.

3.4. On the Acceptable Range of Secular Cooling Rate

[61] The success of a reconstructed thermal history has been judged mostly on the basis of its secular cooling rate. Other aspects such as heat flux and plate velocity are usually difficult to test because of the paucity of convincing observational constraints (though they are sometimes useful [e.g., Korenaga, 2007]). It is easy to reject the case of thermal catastrophe (e.g., Figure 4a, \( U_r(0) = 0.3 \)) because of its divergent nature in the middle of Earth’s history, but how to distinguish between much less extreme candidates is more uncertain. On the basis of the freeboard constraint (i.e., the mean sea level has stayed relatively constant since the Archean), for example, Galer [1991] and Galer and Mezger [1998] argued that the secular cooling rate should be \( \sim 50 \text{ K G}^{-1} \), but this value rests on various factors such as the growth of continental crust, the history of ocean volume, and the evolution of subcontinental lithosphere, all of which are still subject to large uncertainty. In particular, the relation between mantle temperature and the mean depth of ocean basins has to be assumed in this type of calculation, but this relation depends on the adopted heat flow scaling of mantle convection [e.g., Reym er and Schubert, 1984].

[62] In principle, petrology can offer more direct constraints on the secular evolution of mantle temperature, which is likely to be reflected in the composition of igneous rocks at various ages. Historically, the igneous petrogenesis of komatiites has been the prime constraint on mantle potential temperature in the Archean [e.g., Richter, 1985] because their occurrence is mostly restricted to Archean provinces. The origin of komatiites, however, has been widely debated, and depending on which tectonic environment is involved, the estimate of potential temperature varies from very hot (>1800°C, in the case of hot spot magmatism [e.g., Bickle et al., 1977; Herzberg, 1992; Walter, 1998; Arndt, 2003]) to hot (∼1650°C, mid-ocean ridge magmatism [Kelemen et al., 1998]) to only moderately hot (∼1500°C, arc magmatism [e.g., Allegre, 1982; Grove and Parman, 2004]). Moreover, the temperature of source mantle for komatiites does not have to be equal to globally averaged mantle temperature, especially for the hot spot and arc origins [e.g., Jarvis and Campbell, 1983; Campbell and Griffiths, 1992]. Abbott et al. [1994] took a complementary approach and focused on (more abundant) MORB-like rocks to obtain a more continuous constraint on secular cooling. Their estimate on the average cooling rate is ∼70 K G\(^{-1} \) for the last 3 Ga, but because of data scatter, this estimate has a large uncertainty of at least ±30 K G\(^{-1} \).

[63] A new kind of constraint was recently proposed based on the petrological constraint on early Earth dynamics [Labrosse and Jaupart, 2007; Jaupart et al., 2007]. Numerical studies indicate that a transition from an early magma ocean to predominantly subsolidus mantle convection may have taken place when the solid fraction near the surface reached ∼60% [e.g., Abe, 1997]. As the magma ocean is likely to have existed only at the very beginning of Earth’s history, a critical mantle temperature corresponding to this transition can constrain the average rate of secular cooling over the last 4.5 Ga or so. Because the published phase diagrams for Earth’s mantle show that the melt fraction of 40% at the surface is achieved when the potential temperature is ∼200 K higher than the present value [e.g., McKenzie and Bickle, 1988; Zhang and Herzberg, 1994], Labrosse and Jaupart [2007] identified this temperature (∼1550°C) as the critical temperature and argued that Earth must have cooled by only ∼200 K after the magma ocean ceased to exist. This implies that the average cooling rate must be lower than ∼45 K G\(^{-1} \), which is a remarkably tight constraint. Their interpretation of the phase diagram, however, assumes closed-system (batch) melting, in which a melt phase does not move with respect to a residual solid phase. Because of density difference and because of the formation of porous network, the melt is expected to migrate upward, and the compaction of the solid phase results in a very low porosity [e.g., McKenzie, 1984; Spiegelman, 1993]. This so-called fractional melting in Earth’s mantle is well supported by both geochemical and geophysical observations [e.g., Johnson et al., 1990; Sobolev and Shimizu, 1993; Kelemen et al., 1997; Toomey et al., 1998; Evans et al., 1999]. As a familiar example, consider the Iceland hot spot. An excess temperature of ∼200 K or so has commonly been proposed for this hot spot [e.g., White et al., 1995], but no one expects the presence of a partially molten mantle with as much as 40% melt fraction beneath Iceland [e.g., Ito et al., 1999]. One cannot simply read the critical temperature from the phase diagram; it must involve a better understanding of melt generation and migration with realistic mantle rheology. The approach based on the critical temperature will become important when our understanding of early Earth dynamics is substantially improved, but at present it does not yield a useful constraint.

[64] Therefore, the average rate of secular cooling is only loosely constrained to be ∼50–100 K G\(^{-1} \) over the last 3 Ga, which is satisfied by most of the models shown in Figure 5a. Note that the cooling rate can be time-dependent. As Figure 5a indicates, the present-day cooling rate may exceed 100 K G\(^{-1} \), but it does not have to be representative of a long-term average. Decent observational constraints exist only for this average cooling rate.

3.5. A Note on Core Cooling

[65] The global energy balance may be described explicitly with the mantle and core components as [e.g., Stevenson et al., 1983]

\[
C_m \frac{dT_m}{dt} = H_m(t) - Q(t) + Q_s(t)
\] (10)

\[
C_c \frac{dT_c}{dt} = (L + E) \frac{dm_e}{dt} + H_c(t) - Q(t).
\] (11)

where \( C_m \) and \( C_c \) are the heat capacities of mantle and core, \( T_m \) and \( T_c \) are average mantle and core temperatures, \( H_m \) and
$H_c$ are internal heat production in the mantle and the core, and $Q_c$ is the heat flux from the core into the mantle. The mass of the inner core is denoted by $m_{ic}$, and the first term on the right-hand side of equation (11) describes the release of latent heat and gravitational energy associated with the growth of the inner core.

Discussion on internal heat production in this paper has so far been restricted to the heat production of BSE (i.e., the continental crust and the depleted mantle), thus assuming $H_c \approx 0$. This assumption is probably reasonable. Uranium and thorium are both lithophile even at high temperatures and pressures [Wheeler et al., 2006]. The K content of the core is more debatable [e.g., Lewis, 1971; Bukowinski, 1976; McDonough, 2003; Lassiter, 2006]. Note that the successful global mass balance regarding potassium [Lyubetskaya and Korenaga, 2007b] does not exclude the possibility of a K-rich core because potassium is not a refractory lithophile element and the estimate of BSE K content is based on the K/U ratio of silicate rocks (∼10<sup>6</sup> [e.g., Wasserburg et al., 1964; Jochum et al., 1983]). The bulk Earth K/U ratio can be different from this ratio [Humayun and Clayton, 1995], so the core K abundance is out of the scope of global mass balance. In fact, potassium may enter the core at high pressures [e.g., Parker et al., 1996; Lee et al., 2004b] or through sulfide liquid segregation [e.g., Gessmann and Wood, 2002; Ramu Murthy et al., 2003]. Lassiter [2006] pointed out, however, that such behavior of potassium should also be accompanied by other elements with similar chemical nature, which is not supported by existing geochemical data.

With $H_c \approx 0$, the thermal evolution of the core is always controlled by that of the mantle and not the other way around. The core is cooled only passively by the cooling of the overlying mantle; the energy release due to the inner core growth is not an independent heat source because the growth originates in the cooling of the core. Even with some heat production in the core, the significance of this mantle control is unquestionable. With a few exceptions [e.g., Yukutake, 2000], however, previous studies on core evolution tend to overlook the controversy over the parameterization of the surface heat loss $Q$ (sections 3.2 and 3.3) and assume the conventional scaling (usually with a high Urey ratio) [e.g., Labrosse et al., 2001; Nimmo et al., 2004; Davies, 2007]. In order to integrate the set of equations (10) and (11), we also need to parameterize the core heat flux $Q_c$, which depends critically on the (poorly known) lower mantle rheology (section 2.2.1). By modeling the coupled core-mantle thermal evolution, some attempted to evaluate the success of a given model by using its prediction for core-related observables such as the present-day radius of the inner core and the history of the geomagnetic field [e.g., Nimmo et al., 2004], but less attention was paid to the uncertain nature of heat flux parameterization. Directly simulating mantle convection by numerical modeling [e.g., Nakagawa and Tackley, 2005; Butler et al., 2005] does not offer a fundamental solution to this issue unless realistic secular cooling is achieved without assuming a high Urey ratio. Moreover, in order to take into account core-related observables, we have to face other sources of uncertainty such as the amount of ohmic dissipation required for the geodynamo [e.g., Buffett, 2002; Roberts et al., 2003] and the thermodynamic and chemical properties of the core [e.g., Davies, 2007].

The energy release due to the inner core growth would be on the order of ∼2 TW depending on the age of the inner core [e.g., Stevenson et al., 1983], so by neglecting this in addition to heat production in the core, the simple heat balance of equation (8) may be recovered by summing equations (10) and (11), if $dT_c/dt \approx dT_m/dt$. Within this simplified framework, the present-day core heat flux may be estimated as

$$Q_c(0) \approx \frac{C_v}{C_v + C_m} Q(0)[1 - \gamma(0)],$$

where $\gamma$ is the convective Urey ratio. The discussion in section 3.1, together with the heat capacities from Stacey [1981], implies that $Q_c \sim 4.5 \pm 0.8$ TW, comparable to the estimated conductive heat flux along the core adiabat [e.g., Braginsky and Roberts, 1995; Davies, 2007]. Other independent estimates of the present-day core heat flux are marginally comparable (6–12 TW from the expected thermal structure around the core-mantle boundary [Buffett, 2003]) or much higher (9–12 TW based on the postperovskite phase transition [Hernlund et al., 2005], ∼13 TW based on numerical simulation [Zhong, 2006], and 10–30 TW from seismic tomography [Nolet et al., 2006]). Energy release due to the inner core growth is too small to explain this discrepancy, and it is unlikely that the core cooling rate is much greater than the mantle cooling rate. Does this then indicate the need for radiogenic elements in the core? Given our incomplete understanding of element partitioning under extreme conditions [e.g., Bouhifd et al., 2007; Corgne et al., 2007], it is still difficult to rule out this possibility. At the same time, these estimates on core heat flux also suffer from unquoted uncertainty; the estimates of Zhong [2006] and Nolet et al. [2006], for example, depend on their assumption about lower mantle rheology.

Whereas the first-order features of the thermal evolution of Earth (or the mantle) can be modeled reasonably well by parameterized convection using equation (8), the thermal evolution of the core should be regarded as a second-order feature, depending on a larger number of unknowns. Of course, understanding core evolution is still an important goal, but the true power of core constraints on the thermal evolution of Earth may only be revealed by exploring the wide range of permissible model parameters as well as by taking into account the realistic uncertainty of observational constraints. Note that considering core evolution with layered-mantle convection is even more challenging if not hopelessly arbitrary.

4. PRESENT-DAY THERMAL BUDGET OF EARTH: A SUMMARY

On the basis of sections 2 and 3, the present-day thermal budget may be summarized as in Table 1. Heat
generation by tidal dissipation is neglected here because it is estimated to be very small (~0.1 TW [e.g., Verhoogen, 1980]). The assumption of whole mantle convection is vital because it is impossible to define a thermal budget with layered-mantle convection (section 2.1.3). As discussed in sections 2 and 3, the notion of whole mantle convection is compatible with existing geochemical and geophysical observations. Common geochemical arguments for a layered mantle are seen to rest either on mass balance calculations neglecting the importance of uncertainty (sections 2.1.2 and 2.1.4) or on the strong version of the chondrite assumption, which is below the resolution of terrestrial data (section 2.1.5).

[71] The heat production of the bulk silicate Earth is likely to be in the range of 13 to 19 TW, based on the composition model of Lyubetskaya and Korenaga [2007a]. It is not advisable to merge this estimate with previous (slightly different) estimates to modify the mean value or enlarge the uncertainty. For example, the model of McDonough and Sun [1995] implies the BSE heat production of 20 ± 3 TW, but virtually the same data sets and assumptions are used by these two studies (in particular, chemical elements used in subset 5 of Lyubetskaya and Korenaga [2007a] are exactly the same as those considered by McDonough and Sun [1995], and the result from this subset is virtually identical to those from other subsets, which contain more incompatible elements); the only difference is the rigor of statistical inference. As far as heat-producing elements are concerned, the model of Lyubetskaya and Korenaga [2007a] supersedes previous pyrolite-type models such as those of McDonough and Sun [1995] and Palme and O’Neill [2003], and these different models do not coexist with equal credibility. The estimate of the convective Urey ratio varies widely from 0.08 to 0.38. Equation (8) implies that this corresponds to the secular cooling rate of ~100–150 K Ga⁻¹ at the present day. The rate of secular cooling in the past is, however, controlled by a time-dependent balance between heat production and heat loss, and if convective heat loss is maintained at a relatively uniform level as suggested by some recent heat flow scaling laws [e.g., Solomatov, 2001; Korenaga, 2003], the average rate of secular cooling since the Archean may be reduced to ~50–100 K Ga⁻¹ (see Figure 5a and section 3.4).

[74] This understanding of the present-day heat budget, despite its nontrivial uncertainty, is what allows us to integrate the global energy balance backward in time and to constrain the ancient state of Earth with some confidence. This is unique to Earth; a comparable knowledge does not exist for other terrestrial planets and moons. Reconstructing the thermal history of Earth provides a global context to

| TABLE 1. Present-Day Thermal Budget of Earth |
|---|---|---|
| Global heat flux | 46 | ±3 |
| Oceanic heat flux | 32 | ±2 |
| Continental heat flux | 14 | ±1 |
| BSE heat production | 16 | ±3 |
| CC heat production | 7.5 | ±2.5 |
| DM heat production | 8.5 | ±5.5 |
| Convective heat flux | 36.5 | ±4.5 |
| Convective Urey ratio | 0.23 | ±0.15 |
| Secular cooling rate | 124 | ±22 |
| Core heat flux | 4.5 | ±0.8 |
| Internal heating ratio | 0.87 | ±0.02 |

a Heat flux and heat production are in TW (10¹² W), and secular cooling rate is in K Ga⁻¹. All ratios are dimensionless.
b Some of these uncertainties are correlated, so care must be taken for error propagation (e.g., sections 3.1, 3.3.3, and 3.5).
c These estimates assume whole mantle convection.
d These core-related estimates are more uncertain than others (section 3.5).

One wishes to consider them as competing alternatives (see a review by Lyubetskaya and Korenaga [2007a]).
appreciate the significance of different kinds of observations and to relate them in a self-consistent manner. The interpretation of geochemical data, for example, often involves some assumptions about the history of mantle processing, which should be consistent with thermal evolution (section 2.2.2).

Core heat flux, the age of the inner core, and thus the history of the geomagnetic field all depend on the present and past values of the convective Urey ratio (sections 3.4 and 3.5).

[75] Debates over the structure and evolution of Earth’s mantle seem to have become fixated on a stereotyped pattern of whole mantle versus layered-mantle convection, without sufficient efforts to understand the causes and consequences of whole mantle convection. The operation of plate tectonics has considerable bearing on the chemical structure of the mantle (section 2.2), and the relation between magmatism (geochemistry) and convection (geophysics) needs to be examined carefully for a chemically heterogeneous mantle. Why plate tectonics occurs on Earth is also a fundamental issue, as the history of surface heat loss hinges on the physics of plate tectonic convection (section 3.3.3). Our understanding is still in its infancy regarding these chemical and physical aspects of plate tectonics.

5. DISCUSSION

5.1. Internal Heating Ratio

[76] The classical Rayleigh-Bénard convection [Lord Rayleigh, 1916] is heated entirely from below and has no internal heating. The mode of heating, whether heated from below or internally or somewhere in between, is one of the essential factors that controls the style of thermal convection [e.g., McKenzie et al., 1974; Houseman, 1988; Bunge et al., 1997]. Plume-like convective instabilities, for example, cannot be expected for purely internal heating, in which a bottom thermal boundary layer does not exist. The internal heating ratio, $\xi$, is defined as the difference between heat flux out of the top boundary and heat flux into the bottom boundary normalized by the former, so in terms of mantle convection, it may be expressed as

$$\xi = \frac{Q - Q_m}{Q} = \frac{1}{Q} \left( H_m - C_m \frac{dT_m}{dt} \right).$$

The ratio is zero for purely basal heating and unity for purely internal heating. Note that radiogenic heat production and secular cooling are considered collectively as internal heating here. Modeling transient phenomena such as secular cooling depends on a poorly known initial condition, which makes it difficult to draw general and robust conclusions. In the study of mantle convection, therefore, a numerical model is often run for a number of convective overturns, so that the model convection system reaches a (statistically) steady state, in which time-averaged observables do not noticeably fluctuate and do not depend on initial conditions. This steady state approach has to include secular cooling as a part of internal heating in order to make modeling results relevant to Earth’s mantle.

[77] The discussion in section 3.5 suggests that the present-day internal heating ratio is $\sim 0.87$ (Table 1) but with a large uncertainty because the core heat flux is not tightly constrained. Nonetheless, equation (12) should be approximately valid, which implies the following relation between the internal heating ratio and the convective Urey ratio:

$$\xi(t) \approx 1 - \frac{C_e}{C_m + C_e} [1 - \gamma(t)].$$

A thermal history consideration suggests higher Urey ratios in the past (e.g., Figure 5d), so the internal heating ratio must have been closer to unity than at present.

[78] Thus, the convection of Earth’s mantle thus has been predominantly internally heated throughout its history. When the dynamics of the core-mantle boundary region is studied by numerical modeling, however, internal heating is sometimes neglected or much reduced in order to highlight or focus on processes related to basal heating [e.g., Nakagawa and Tackley, 2004; Matyska and Yuan, 2005]. Whether this sort of simplification can be justified depends on how modeling results are interpreted. Discussing the mantle-wide thermal structure, for example, is not meaningful because such large-scale features are very sensitive to the heating mode. Designing a numerical study relevant to Earth’s mantle may not always be easy, particularly when realistic complications (e.g., mantle rheology, phase transitions, and chemical heterogeneities) have to be incorporated, but it is important to keep in mind that the heating mode is not an arbitrary factor as its likely range is constrained by the thermal history of Earth.

5.2. Urey Ratio: A Historical Perspective

[79] The Urey ratio, whether being bulk Earth or convective, does not have to be close to unity and is, in fact, far from unity at present (section 3.1). However, the notion of “thermal equilibrium” (i.e., $\gamma \approx 1$) appears to have long been entrenched in the geochemical and geological literature [e.g., Urey, 1955; O’Nions and Oxburgh, 1983; Allegre, 2002]. This misconception is not harmless; it can be a source of misleading arguments such as helium-heat imbalance (see discussion by Lyubetskaya and Korenaga [2007b]). Understanding the historical origin of this notion is probably as important as knowing the likely range of the present-day Urey ratio.

[80] It is well known that in the late 19th century the geological community was disturbed by the estimate of Earth’s age by William Thomson (Lord Kelvin), perhaps the greatest physicist at the time [Burchfield, 1975]. Kelvin’s estimate of $\sim 100$ Ma old Earth is based on his interpretation of continental heat flow data in terms of the conductive cooling of an initially hot sphere [Lord Kelvin, 1863]. Surface heat flux from such a simple cooling system is inversely proportional to $\sqrt{t}$ [Turcotte and Schubert, 1982], and a fairly young Earth had to be assumed to explain the measured heat flow. The discovery of radioactivity at the turn of the century provided a new kind of energy source,
and continental crust was then found to contain enough radiogenic elements to support surface heat loss. The notion of thermal equilibrium for Earth’s heat loss was probably born around this time, and for continents, it is approximately valid even today; most of continental heat flux originates in radiogenic elements in continental crust (section 2.1.2). In the 1950s, heat flow measurements at ocean basins showed that oceanic heat flux was comparable to continental heat flux, which was puzzling because oceanic crust was already known to be much more depleted in radiogenic elements. It was before the theory of plate tectonics was born, and mantle convection was not seriously considered then. It is thus understandable that people speculated that suboceanic mantle may contain sufficient radiogenic elements to balance the high oceanic heat flow [e.g., Bullard, 1954].

[Urey [1955, 1956] pointed out that the abundance of radiogenic elements in chondritic meteorites might be used to argue for thermal equilibrium on a global scale. In particular, Urey [1955, p. 30] described a possible physical mechanism behind such equilibrium as follows:

> Dr. Roger Revelle suggested to me that the rate of heat loss of the Earth is now and perhaps always has been equal to the rate of generation of heat in the Earth. He suggested that a rise in temperature of the mantle will cause more rapid convections and a more rapid dissipation of heat, and a fall in temperature will produce opposite effects until equilibrium is attained again.

This feedback mechanism was later investigated quantitatively by Tozer [1972] with temperature-dependent viscosity, and he derived a relaxation timescale of $\sim 10^8$ years, which may be considered to be short enough compared to the thermal evolution of Earth. Tozer’s argument for a “self-regulating” Earth is intuitively appealing and helped to popularize the notion of thermal equilibrium. Its influence on subsequent studies can be readily recognized [e.g., McKenzie and Weiss, 1975; Bickle, 1986].

The validity of thermal equilibrium was questioned by Schubert et al. [1980], who noted that it would imply no change in internal temperature according to the global energy balance (equation (8)). As heat production should decrease with time (because of radiogenic decay), thermal equilibrium demands that heat loss should also decrease with time even though internal temperature does not change. Schubert et al. [1980] argued that this contradicts convection experiments, which exhibit a positive correlation between internal temperature and heat flux. To prove their point more quantitatively, they constructed parameterized convection models, but as explained in section 3.2, their study indicated that the Urey ratio should be as high as $\sim 0.7$. This value should be sufficient to reject the notion of “exact” thermal equilibrium, but one could also use it to support a crude version of $\gamma \approx 1$. The high-Urey-ratio models of Schubert et al. [1980] and Davies [1980] have deeply pervaded the solid Earth community as is evident from the large number of their variants in the literature (section 3.2).

The popularity of high-Urey-ratio models may also come from their prediction of modest secular cooling at present. Low Urey ratios such as 0.2 indicate the present-day cooling rate of $>100$ K Ga$^{-1}$ (Table 1), which apparenently contradicts geological constraints [e.g., van Keken et al., 2002]. This argument, however, does not acknowledge an important difference between the present-day cooling rate and the average cooling rate (section 3.4). Given these various kinds of historical inertia, it may take a while for a proper understanding of Earth’s thermal budget to prevail.

5.3. Geoneutrinos

The possibility of remotely sensing radiogenic decay processes in Earth’s interior by detecting antineutrinos (equations (1)–(4)) has long been speculated [e.g., Eder, 1966; Krauss et al., 1984; Kobayashi and Fukao, 1991; Rothschild et al., 1998], and the first result by the Kamioka liquid scintillator antineutrino detector (KamLAND) appears promising [Araki et al., 2005; McDonough, 2005]. Neutrino geophysics certainly has a potential to provide a completely new kind of constraint on Earth’s thermal budget, though it will probably take more time and resources for such efforts to yield useful and convincing results. The KamLAND result, for example, emphasizes its statistical upper bound on internal heat production ($\sim 60$ TW), but what is more relevant to debates over the thermal budget is, instead, the lower bound, which is currently zero [Araki et al., 2005; Enomoto, 2006]. Also, the interpretation of antineutrino data is not entirely independent from geochemistry and geophysics. If a detector is located on a continent and if one wants to constrain the heat production of the mantle and core, the heat production of continental crust should be taken into account using regional geochemical and geophysical data (i.e., crustal composition and heat flow). The deployment of a detector on ocean basins (or on oceanic islands) can minimize this type of complexity [Dye et al., 2006]. Antineutrino detection is an expensive endeavor, which naturally limits the likely number of detectors on the globe. A rule of thumb is that the number of model parameters that can be determined is equal to the number of detector locations; if we combine three locations’ data, for example, we may be able to constrain the average heat production of the crust, the upper mantle, and the lower mantle, assuming a spherically symmetric distribution of heat-producing elements. In other words, antineutrino detection alone would not be able to infer the detailed distribution of heat-producing elements, and it is probably best regarded as a means of hypothesis testing, i.e., to discriminate between various hypotheses put forward by geochemists and geophysicists [e.g., Mantovani et al., 2004]. The directionality of geoneutrinos could potentially provide further constraints, but it is difficult to detect [Batygov, 2006]. Note that the resolving power of antineutrino detection decreases as the source depth increases and that the energy level of the antineutrino produced by the decay of $^{40}$K is below the detection limit of currently operating detectors. Constraining the heat production of the lower mantle and the core will thus remain challenging at least for a while.

Neutrino geophysics is probably the ultimate way to test the assumption of whole mantle convection, and even if it disproves the assumption, the heat production of Earth would not be unbounded because the total heat production
will be directly constrained by detector data. The heat production of continental crust and subcontinental lithosphere may also be better constrained, which would be essential to reduce the uncertainty of convective heat flux and thus the convective Urey ratio.

5.4. Outlook: Meaning of the Low Urey Ratio

Despite various uncertainties involved in the preceding discussion on Earth’s thermal budget, it is probably safe to regard the low convective Urey ratio (~0.2) as one of the first-order features of the present Earth. This by itself may not be striking, but if considered together with the likely history of the Urey ratio, it is intriguing to ask why the Urey has become so low. The thermal evolution models with continental growth seem to suggest that the Urey ratio started to fall below unity at ~3 Ga ago (Figure 5d), and this timing seems to be comparable with the emergence of plate tectonics as indicated by the history of supercontinental aggregation [e.g., Hoffman, 1997]. Note that there are some (indirect) geochemical arguments based on zircon data for the operation of plate tectonics as early as ~4.4 Ga ago [e.g., Wilde et al., 2001; Harrison et al., 2005], but the kind of plate tectonics similar to what we observe today may not have started until ~3.5 Ga ago [e.g., Van Kranendonk et al., 2007; Shirey et al., 2008]. Plate tectonics may have been efficiently cooling Earth at a nearly constant rate, regardless of its steadily decreasing heat production. By decoupling surface heat flux from internal heat production, Earth’s thermal state was able to deviate from an equilibrium, resulting in secular cooling. Note that secular cooling is essential for the generation of the geodynamo, and without the geomagnetic field, Earth’s surface water may have evaporated into space [e.g., Hunt and Donahue, 1976; Chassefière, 1997]. A departure from Tozer’s [1972] self-regulating Earth seems to be the essential nature of Earth’s evolution, which may ultimately be responsible for sustaining a habitable planet. This is very speculative because we still do not know the fundamental physics of plate tectonic convection and how mantle dynamics should actually interact with the evolution of the atmosphere, the oceans, the continental crust, and the core. The presence of surface water, for example, is often believed to be essential for the operation of plate tectonics [e.g., Tozer, 1985; Bercovici et al., 2000], but plate tectonics may be responsible for the long-term stability of surface water through core dynamics as mentioned above. Understanding the causes and consequences of plate tectonic convection will be both helped and hampered by the interconnectedness of different components in the Earth system, and this may be what we should expect for the emergence of complexity in the integrated Earth system. Reconstructing a geologically sensible thermal history is only a very small step in this exciting, multidisciplinary enterprise.

GLOSSARY

**Bulk Earth Urey ratio:** Internal heat production within the entire Earth divided by total surface heat flux.

**Bulk silicate Earth (BSE):** Solid Earth excluding the core.

**Chondrites:** Stony meteorites that have been least modified by the melting or differentiation of the parent body. Three major types are carbonaceous, ordinary, and enstatite chondrites, with the first type considered to be the most primitive.

**Chondrite assumption:** Assumption that the relative abundance of refractory lithophile elements in BSE should be close to (but not necessarily identical to) that of chondrites. Note that the ratio of these elements is relatively constant among different types of chondrites, so this assumption does not impose a particular type of chondrites as building blocks for Earth.

**Continental crust (CC):** Uppermost layer of the solid Earth corresponding to continents and continental shelves, with the average composition being andesitic.

**Convective Urey ratio:** Internal heat production within the mantle divided by mantle heat flux.

**Depleted mantle (DM):** Mantle after the extraction of continental crust. On average, this mantle is depleted in terms of ( incompatible) elements enriched in continental crust, but there can be spatial variation in the degree of depletion. Different source mantles for MORB and OIB may represent such spatial variation.

**Depleted mantle lithosphere (DML):** Residual mantle after the extraction of oceanic crust. DML is even more depleted than DM.

**Geoneutrinos:** Neutrinos produced by radiogenic elements in Earth’s interior. The decay chains of all of Earth’s major heat-producing elements (K, U, Th, and 40K) include beta decays, which emit electron antineutrinos.

**Internal heating ratio:** Difference between heat flux out of the top boundary layer and heat flux into the bottom, normalized by the former. At steady state, the ratio is zero for purely basal heating and is unity for purely internal heating.

**Layered-mantle convection:** Notion that Earth’s mantle is divided into two layers, with an impermeable boundary between them. It used to be thought that this boundary corresponded to the 660-km seismic discontinuity.

**Mid-ocean ridge basalt (MORB):** Basaltic rocks found at mid-ocean ridges, representing the top portion of oceanic crust. This term is usually used to signify its relatively depleted nature compared to ocean island basalt.

**Oceanic crust (OC):** Uppermost layer of the solid Earth corresponding to ocean basins, with the average composition being basaltic.

**Ocean island basalt (OIB):** Basaltic rocks found at ocean islands such as Hawaii. They tend to be more enriched than mid-ocean ridge basalt and also to exhibit greater isotopic variabilities.

**Parameterized convection models:** Highly approximated models for the cooling of a planetary body, in which heat flow scaling for convection is parameterized simply as a function of average internal temperature.

**Primitive mantle (PM):** Primordial mantle formed after core formation but before the extraction of continental crust. This is synonymous with BSE.
**Refactory lithophile element (RLE):** Elements that have very high condensation temperature (refractory) and are also favored by silicate phases (lithophile). They include Be, Al, Ca, Sc, Ti, V, Sr, Y, Zr, Nb, Ba, rare earth elements, Hf, Ta, Th, and U.

**Secular cooling:** Gradual loss of the internal heat content (of a planetary body).

**Whole mantle convection:** Notion that Earth’s mantle convects as a single layer. In this model, subducting slabs eventually go all the way down to the core-mantle boundary. Mantle plumes also originate at this boundary.

[88] **ACKNOWLEDGMENTS.** This work was sponsored by the U.S. National Science Foundation under grant EAR-0449517. The author thanks Mark Harrison for pointing out how uncertain the composition of the continental lower crust is. This paper has benefited from thoughtful reviews by Editor Michael Manga, Norm Sleep, Dave Stevenson, and Bill McDonough.

[89] The Editor responsible for this paper was Michael Manga. He thanks technical reviewers Norm H. Sleep and David Stevenson and one anonymous technical reviewer.

**REFERENCES**


Burchfield, J. D. (1975), Lord Kelvin and the Age of the Earth, Univ. of Chicago Press, Chicago, Ill.


Campbell, I. H. (2005), Large igneous provinces and the mantle plume hypothesis, Elements, 1, 265–269.


Rayleigh, Lord (J. W. Strutt) (1916), On convective currents in a horizontal layer of fluid when the higher temperature is on the under side, Philos. Mag., 32, 529–546.


Yukutake, T. (2000), The inner core and the surface heat flow as clues to estimating the initial temperature of the Earth’s core, *Phys. Earth Planet. Inter.*, 121, 103–137.


J. Korenaga, Department of Geology and Geophysics, Yale University, P.O. Box, 208109, New Haven, CT 06520-8109, USA. (jun.korenaga@yale.edu)