EFFECTS OF PHASE TRANSITIONS ON MANTLE CONVECTION

Ulrich Christensen

Institut für Geophysik, Universität Göttingen, Herzberger Landstrasse 180, 37075 Göttingen, Germany

KEY WORDS: seismic discontinuities, numerical modeling, mantle layering, episodicity

INTRODUCTION

Mantle convection—the creeping flow of solid mantle rock driven by buoyancy forces predominantly arising from thermal expansion—is the accepted paradigm for explaining the internal geological activity of the Earth. Its theoretical understanding is largely based on the simple concept of Rayleigh-Bénard convection. Here, one assumes a fluid layer of infinite lateral extent and constant material properties, which is kept at different temperatures at its top and bottom boundaries. Obviously the Earth’s mantle is not so simple, and the consequences of various complications must be considered. A question that has attracted the interest of earth scientists since the early days of plate tectonics is how structural phase changes in the transition zone of the Earth’s mantle between 400 and 700 km depth influence the convective flow. It was often assumed that these phase changes would constitute a serious obstacle for convection currents (e.g. Knopoff 1964). This notion could explain why deep earthquakes, which are associated with subducting lithospheric slabs and hence the sinking branch of mantle convection, do not occur below 700 km, although alternative explanations for the apparent or real failure of slabs to penetrate the lower mantle have been suggested, such as a difference in bulk chemistry. Any restrictions on the mass flux between upper and lower mantle would have important consequences for the thermal and chemical evolution of the Earth and for our interpretation of surface observables such as global geoid anomalies.

The understanding of how phase transitions interact with convection has benefited greatly from numerical model calculations, starting with Richter (1973) and followed by Christensen (1982) and Christensen & Yuen (1984, 1985). These calculations demonstrated that under conditions relevant for the Earth
an exothermic transition (in the sense that latent heat is released during transition to the high-pressure form) enhances convection, whereas an endothermic phase change impedes it. The transition of various minerals into the perovskite structure is now identified as the cause of the seismic discontinuity at about 660 km depth. Because this transition is endothermic, a two-layered mode of mantle convection might be possible without need for significant compositional differences between the upper and lower mantle. However, based on the early numerical results it was concluded that the properties of the transition to perovskite are probably insufficient to constitute an efficient barrier for mantle convection. During recent years this hypothesis has changed somewhat, because of revised estimates for important thermodynamic parameters, and because less-idealized convection calculations became feasible. In a series of papers beginning with Machetel & Weber (1991) it was suggested that the Earth could be in a hybrid mode, intermediate between the end-member models of layered convection and whole-mantle convection. In the model calculations, periods of layered circulation are interrupted by episodes of extensive mass flux across the phase boundary, termed flushing or avalanche events. Alternatively, the circulation can be semilayered in the sense that it crosses the phase boundary in some regions but not in others.

In parallel to the developments in convection theory, advances in the seismological mapping of deep subducted slabs also suggest a compromise solution to the long-standing question of whether slabs descend into the lower mantle: Some do and others do not (e.g. van der Hilst et al 1991). The evidence for the fate of the subducted lithosphere has recently been reviewed by Lay (1994) and so is not discussed in great detail here. This review concentrates on the principal mechanisms by which phase transitions influence mantle convection and on the various controlling influences from a perspective of geodynamical theory and numerical convection modeling.

THERMODYNAMIC PROPERTIES OF MANTLE PHASE TRANSITIONS

According to the generally accepted petrological model the upper mantle consists of the minerals olivine (63%), orthopyroxene (8%), clinopyroxene (16%), and garnet (13%) (approximate proportions in parentheses). This assemblage undergoes a sequence of structural phase transitions at pressures that correspond to a depth range of roughly 300–800 km (Figure 1a). Few geophysicists disagree that transitions of the olivine component are the cause for the two main seismic discontinuities. The 410-km discontinuity is caused by the transition from the α-(olivine) structure to the β-(Wadsleyite) structure and the 660-km discontinuity by the transition from γ-spinel to an assemblage of perovskite (pv) and magnesiowüstite (mw). These transitions occur for the expected mantle temperatures at the appropriate pressure and are accompanied by increases in
Figure 1  Phase relationships in the mantle. (a) Proportion of various minerals as a function of pressure for the pyrolite model of upper mantle composition. (b) Phase boundaries for the olivine component. The broken line is the 1573 K adiabat. OI = olivine, Wa = Wadsleyite, Sp = spinel, Pv = perovskite, Mw = magnesiowüstite, Opx = orthopyroxene, Cpx = clinopyroxene, Gt = garnet, Maj = majorite. Simplified from Ita & Stixrude (1992); copyright by the American Geophysical Union.
both density and seismic velocity of the right order (Jeanloz & Thompson 1983). However, it is disputed whether a slight change in bulk composition is superimposed, especially at the 660-km discontinuity, where the Mg:Si ratio and/or the Mg:Fe ratio might change (Knittle & Jeanloz 1989, Stixrude et al 1992). For most of the following discussion we assume that this is not the case.

As we shall see in the next section, in a dynamic context the two most important parameters are the difference of density $\Delta \rho$ between the phases and the Clapeyron slope $\gamma = dp/dT$ of the transition. Strictly speaking, the Clapeyron slope is defined only for a univariant transition in a single-component system, where it is linked through the law of Clausius-Clapeyron to the latent heat $Q_L$ or entropy change $\Delta S$:

$$\gamma = \frac{\Delta S}{\Delta V} = \frac{\Delta S \rho^2}{\Delta \rho} = \frac{Q_L \rho^2}{\Delta \rho T},$$

where $T$ is absolute temperature, $\rho$ is the (geometric) mean density of the two phases, and $\Delta V$ is the specific volume difference. For divariant transitions, for which both phases coexist in a finite range of $p$, $T$-space, the Clapeyron slope can be defined in some appropriate way (e.g. as the average of the $p$-$T$ slopes for the points where the transition begins and where it is completed).

The equilibrium phase boundaries for olivine (Figure 1b) have been experimentally determined using the multianvil press and more recently also in the diamond-anvil press. For the $\alpha$-$\beta$ transition, the Clapeyron slope is approximately +3.0 MPa/K (Katsura & Ito 1989). The slope for the $\beta$-$\gamma$ transition could be somewhat higher, on the order of +5 MPa/K (Katsura & Ito 1989). Most important in our context is that the decomposition of the $\gamma$-phase to perovskite, and magnesiowüstite certainly has a negative Clapeyron slope, determined to be $-2.8$ MPa/K by Ito & Takahashi (1989) in the multianvil press. In a different approach, calorimetric data obtained at high temperature and room pressure for the various phases are employed to calculate phase diagrams. The results are generally consistent with the direct determination of phase boundaries, though uncertainties may be somewhat larger. For the $\gamma$ to $\gamma''$ transition, Ito et al (1990) suggested that the Clapeyron slope could be as negative as $-4(\pm 2)$ MPa/K, but based on improved calorimetric results this was later revised to $-3 \pm 1$ MPa/K (Akaogi & Ito 1993). Bina & Helffrich (1994) reassessed the available thermodynamic data and calculated values in the range of $-1.9$ to $-2.7$ MPa/K. Chopelas et al (1994) used thermodynamic properties derived from Raman spectra to arrive at a slope of $-2.5$ MPa/K, in agreement with a limited set of data on the transformation obtained in a diamond-anvil press (Boehler & Chopelas 1992).

The room-temperature densities of most of the high-pressure phases are accurately determined from in-situ measurements at the relevant pressures (see Ita & Stixrude 1992 for a recent review). The density contrast $\Delta \rho/\rho$ is approximately 6%, 2%, and 9% for the $\alpha$-$\beta$, $\beta$-$\gamma$, and $\gamma$-$\gamma''$ transitions, respectively.
We shall see that the influence of a phase transition on convection depends on the product of the density difference and the Clapeyron slope. Because the $\beta$-$\gamma$ transition is accompanied by only a small density increase, it is less important than the other two transitions, even though it may have a larger absolute value of $dp/dT$.

The effect of mantle phase transitions on convection has in most cases been determined on the basis of a pure olivine system. However, the other constituents make up 35–40% of the volume and must not be neglected. The transitions in the pyroxene-garnet components are more gradual and do not lead to sharp discontinuities (Figure 1a). The pyroxenes transform to garnet-majorite in the depth range of 350–450 km. According to Irifune (1987) the increase in zero-pressure density is 5.7% and the Clapeyron slope is $+1.5 \pm 1$ MPa/K. At a pressure slightly higher than in the olivine system garnet-majorite transforms to perovskite, but large uncertainties remain. For example, it is not clear which phase takes up the aluminum in the lower mantle. The properties of the garnet-perovskite transitions are poorly known, but the results by Sawamoto (1987) and Gasparik (1990) suggest a positive Clapeyron slope. Subducted oceanic crust consists of over 80% garnet-majorite at pressures around 20 GPa (550 km). Experiments by Irifune & Ringwood (1993) suggest that for a basaltic composition the transformation of garnet to perovskite would not be completed above a depth of 800 km, which leaves the subducted crust less dense than ambient mantle from 660 km to about 800 km depth or more.

**PRINCIPAL INFLUENCE OF PHASE TRANSITIONS ON CONVECTION**

The two main effects of phase transitions on convection arise from deflections of the phase boundary from its normal depth and from the release of latent heat. The latter modifies the temperature and hence the buoyancy forces that drive the flow. The first effect is more important. When the mantle is in thermodynamic equilibrium, deflections of the phase boundary are controlled by lateral temperature differences, such as those that occur in cold slabs or hot plumes. Figure 2 demonstrates the basic mechanism for a descending slab. In the exothermic case ($dp/dT > 0$) the transition occurs at lower pressure, hence shallower depth, inside a cold slab as compared to the ambient mantle. The elevated region of the denser phase exerts a strong downward pull on the slab; therefore, an exothermic reaction generally helps to drive convection. For an endothermic transition ($dp/dT < 0$) a depression of the low-pressure low-density phase occurs and its buoyancy opposes the sinking of the slab. Similar considerations apply to a hot plume. If the Clapeyron slope is sufficiently negative and the density contrast sufficiently high, convection currents are unable to penetrate the phase boundary.
For a more quantitative consideration, let us assume a Rayleigh-Bénard-type convective system, which contains a single divariant phase change, and let us make some simplifying assumptions about the phase relationship. We employ a phase function $\Gamma(0 \leq \Gamma \leq 1)$ indicating the relative proportion of the dense phase. With linear boundaries in the $p$-$T$ phase diagram and constant width of the phase loop, $\Gamma(p, T)$ is assumed to depend only on the "reduced hydrostatic pressure" given by

$$\pi = p_H - (p_0 + \gamma T),$$

where $p_H$ is the hydrostatic pressure, $p_0$ is the transition pressure at 0 K, and $T$ is the absolute temperature. The transition pressure ($\pi = 0$) is defined such that $\Gamma(0) = 0.5$. Density varies as

$$\rho = \rho_r(1 - \alpha T) + \Delta\rho \Gamma,$$

where $\rho_r$ is a reference density, which may generally depend on depth, but which is assumed to be constant here for simplicity, and $\alpha$ is the thermal expansion coefficient. With these assumptions, the key equations for convection through a divariant phase boundary in nondimensional form are

$$\nabla p' = \frac{\partial \tau_{ij}}{\partial x_j} - Ra \left( 1 + P \frac{d\Gamma}{d\pi} \right) T e_z$$

and

$$\left( 1 + Di \frac{p^2}{R_p} \frac{d\Gamma}{d\pi} \right) \left( \frac{\partial T}{\partial t} + u \cdot \nabla T \right) + \left( 1 + P \frac{d\Gamma}{d\pi} \right) Di u_z T = -\nabla^2 T + \frac{Di}{Ra} \tau_{ij} \frac{\partial u_i}{\partial x_j},$$

where $u$ is the velocity, $\tau$ is the deviatoric stress, $p'$ is the nonhydrostatic part of the pressure, and the Einstein summation convention is used. A derivation of a similar set of equations and the nondimensionalization scheme are found in Christensen & Yuen (1985). The nondimensional parameters in Equations 4 and 5 are the Rayleigh number

$$Ra = \frac{\rho \alpha g \Delta T h^3}{\kappa \eta},$$
the Phase boundary number

\[ P = \frac{\Delta \rho \gamma}{\rho^2 \kappa g h}, \]  
(7)

the Dissipation number

\[ D_i = \alpha gh / c_p, \]  
(8)

and the Density ratio

\[ R_\rho = \frac{\Delta \rho}{\rho \alpha \Delta T}, \]  
(9)

where \( g \) is the gravitational acceleration, \( \Delta T \) is the temperature contrast over the convecting layer, \( h \) is the height of the layer, \( \kappa \) is the thermal diffusivity, \( \eta \) is the viscosity, and \( c_p \) is the heat capacity. If these properties vary with \( p \) and \( T \), some mean reference value is employed in Equations 6–9. With typical mantle values, and for the transition from spinel to perovskite plus magnesiowüstite, approximate values for the various numbers are \( Ra = 10^7 \), \( P = -0.12 \), \( D_i = 0.5 \), and \( R_\rho = 1.5 \), with an uncertainty of at least several tens of percent.

The equations for convection with a single phase are obtained by setting the phase boundary number to zero in Equations 4 and 5. In the equation of motion (4), \( P \) describes the influence of buoyancy arising from deflections of the phase boundary relative to the buoyancy due to thermal expansion. Because the integral of \( dG/d\pi \) over depth equals unity, an absolute value of \( |P| = 1 \) would indicate equal importance of both sources of buoyancy in the system as a whole. Because the buoyancy due to phase-boundary deflection is localized, whereas thermal buoyancy is distributed over the whole depth range, a value for \( |P| \) as small as 0.1 can have a significant effect on convection. In the energy equation (5) the various terms have the following physical meaning: The first term in parentheses describes an enhanced effective heat capacity, which arises because in the two-phase region temperature changes are buffered to some extent by the latent heat of the reaction (compare the classical Stefan problem). Because \( P^2 D_i / R_\rho \ll 1 \), this term is of little importance in our context. The second term on the left-hand side describes the thermal effects of adiabatic compression, and, through the contribution of \( P dG/d\pi \), the release of latent heat. On the right-hand side are the thermal diffusion and frictional heating terms.

In numerical convection calculations the Boussinesq approximation is frequently employed, in which the flow is treated as incompressible \((\nabla \cdot \mathbf{u} = 0)\), and all effects of density changes are ignored except buoyancy effects. This implies setting the dissipation number \( D_i \) to zero. Because the dissipation number controls not only adiabatic and frictional heating but also the latent heat term,
either all three effects must be accounted for, or they must all be neglected to avoid unbalanced sources or sinks of thermal energy. Only slight differences in mass flux across an endothermic phase boundary have been found between the Boussinesq case and either the so-called extended Boussinesq approximation \((Di \neq 0, \nabla \cdot \mathbf{u} = 0)\) or a fully compressible formulation \((Di \neq 0, \nabla \cdot \mathbf{u} \neq 0)\) (Christensen & Yuen 1985, Ita & King 1994). Hence the latent heat effect is of secondary importance compared to that of phase-boundary deflection, at least at the high Rayleigh numbers appropriate for the Earth, but it does have a more profound effect near the critical Rayleigh number (Schubert et al 1975). As long as the vertical mass flux through a phase boundary is not shut off, the effect of latent heat is mainly to introduce a local kink or step into the adiabatic (or isentropic) temperature gradient that characterizes the interior of convection cells (see Figure 1b). Typically, the temperature step \(\delta T = Q_L/\rho c_p\) is of order 100 K or less for mantle phase transitions. As long as latent heat release does not introduce additional lateral temperature differences, i.e. as long as the temperature step is equal inside and outside slabs or plumes, buoyancy is not affected. Because the amount of latent heat depends on absolute temperature, some difference in \(\delta T\) exists, which introduces a minor effect opposing that of phase-boundary deflection.

**STYLE OF CONVECTION THROUGH AN ENDOTHERMIC PHASE BOUNDARY: RESULTS FROM NUMERICAL MODELING**

From the previous discussion it is clear that an exothermic transition enhances convection and that an endothermic transition represents an obstacle. The endothermic case can clearly have a more profound influence on the style of convection and has been studied extensively in recent years. The exact conditions for layered convection cannot be estimated from simple considerations. Olson & Yuen (1982) assumed that a phase boundary number \(P\) of order \(-1\) is required; however, later numerical modeling has demonstrated that two-layer convection is obtained at much lower absolute values of \(P\). The exact threshold value of \(P\) is dependent on a number of parameters, most notably the Rayleigh number (see below). Originally the discussion on the style of mantle convection was in terms of the two end-member models: whole-mantle convection with virtually no impedance of mass flux between upper and lower mantle, or layered convection with very little mass flux, if any. However, the recent numerical results have indicated that there is no abrupt change from one regime to the other, but rather that a hybrid regime of episodic layering and penetration exists at intermediate values of \(P\). The intermittent regime has been found for various geometries: 2-D Cartesian (Weinstein 1993), 3-D Cartesian (Honda et al 1993), spherical axisymmetric (Machetel & Weber 1991, Solheim & Peltier 1994a), and 3-D spherical (Tackley et al 1993). The typical evolution, which is seen
in most numerical models, starts with the accumulation of descending cold material above the endothermic phase boundary, where it piles up and spreads laterally. After some critical mass has been accumulated, the pile overcomes the buoyancy of phase-boundary deflection and penetrates into the lower mantle (Figure 3). (Obviously, to conserve mass, this must be accompanied by a flux of hot material from the lower into the upper mantle, but hot plumes have not been observed to trigger the mass exchange across the phase boundary in the numerical models.) Once the penetration of the lower mantle starts, the large blob of cold material is found to sink rapidly; mass exchange occurs in a short burst, which has been termed a "flush event," "avalanche," or "catastrophic overturn" by various authors. The intermittent regime is observed only at high (>10^6) Rayleigh numbers, such as is realistic for the Earth. At lower Rayleigh numbers, a kind of leaky two-layer convection is found when the phase boundary number \( P \) is below a threshold value (Christensen & Yuen 1985, Ita & King 1994).

**Figure 3** Isotherms for two snapshots from a axisymmetric spherical convection model with two phase boundaries. At 200 Ma convection is predominantly layered; at 300 Ma an avalanche descends into the lower mantle. From Solheim & Peltier (1994a); copyright by the American Geophysical Union.
A quantitative measure of the “degree of layering” is the ratio of the mass flux across the phase boundary to the depth-averaged vertical mass flux (Peltier & Solheim 1992). In most of the published models exhibiting a hybrid regime of convection, the time-averaged flux across the phase boundary is only slightly lower than the flux at other depths (Figure 4), even though the circulation may appear predominantly layered for long intervals of time. Probably the very strong mass flux during the avalanche events makes up for the longer periods of relative quiescence, or, in other words, although cold descending mantle rock is temporarily arrested in the transition zone, most of it sinks eventually into the lower mantle and is not reheated and recycled into the upper mantle. While from the perspective of the time-averaged mass flux between upper and lower mantle the intermittent mode of convection seems closer to whole-mantle convection, profiles of the radially averaged temperature typically indicate a significantly superadiabatic temperature gradient in the depth range of the endothermic phase boundary (Figure 5), with a temperature increase on the order of up to a few hundred degrees. Such a thermal boundary layer is more typical for layered convection and could have implications for the origin of hot plumes in the upper mantle.

For a given value of the phase boundary number the degree of layering of the circulation is found to increase with increasing Rayleigh number (Christensen & Yuen 1985, Zhao et al 1992). Christensen & Yuen (1985) proposed that the critical phase boundary number for the transition between the layered and unlayered convection obeys the relation

$$P_{\text{crit}} = a \, Ra^{-0.2} \quad (10)$$

with a constant $a = -4.4$. In light of the more recent modeling results, this value would slightly underestimate the propensity for layering and may approximately...
indicate the condition for the transition from the intermittent to the completely layered regime. Because modeling assumptions differ strongly in recent publications, and because the parameter space has rarely been investigated in great detail, it is hardly possible to quantify precisely the conditions for the different convective regimes. However, from various model results it can be estimated that the style of convection changes from virtually unlayered to intermittently layered to permanently layered for $P$-values in the range of $-0.1$ to $-0.25$ when the Rayleigh number is of the order $10^6-10^7$. This is compatible with the parameter values estimated for the transition of $\gamma$-spinel to perovskite in the Earth's mantle.

To understand the dependence of the degree of layering on the Rayleigh number we shall elaborate on a simple concept by Christensen & Yuen (1985), which is also useful for understanding the influence of other parameters. For a vertical convective boundary layer (cold slab or hot plume) to penetrate, the obstructive buoyancy of the deflected phase boundary must be overcome by a driving force from thermal expansion. If we consider the cartoon in Figure 1b, the question arises of how much slab length is effectively available to balance the phase-boundary deflection. Boundary layers become thinner and more "fragile" as the Rayleigh number increases, and it can be expected that a thin slab would more easily fold and buckle rather than transmit stress over a long distance to the phase boundary where it is needed to push the slab through. The effective slab length available for balancing phase-boundary deflections should depend on slab thickness. From this reasoning it is expected that at a given value of $P$ the thickness of thermal boundary layers, and their rheological properties, control the degree of convective layering. Although this concept is certainly simplistic and does not account explicitly for the piling of cold material above
the phase boundary, which was found to precede an avalanche event, it provides a basis to understand most of the findings from numerical modeling. Tackley (1994) presents a quantitative elaboration of this concept.

The controlling influence of boundary layer thickness on penetration can provide a simple explanation of why intermittency between layered and unlayered states occurs. In the layered state flow velocities are comparatively low and thick boundary layers can grow. When a critical thickness is reached, an avalanche event occurs. During this event the velocity is much higher than before and the newly formed boundary layers are thinner, which prevents further penetration. The cycle repeats after diffusion has been given sufficient time to grow a thick boundary layer again. A different concept to explain intermittency is based on the spreading of ponded material above the phase boundary (Bercovici et al 1993).

It is commonly assumed that earlier in the Earth's history the mantle was hotter and less viscous, hence the Rayleigh number higher. The sensitivity of relative mass flux through an endothermic phase boundary on the Rayleigh number opens the perspective that mantle convection was predominantly or totally layered in the Archean and became subsequently more and more penetrative, as was demonstrated in convection models that allow for secular cooling (Steinbach et al 1993). For a system evolving in this way, the first overturn would be especially dramatic. It has been speculated that the major resurfacing event on Venus about 500 Ma ago, which was derived from cratering statistics (Bullock et al 1993), might have been caused by such catastrophic overturn of the Venusian mantle (Steinbach & Yuen 1992).

VARIous INFLuENCES ON THE DEGREE OF LAYERING

In addition to the phase boundary number and the Rayleigh number various other parameters influence the degree of layering. Decreasing width of the phase loop for a divariant transition slightly increases the degree of layering (Christensen & Yuen 1985, Peltier & Solheim 1992). In many numerical calculations, the transition interval was larger for technical reasons than the probable width for the \( \alpha-\beta \) and the \( \gamma-(pv+mw) \) transitions, which may be very narrow in the latter case (Ito & Takahashi 1989); therefore, these calculations slightly overestimate the relative mass flux between upper and lower mantle. However, as long as the transition width is much smaller than the thickness of thermal boundary layers, one would not expect that its precise value (or the question of whether the transition is univariant or divariant) has any influence on the dynamics.

Zhao et al (1992) and Solheim & Peltier (1994a) found that internal heating as compared to pure bottom heating slightly increases the propensity for layered convection. In their models, internal heating was added to a specified temperature contrast across the convecting region. This increases the effective Rayleigh number of the system, which may provide a simple explanation for their results.
While so far most of the numerical modeling has been done in two-dimensional geometry, recently Honda et al (1993) and Tackley et al (1993, 1994) presented models in three-dimensional Cartesian and spherical geometry, respectively. Basically, the dynamics is similar to that in two-dimensional simulations. The geometry of downwelling currents is that of interconnected cold sheets, and it is found that the piling of material in the transition zone occurs predominantly at the junctions of such sheets. The accumulated material is then suddenly released into the lower mantle in a large circular downwelling. Whether the geometrical details observed in these simulations are relevant for the Earth remains doubtful, because the strong concentration of descending flow at junctions of descending sheets is due to the assumption of constant viscosity in the model and is hardly expected for subducting plates. Another observation by Tackley et al (1993, 1994) is that the global mass flux across the endothermic phase boundary, although reduced compared to the radial mass flux at other depth levels, does not show the strong variations with time that were observed in most of the two-dimensional simulations. They find that although on a regional scale the flow between upper and lower mantle is episodic, the avalanches do not occur in a globally synchronous manner. Their model system is large enough for several avalanches to be active at a given time with some phase shift, so that global averaging gives a mass flux that is fairly constant in time.

In many model calculations only the endothermic transition at 660 km has been taken into account. Solheim & Peltier (1994b) and Tackley et al (1994) found a slight enhancement of the mass flux between upper and lower mantle when the exothermic olivine-spinel transition is also accounted for, in accord with expectation. However, Zhao et al (1992) and Steinbach & Yuen (1992) reported the opposite finding. The reasons for this discrepancy are not resolved.

In most models that included phase boundaries, constant viscosity or a smooth variation of viscosity with depth was assumed. However, the differences in mechanical properties of slabs and plumes, and the plate-like character of motion at the surface can have a significant influence on the mass flux through an endothermic phase boundary. Christensen & Yuen (1984) employed a temperature- and stress-dependent rheology to simulate oceanic lithosphere in a convecting system, but made no detailed comparison with the isoviscous case. Zhong & Gurnis (1994), in a more detailed study, use an ad-hoc distribution of viscosity in the surface layer of their model to obtain plate-like motion. One consequence of the existence of plates is that much larger width-to-height ratios for the convection cells are possible. Zhong & Gurnis find that the relative mass flux between upper and lower mantle increases very strongly with the ratio of plate length to the height of the model box. In their calculations, convection can be almost perfectly layered for aspect ratio one and not layered at all for aspect ratio five, with all other parameters equal. These results are easily understood
on the basis of our assumption that boundary layer thickness is a major controlling parameter for the penetration of the endothermic boundary, because longer plates imply greater thicknesses from conductive cooling. In a further set of calculations, Zhong & Gurnis (1994) also include temperature-dependent viscosity to account for the higher mechanical strength of the subducted part of the plate. Compared to an isoviscous descending flow, penetration of slabs is favored by the high viscosity and proceeds in a more steady fashion not exhibiting the strong intermittency usually observed in isoviscous models. Apparently, the increased viscosity of the slab makes it a more efficient stress guide and allows the driving (negative) buoyancy by a longer part of the slab to be balanced against the impeding buoyancy of the deflected phase boundary. However, Zhong & Gurnis also find that the overall mass flux between upper and lower mantle is not enhanced by the temperature dependence of viscosity; outside the region of slab penetration the circulation can be layered. These calculations help indicate how the existence of mechanical plates affects the picture derived from essentially isoviscous models. However, one of the drawbacks of the models by Zhong & Gurnis (1994) and Christensen & Yuen (1984) is that slabs descend vertically at a sidewall of the model system where mirror symmetry is assumed. A shallower dip angle of slabs would probably decrease the propensity for slab penetration, as was demonstrated in laboratory experiments of slab subduction into a chemically layered “mantle” (Kincaid & Olson 1987).

Phase transformations cause discontinuous changes in material properties, which could influence the convective flow. For example, Rubie (1984) argued that diffusion-controlled superplasticity in the fine-grained reaction product of a transformation would weaken the material. The effect was modeled by Christensen & Yuen (1985) in a crude way by introducing a low-viscosity zone in the vicinity of an endothermic phase boundary, which slightly favors layered convection.

REACTION KINETICS

So far it has been assumed that mantle phase changes occur at or very close to thermodynamic equilibrium. However, in the coldest parts of subducting slabs the reaction kinetics might be so slow that the low-pressure phase is metastably preserved to great depth (Sung & Burns 1976, Daessler & Yuen 1993). Based on kinetic data for the olivine-to-spinel transition mainly in analogue materials, Rubie & Ross (1994) concluded that at temperatures around 550°C, which are expected in the core of rapidly subducting slabs, olivine would be preserved down to a depth of about 550 km, whereas for temperatures of
700°C or higher the transformation would occur close to equilibrium. Estimates for the kinetics of the perovskite-forming transformations are not available. If slow reaction kinetics plays an important role, the 410-km discontinuity should be depressed rather than elevated in the coldest parts of subducting slabs. The seismic evidence concerning the deflection of the 410-km discontinuity in slabs is ambiguous: Both elevation (Vidale & Benz 1992) and depression (Iidaka & Suetsugu 1992) have been reported. For the 660-km discontinuity, Richards & Wicks (1990) and Vidale & Benz (1992) reported a depression of the order of 20–30 km in the Tonga slab, which according to regional tomographic studies (Zhou 1990) probably penetrates into the lower mantle. This amount of depression is expected for an equilibrium transition and does not hint at an important delay of the spinel→(pv + mw) transition. If slow reaction kinetics is important for the 660-km boundary, it may provide another mechanism to introduce episodicity of penetration into the lower mantle. The sinking of slabs into the lower mantle would be controlled by the thermal diffusion time needed to heat up the coldest parts of accumulated slab material sufficiently for the phase transformation to occur.

ADDITIONAL CHEMICAL DIFFERENCES

It is unresolved whether the seismic properties of the lower mantle are consistent with a major element composition that is the same as that in the upper mantle (Chopelas & Boehler 1992, Wang et al 1994) or whether a slight chemical difference is required (Knittle & Jeanloz 1989, Stixrude et al 1992). Christensen & Yuen (1984) and Kincaid & Olson (1987) found that without phase-boundary effects a chemical density difference on the order of 5% is needed to prevent slabs from dipping deeply into the lower mantle. Because the more recent petrological models of the lower mantle limit the possible intrinsic density contrast to approximately 2% or less, the compositional stratification alone would not prevent overturn, and the effects of the endothermic phase transition would remain essential. However, if a difference in the bulk composition of the upper and lower mantle could be proven, this would put a severe upper limit on the mass flux across the 660-km discontinuity during the Earth’s history. Silver et al (1988) speculated that slabs may penetrate temporarily into a lower mantle with a slightly higher intrinsic density and after some time pop up again, but whether such a scenario would be dynamically plausible has not been fully explored.

The cause for a possible chemical difference between upper and lower mantle is also debated. The hypothesis that it results from the crystallization of an early magma ocean (Ohtani 1985) was challenged by Kato et al (1988), who find that this contradicts the observed chondritic relative abundances of some trace elements in the upper mantle. These elements should be strongly fractionated from one another during solidification of the magma ocean. Recently, processes
for the chemical differentiation of the mantle have been proposed that invoke the solid-solid phase boundaries directly. Bina & Kumazawa (1993) point out that chemical equilibrium across phase transitions with a superimposed temperature gradient implies a difference in composition. However, the chemical stratification produced by a process based on this effect would require drastically enhanced mass diffusion. To test a different hypothesis, Weinstein (1992) presented a numerical convection model containing distributed chemical heterogeneity and a phase boundary. In the imperfectly layered convection mode, the phase boundary acts as a filter: Chemically dense rock in the upper mantle will preferentially pass the boundary and sink in a trap-door fashion into the lower mantle, whereas chemically buoyant rock is retained in the upper mantle. Gradually a compositional stratification builds up, which coincides with the phase boundary. However, in these calculations the initial distribution of chemical heterogeneity was rather arbitrary; it was not attempted to model the compositional stratification of subducting lithosphere, which is the main known source of heterogeneity.

The dynamical effects of phase transitions on sinking slabs, which are thought of as a sandwich structure of former oceanic crust, a depleted harzburgite layer, and fertile mantle, are complicated even more because the transition pressure to the perovskite structures depends on chemical composition. Irifune & Ringwood (1993) find that the transformation from the garnet structure to perovskite in the former crustal layer is not completed above a depth of about 800 km. Accordingly, it would be less dense than the surrounding lower mantle and add to the impeding buoyancy of the downward deflection of the phase boundary in the slab due to low temperatures. The importance of this additional source of buoyancy, which is restricted to a rather thin layer, has been questioned in model studies of Richards & Davies (1989) and Gaherty & Hager (1994). However, these model calculations did not consider explicitly the combined action of phase transitions and compositional heterogeneity of the slab. When the Clapeyron slope is sufficiently negative to drive the system close to a layered state, the additional buoyancy of the crustal layer in the depth range from 660 to 800 km might be decisive to arrest the slab completely.

APPLICATION TO THE EARTH AND COMPARISON WITH OBSERVATION

The concept of a hybrid mode of mantle convection, intermediate between whole-mantle convection and layered convection, is attractive, as it might reconcile various observations that appear otherwise conflicting. Fluid dynamical modeling has demonstrated the possibility of such a mode. In most of these models the phase boundary number $P$ was somewhat more negative than what our best estimates for the relevant thermodynamic properties in the mantle
would indicate, especially when one considers that the Clapeyron slope for the transition of garnet-majorite to perovskite may be positive. However, other effects such as kinetic hindrance of the phase transition or the stability of garnet in the former crustal layer of the slab to depths much greater than 660 km could help to drive the system to a more layered state.

The best evidence indicating regional differences in the material flow across the 660-km discontinuity comes from regional seismic tomographic studies of the subduction zones in the western Pacific (van der Hilst et al 1991, 1993; Fukao et al 1992). Elongated subhorizontal high-velocity structures above the 660-km discontinuity, extending from deep Benioff zones into the direction of subduction, are found for the Southern Kurile, the Japan, and the Izu-Bonin subduction zones. For the Kamtchatka-Northern-Kurile and the Tonga slab, high velocities are found to extend from the deepest seismicity well into the lower mantle. The straightforward interpretation is that in the former cases the slabs fail to penetrate and flatten in the upper mantle, whereas in the latter cases they steepen and descend into the lower mantle. It has been suggested that high-velocity anomalies in the lower mantle below deep subduction zones may represent cold sinking convection currents that originate below the 660-km discontinuity in the so-called thermally coupled mode of layered convection. However, a slab would not dissolve without trace, and the lack of evidence for flat aseismic extensions of the slabs above the discontinuity in these cases makes this interpretation unlikely.

Although the tomography results indicate that different slabs may behave quite differently when they encounter the 660-km discontinuity, there is no clear evidence that slab material piles up above the discontinuity. Rather, slabs that fail to penetrate into the lower mantle seem to move horizontally, or, they may just “lie down” on the discontinuity when fast trench rollback occurs. van der Hilst & Seno (1993) argue that rapid trench rollback and the subduction of relatively young lithosphere explains the flat-lying extension of the Izu-Bonin slab above the discontinuity, whereas for the Mariana trench rollback was slower and the lithosphere older, leading to the present penetration of cold flow into the lower mantle. At least partial layering of mantle convection is also suggested by large-scale topography of the 660-km discontinuity derived from travel times of S-waves reflected from it (Shearer & Masters 1992, Shearer 1993); the discontinuity is some 15 km below its average depth in the region of western Pacific subduction. Although the actual deflection inside a slab penetrating the discontinuity would be larger, this would hardly be visible on large scales because of the small spatial extent of the slabs. Low temperatures on a regional scale are probably required to explain the large-scale depression, in accord with the presence of flattened slabs in the transition zone. Solheim & Peltier (1994b) compared phase-boundary deflection in a convection model with partial layering and obtained similar values of the order of 10–20 km for
the average deflection of the 660-km discontinuity. On the other hand, Phipps Morgan & Shearer (1993) found that the buoyancy connected with the deflection of the 660-km discontinuity (as determined by seismology) is insufficient to suppress mass flux across the boundary, which in their model is driven by density anomalies in the mantle inferred from global tomography. However, aside from the uncertainties of the tomographic models, the result is sensitive to the conversion factor from velocity perturbation to density perturbation. A detailed study of the topography of the 660-km discontinuity beneath the Izu-Bonin trench (Wicks & Richards 1993) shows a 40–60 km depression, which is probably 200–300 km wide, but its extent towards the Phillipine sea, i.e. the region where tomography suggests a flat-lying slab, is not well constrained. The data would be compatible either with flattening of the slab in the upper mantle or with advective thickening of the slab when it enters the lower mantle.

Although regional tomography results support a partially layered mode of convection, several analyses of global tomographic inversions suggest that mantle convection is largely unlayered at present. Numerical models in which convection is partially or fully layered exhibit a distinct peak in the distribution of power of long-wavelength thermal heterogeneity versus depth around the endothermic phase boundary (Peltier & Solheim 1992, Tackley et al 1994), which is typically not found in global tomography models of the mantle (Woodward et al 1994). However, it is possible that the effects of stagnant slabs in the transition zone on seismic velocities, which would cause an enhanced lateral heterogeneity, are partially compensated by the opposite effect arising from the induced phase boundary topography. A joint inversion for both sources of seismic heterogeneity has not yet been done—they may cancel each other to some extent (Wahr et al 1992) and reduce the apparent power of heterogeneity around the 660-km discontinuity. Jordan et al (1993) find that in a convection model with partial layering the correlation of thermal anomalies across the phase boundary is poor; the radial correlation length is much smaller near the boundary than in other depth ranges. Such a minimum in correlation length is not found in tomographic models of the whole mantle, and the authors conclude that the large-scale circulation in the mantle is not seriously inhibited at the 660-km discontinuity. However, the overall radial correlation length in mantle tomography is clearly less than it is in their whole-mantle convection model, which makes the argument less convincing.

While there is at least some evidence from seismology that spatial differences in the mass flux between upper and lower mantle exist, it is more difficult to test if there are strong changes with time, and if so, whether they occur on a global scale. Cadek et al (1994) compared tomographic models of the lower mantle with the sites of Cenozoic and Mesozoic subduction (Richards & Engebretson 1992) and found significant correlations between seismic anomalies in the upper
portions of the lower mantle with Cenozoic subduction and between anomalies in the deepest mantle with Mesozoic subduction. They interpret the minimum in correlation in the depth range 2000–2500 km as evidence for a hiatus in the downward mass flux that followed an avalanche event in the Cretaceous. There have also been speculations on a link between the increased plume activity during the Cretaceous documented by the formation of oceanic plateaus (Larson 1991) and a possible avalanche event through the 660-km boundary (Weinstein 1993); however, the three-dimensional model results of Tackley et al. (1994) warn that the episodes of strong mass flux across the endothermic phase boundary may be of regional scale and not as global as commonly observed in two-dimensional models.

Seismology can now provide fairly detailed images of slabs, but the deep structure of hot plumes remains more elusive (but see Nataf & VanDecar 1993). Often it is assumed that plumes rise from the core-mantle boundary (Davies & Richards 1992). However, rheologically stiff slabs do not easily penetrate the 660-km discontinuity, so penetration would be even more difficult for a low-viscosity plume. On the other hand, additional obstacles such as slow reaction kinetics are not relevant for plumes. In some 2-D numerical models (Steinbach & Yuen 1994; U Christensen, unpublished results) it was observed that hot plumes, which rise in the upper mantle in response to sinking avalanches, draw their material from the upper parts of the lower mantle without direct connection to the lower thermal boundary layer. However, if this were the origin of plumes in the mantle, their relative fixity and persistence would be hard to understand. More work on the behavior of low-viscosity plumes in a semilayered convection regime is required.

Geochemistry can provide some contraints on the vertical mass exchange in the mantle. If continental crust is considered to be the complement to the so-called MORB-reservoir (the part of the mantle from which mid-ocean ridge basalts are derived and which is depleted in incompatible trace elements), then mass balance calculations for various trace elements suggest that the mass of the depleted mantle is approximately 50 ± 20% of the total mass of the mantle (Hofmann 1989). Rare gas abundances and isotope ratios in the atmosphere and in oceanic basalts (Allègre et al. 1987) suggest that the strongly degassed part of the mantle forms approximately one half of the total mantle. The uncertainties are rather large, but taken at face value these estimates would exclude both perfectly layered convection or simple whole-mantle convection, and would agree with episodic or imperfect layering. However, as discussed above, in most numerical models with episodic layering, the time-averaged mass flux between upper and lower mantle is not strongly reduced compared to unimpeded whole-mantle convection; for example, Tackley et al. (1994) estimate from their models time scales of 1.1–2.5 Ga for exchanging the entire mass of the mantle. If the depleted mantle forms only a fraction of the whole mantle, convection must
have been quite close to the layered model on average over the Earth's history. One way to reconcile this with the evidence of significant recent mass flux through the 660-km discontinuity would be that in the past a higher Rayleigh number made the convection more layered.

SUMMARY

Results of numerical modeling suggest that some form of compromise between the long-standing end-member models of whole-mantle convection and layered convection is dynamically plausible. The influences of various parameters on the style of convection can be understood from a balance of buoyancy forces arising from 1. phase boundary deflection, which in the case of an endothermic transition impedes convection, and 2. thermal buoyancy, which drives convection. Given current values for the Clapeyron slope of the transition to perovskite and for other thermodynamic parameters, the conditions for an episodic regime are marginally satisfied. However, sluggish reaction kinetics in slabs and the stability of garnet in the subducted oceanic crust below 660-km depth could enhance the forces resisting slab penetration. Seismic tomography of subduction zones suggests that some slabs are deflected and remain at least temporarily in the transition zone, whereas others descend straight into the lower mantle. The distribution of power with depth and the radial correlation in global models of lateral seismic heterogeneity suggest that presently convection may not be strongly layered, whereas geochemical arguments require some but not too much mass flux between upper and lower mantle over the age of the Earth.

The question to be resolved in the future may no longer be whether the model of layered convection or that of whole-mantle convection is correct, but rather what the relative mass flux between upper and lower mantle is and how it changes on short and long time scales. Seismic imaging of the mantle, mineral physics, geochemistry, and convection modeling will combine to resolve these issues. In order to be able to compare the results of convection modeling more directly with observations, the most important step is to model slabs in a more realistic way. Their rheology, petrological structure, and geometry of descent play an important and perhaps decisive role in what happens when slabs meet the 660-km discontinuity.

ACKNOWLEDGMENTS

Comments by Carolina Lithgow, Neil Ribe, and Lars Stixrude helped to improve the manuscript.
PHASE TRANSITIONS AND CONVECTION 85

Literature Cited


Christensen UR, Yuen DA. 1984. The interaction of a subducting lithospheric slab with a chemical or phase boundary. J. Geophys. Res. 89:4389-402


Gasparik T. 1990. Phase relations in the transition zone. J. Geophys. Res. 95:15,751-69


Ito E, Takahashi, E. 1989. Post-spinel transformations in the system Mg$_2$SiO$_4$-Fe$_2$SiO$_4$ and some geophysical implications. J. Geophys. Res. 94:10,637-46

Ito E, Akaogi M, Topor L, Navrotsky A. 1990. Negative pressure-temperature slopes for reactions forming Mg$_2$SiO$_4$ perovskite from calorimetry. Science 249:1275-78


Katsura T, Ito E. 1989. The system Mg$_2$SiO$_4$-Fe$_2$SiO$_4$ at high pressures and temperatures: precise determination of stabilities of olivine, modified spinel, and spinel. J. Geophys. Res. 94:15,663-70

Kincaid C, Olson P. 1987. An experimental


van der Hilst R, Engdahl ER, Spakman W. 1993. Tomographic inversion of P and pP data for aspherical mantle structure below the north-


