MANTLE CONVECTION AND THE STATE OF THE EARTH'S INTERIOR

Bradford H. Hager
Michael Gurnis
Seismological Laboratory
California Institute of Technology
Pasadena, California 91125
Introduction

During 1983-1986, the four year period covered by this review, emphasis in the study of mantle convection shifted away from fluid mechanical analysis of simple systems with uniform material properties and simple geometries, toward analyzing the effects of more complicated, presumably more realistic models. This shift received impetus from developments in other branches of earth sciences, particularly seismology, geodesy, geochemistry, and mineral physics, which began to provide more powerful observational constraints on mantle convection. The ability to address previously intractable problems like the origin of long-wavelength geoid anomalies attests to the power of this multidisciplinary approach. The stage is now set for rapid acceleration in our understanding of mantle dynamics through comparison of improved observations and more sophisticated models.

This paper is organized primarily around what we consider to be important processes related to mantle convection and secondarily around important technical developments. Due to space limitations, it is impossible to do justice to all the important developments. More comprehensive recent reviews of mantle convection and the thermal state of the earth include those by Loper (1984a, 1985) and Jeanloz and Morris (1986). Stevenson (1983; Stevenson et al, 1983) and Loper and Roberts (1983) have reviewed core convection and dynamics. Anderson (1984) has provided a provocative discussion of outstanding problems and paradoxes.

Developments in Seismology

In the past four years, determination of earth structure using seismology has become more relevant to mantle dynamics, with emphasis shifting from determination of spherically symmetric models to the resolution of lateral heterogeneities. The lower mantle has been probed using body wave tomography (Dziewonski, 1984; Clayton and Comer, 1984); the upper mantle has been mapped globally using surface wave tomography (e.g. Woodhouse and Dziewonski, 1984; Tanimoto and Anderson, 1984; Tanimoto, 1986). Variations in anisotropy, presumably due to alignment of crystals by mantle flow (e.g. Christensen, 1984), have also been reported (e.g. Anderson and REgan, 1983; Nataf et al, 1984; 1986; Nakanishi and Anderson, 1984; Tanimoto and Anderson, 1985). Even the inner core has been probed, using both body wave phases (Morelli et al, 1986) and splitting of the Earth's free oscillations (Woodhouse et al, 1986). All global tomographic studies to date have been hampered by less than optimal data coverage. As a result, the velocity variation at a given point in an Earth model is often a strong function of the parameterization used (e.g. discrete blocks vs low order polynomials), particularly near the 670 km discontinuity and the core-mantle boundary, making detailed dynamic interpretation of these models hazardous. Still, much progress has been made, regions where sampling is good are increasingly well resolved, and more reliable determinations of structures are to be expected. Variations in seismic velocity have also been determined on regional scales (e.g. Walck, 1984, 1985; Grand and Helmberger, 1984ab), with contrasts between tectonically active and stable regions being quite large above 200 km and extending to 400 km depth. The large amplitudes of these contrasts may be indicative of fairly deep seated thermo-chemical heterogeneity. Grand (1986)
presented a tomographic analysis of shear wave structure beneath the North American plate. He traced a high velocity anomaly which may be the remnant of the old subducted Farallon plate extending to over 1500 km depth beneath the east coast of North America.

The Geoid and Mantle Viscosity

The long-wavelength geoid has been interpreted in terms of mantle convection, both empirically and dynamically. Eckhardt (1984) presented a tutorial about correlations on a sphere. Chase and Sprowl's (1983) study pointed out that major present-day geoid lows correspond to regions where subduction was occurring ~100 Myr ago, regions which are now also mostly devoid of hotspot activity. This suggests a rather sluggish response of the lower mantle to subduction.

From a dynamic standpoint, mantle convection results in significant dynamic topography at the surface, at the core-mantle boundary (CMB), and at any other boundaries in chemical composition that might exist, leading to a sort of "dynamic isostasy." This dynamic topography has an effect on the geoid opposite in sign and comparable in magnitude to that of the interior density contrasts driving flow.

The net geoid anomaly is a small difference of large numbers. Parsons and Daly (1983) presented results for two-dimensional dynamic geoid models; Richards and Hager (1984) and Ricard et al (1984) presented results valid for spherical geometries. These studies show that due to dynamic compensation the geoid is insensitive to interior density contrasts near the boundaries of the flow system, but is sensitive to density contrasts in the interior. The predicted geoid anomaly for a given density contrast (e.g., one inferred from seismic tomography) is a strong function of mantle viscosity structure and the presence or absence of chemical stratification.

Hager et al (1985) showed, using simple models with two viscosity layers, that density contrasts inferred from tomographic models of lower mantle structure can account for most of the long wavelength geoid, with subducted slabs also making an important contribution (Hager, 1984). Incorporating results from upper mantle surface wave tomography, in a more detailed, four-layer viscosity model, allows 90% of the variance in the geoid at wavelengths longer than 4,000 km to be explained (Hager and Richards, 1985; Hager and Clayton, 1986).

Dynamic geoid models can resolve relative variations, but not absolute values of mantle viscosity. A lower bound can be inferred by requiring that the advected heat flux carried by flow driven by the density contrasts causing the geoid anomalies be less than the observed surface heat flux (O'Connell and Hager, 1984). This approach led Hager and Clayton (1986) to infer an average asthenospheric viscosity of $10^{20}$ Pa s, with viscosity increasing through the transition zone to $3 \times 10^{22}$ Pa s, remaining essentially constant below that depth for their preferred viscosity model. (There is, however, a strong tradeoff in models between the thickness and relative viscosity of a given layer.) On the basis of solid state theory, Poirier and Liebermann (1984) show that if the geotherm is adiabatic, the viscosity should increase by a factor of 100 near the top of the lower mantle and then become essentially isoviscous due to the decrease in activation volume with depth. Much of the effort to infer the viscosity
structure of the mantle from its response to deglaciation was focused on its rotational response (e.g. Yuen and Sabadini, 1984; Wu and Peltier, 1984). Satellite geodesy made possible detection of changes in the observed Earth ellipticity in the eighth decimal place of precision (Yoder et al., 1983; Rubincam, 1984). The observed value, if attributed to postglacial rebound alone, suggests a lower mantle viscosity of either $\sim 3 \times 10^{21}$ Pa s or $\sim 10^{24}$ Pa s, depending upon which root of an equation with two solutions is chosen (e.g. Peltier, 1983; Yuen and Sabadini, 1984; 1985). However, as Yoder and Ivins (1985) point out, present day melting of small glaciers (Meier, 1984) must also be accounted for, leading to a preferred lower mantle viscosity of $\sim 2 \times 10^{22}$ Pa s. This is in close agreement to that inferred from heat flux arguments.

To date most models of viscosity inferred from postglacial rebound have been quite simple, employing only two viscous layers, as is appropriate for the resolution of that data set. It appears likely, although yet to be demonstrated quantitatively, that the more detailed mantle viscosity structure inferred by matching the geoid can satisfactorily explain the postglacial rebound observations. If not, the discrepancy may indicate that postglacial rebound is sampling a transient response (Peltier, 1985a; Sabadini et al., 1985).

Another topic is the detection of effective chemical stratification (nonadiabatic density contrasts). Fjeldskaar and Cathles (1984) pointed out that the common use of the correspondence principle to solve the viscoelastic equations (e.g. Peltier, 1985b) results in density increases with depth being treated implicitly as nonadiabatic. It is not yet resolved whether nonadiabatic density contrasts are required, and if required, whether they are the result of true chemical layering, or to delayed phase transitions.

Is the Mantle Chemically Stratified?

This question continues to be hotly debated, although to some extent it is a semantic one. One reason for invoking a stratified mantle is to keep geochemical reservoirs separated for eons, yet the fact that materials from these "distinct" reservoirs are seen at the surface indicates some connection with them. The meaningful questions appear to be not whether circulation penetrates the 670 km discontinuity, but to what extent, and what are the details of the plumbing. Does the lower mantle have a chemically distinct composition from the upper mantle, with interaction due to dynamic entrainment (e.g. Olson, 1984), or does a substantial viscosity increase with depth slow the interchange of material between upper and lower mantle? Are distinct reservoirs maintained by differences in intrinsic density, or by differences in mechanical properties?

One approach to resolving this problem has been through comparison of models of seismic velocity variations with depth with predictions of models from mineral physics (e.g. Jeanloz and Thompson, 1983). Although some crucial material properties are as yet undetermined, recent results have been provocative. Anderson and Bass (1984; 1985; Bass and Anderson, 1984) showed that an olivine-eclogite composition dubbed "piclogite" may fit the seismological models. They cast doubts on the traditional pyrolite model, but Weidner (1985), making different assumptions about as yet unmeasured material properties, leapt to its defense. More measurements will
help to resolve these controversies. However, given the large regional variations observed in seismic velocities, it appears that future comparisons will have to address a range of possible temperatures and compositions.

Bock and Ha (1984) attributed S-P seismic phase conversions to a sharp interface at ~ 700 km depth near the Tonga slab, while Lees et al (1983) interpreted P'P' precursors occasionally observed as due to reflections from a very sharp (< 2 km thick) 670 km seismic discontinuity (an interpretation challenged by Muirhead (1985)). Such a sharp interface is not expected due to a phase change alone given available (extrapolated) phase diagrams. If the phase diagrams hold up, a chemical boundary seems required, which would then lead to the puzzle of why these phases are not observed more commonly.

If the 670 km seismic discontinuity is a chemical discontinuity, there should be substantial dynamic topography on it, reaching an amplitude of hundreds of km beneath subducted slabs (Christensen and Yuen, 1984; Hager, 1984). Such topography might explain the general absence of P'P' precursor reflections. Mapping the topography of the 670 would be the most direct way to determine whether chemical stratification might exist. In the absence of detailed mapping, subducted slabs may provide the best probe. The deepest earthquakes are all shallower than 680 km (Stark and Frohlich, 1985), not significantly deeper than 670 km. This is consistent with a phase change at that depth shutting off seismicity and is not suggestive of the hundreds of km of dynamic deformation expected for a chemical discontinuity.

Creager and Jordan (1984; 1986a) analyzed long-wavelength variations in projections of travel time residuals of P waves onto the focal spheres of deep earthquakes in subducted slabs. They proposed that slabs penetrate essentially unaltered in form, but with varied dip, to ~ 1400 km depth. Silver and Chan (1986) explained variations in S-wave wave forms as due to multipathing caused by a similar structure. These observations suggest that convection is mantle-wide, although triggering of cold downwellings in the lower mantle by impinging of subducted slabs on the 670 might also be a possibility.

Giardini and Woodhouse (1984) showed that the deep Tonga slab is telescoping under compression as it approaches 670 km, indicating substantial shortening due to resistance to penetration. Vassiliou et al (1984) showed that the resistance to penetration due to a viscosity jump of a factor of 10 or more leads to down-dip compression in deep slabs not resolvable different from that due to a chemical barrier. The resulting flow patterns are quite different, however, with slabs tending to bend over and lie flat along the boundary if the mantle is stratified (Hager et al, 1983; Garfunkel et al, 1984). For whole-mantle flow with a viscosity increase, the slab thickens by pure shear while penetrating the 670, in accord with the seismic observations. Comparing the degree of shortening determined from deep seismicity or focal sphere analysis to that predicted for convection models should place useful constraints on slab and mantle rheology.

D" and the Core-Mantle Boundary

The structure of D", the region of anomalous seismic velocity gradient just above the core-mantle boundary (CMB), is of obvious
importance to convection, since heat must pass through it from the core; it has traditionally been interpreted in terms of a simple thermal boundary layer. Loper (1984b; Loper and Stacey, 1983; Loper and Eltayeb, 1986) has interpreted the region of anomalous velocity gradient in a global seismic reference model in terms of a thermal boundary layer and the source of mantle plumes. The dynamic stability and behavior of such a thermal boundary layer has been explored numerically by Christensen (1984a) and analytically by Marshall and Morris (1986), who show that initial small-scale convection within the boundary layer will be enhanced in a fluid with temperature-dependent viscosity, as previously thought, but that the time scale needed for hot-fluid to “break out” and form a plume will be controlled by the ambient viscosity. The small-scale convection would give rise to CMB topographic variations with a wavelength on the order of the layer thickness, \( \sim 100 \) km (Olson et al, 1986); such high frequency undulations could be the cause of seismic velocity variations inferred to exist over scales as small as 100 km at the CMB (Lavely et al, 1986).

More refined seismic models (e.g. Lay and Helmberger, 1983ab) have D'' structures that include shear velocity increases with depth, incompatible with a simple boundary layer interpretation. It seems likely instead that D'' is a chemically distinct layer. The thickness of this layer appears to vary substantially (Lay, 1986a); indeed it may not even be a globally continuous feature. Craeger and Jordan (1986b) speculate on the basis of analysis of PKP travel times that features analogous to continents may exist at the CMB. Davies and Gurnis (1986) discuss simple models of the fluid dynamics of these inferred “mantle dregs” and show that they would be expected to have substantial lateral variability in thickness. If plumes do indeed originate from this complex thermal-chemical layer, then the rate at which hot-fluid “breaks out” of the layer and forms plumes will be reduced by the chemical density contrast (Christensen, 1984a).

Dynamical models of the geoid have not yet addressed the complexities of D'' existing either as a low viscosity thermal boundary layer or as a chemically distinct layer. If D'' is chemically distinct, the mass anomalies attributed to dynamic topography at the CMB in geoid models would instead be the result of much larger dynamic topography at the top of D'', since the amplitude of dynamically maintained topography is inversely proportional to the density contrast across that boundary. One way of addressing the structure of this region is through the mechanical interaction of flow in the core with bumps on the core-mantle boundary. Gwinn et al (1986) have interpreted an observed shift in the resonance frequency of the free-core nutation as being the result of coupling of core motion to the mantle through excess ellipticity of the CMB. The required topography is a factor of 5 smaller than that predicted by Hager et al (1985) on the basis of geoid modelling. Hide (1986) suggested that the observed decade length fluctuations in length of day might be due to interactions of flow in the mantle with bumps at the CMB. Speith et al (1986), using the geostrophic core flow models of Le Mouel et al (1985) showed that changes predicted using available core topography models are too large. These geophysical data sets may require a chemically distinct D'' to explain them. Brown (1986) argues on
the basis of expected thermal conductivity and core heat flux that D" is more likely to be compositional than thermal in origin. Estimates of temperature increase from lower mantle adiabat to core are 300° - 800°C (e.g. Shankland and Brown, 1985). The dynamics of a chemically distinct, possibly low viscosity D" layer are not yet sufficiently well understood to make this temperature jump a useful constraint on the convective state of D".

A chemically distinct D" would require revision of estimates of mantle composition based on mass balance calculations. D" presents a possible reservoir comparable in mass to continental lithosphere, with little constraint upon its composition.

Fluid Dynamics of Convection

Systematic fluid dynamical studies, both numerical and laboratory, elucidate the basic physics of convection. They also establish the physical plausibility of "mantle models" and suggest new types of observations. Over the last few years much work concentrated on the effects which "realistic" rheologies have on thermal convection and on problems related to subduction, hot-spots and plume-style convection.

Effects of "Realistic" Rheologies: The temperature-dependence of silicate rheology has been thought responsible for the "plate-like" characteristic of mantle convection. However, in fluid dynamical models attempting to simulate this process the cold top thermal boundary becomes stiff and globally immobile, unlike the oceanic lithosphere which slides past the overriding plate. Christensen (1984b) showed that at a moderately high Rayleigh number, if the viscosity is stress dependent as well as temperature dependent, the top surface could "subduct," allowing "plate motion." At the Rayleigh numbers appropriate for whole mantle convection, however, the top surface remained immobile even for a strongly nonlinear viscosity (Christensen, 1985a). More work along these lines is needed. Christensen (1983) showed that if the rheology is stress dependent, the strong coupling between flow and viscosity tends to homogenize the viscosity, with variations generated by temperature and pressure dependencies becoming smeared out. Lin and Parmenter (1985) showed that even with temperature dependent viscosity, topography and gravity over upwellings is positive.

Hot-Spots and Plume-Style Convection: Hot-spots are generally attributed to some sort of axial convection extending deep into the mantle. Studies related to the fluid dynamics of hot-spots flourished during the last several years, particularly with regard to the initiation of plumes on thermal boundary layers, as discussed above, the ascent of plumes through the mantle, and the interaction of plumes with the lithosphere.

Important fluid dynamical research was presented on the behavior of plumes or diapirs after they emerge from their source. In the sub-solidus convection regime, three modes of ascent are possible: (1) ascent as an isolated source of buoyancy, (2) ascent as a starting plume, and (3) ascent as a steady-state (or fully developed) plume. Fluid dynamical results were presented for cases (1) and (2) during the last few years.

Numerical (Daly and Raefsky, 1985) and experimental (Ribe, 1983; Ansari and Morris, 1985) results were presented for the case
of a rigid sphere ascending through a fluid with a strongly temperature dependent, Newtonian viscosity. Two limits were found: at large Peclet numbers the sphere ascends at the Stokes velocity defined in terms of the far-field viscosity, while at small Peclet numbers the sphere ascends faster than predicted by this Stokes velocity because the diapir is lubricated by a thin, low-viscosity layer surrounding it. Daly and Raefsky (1985) argued that for reasonable mantle parameters, the diapir would only ascend a few radii before losing most of its initial heat.

Low Reynolds number diapirs were studied in laboratory experiments. Griffiths (1986a,b) studied diapirs with thermal buoyancy which were characterized by high Rayleigh numbers and moderately high Peclet numbers. The thermals maintained constant buoyancy (i.e. temperature anomaly times volume) through their ascent. As a thermal ascends, heat diffuses out and heats the surrounding fluid; consequently background fluid is entrained into the thermal. Thermals increase in radius but decrease in velocity as they ascend. Olson and Singer (1985) studied the “end member” case of thermals which had very high Peclet numbers (i.e. very high temperatures) so that diffusion and entrainment only played a minor role. Extrapolation of these experiments to the mantle revealed that only large (>200 km) and hot (> 200K) thermals would be able to traverse the entire mantle in 100 m.y. or less (Griffiths, 1986b).

These isolated spherical diapirs are only one of a number of styles mantle plumes could have; if the hot-spot source provides a constant flux of hot mantle, then more steady plume convection could occur. Unfortunately, well designed fluid dynamical studies have yet to be carried out on such steady state plumes with realistic rheologies. Similarity solutions were developed for steady-state plumes with temperature-dependent viscosity (Loper and Stacey, 1983), and with power-law viscosity (Loper, 1984b).

Hot-spots are associated with topographic swells (Crough, 1983) and it is envisaged that the geophysical signatures of these features will place constraints on the geometry and heat and mass fluxes of mantle plumes. A necessary prerequisite will be an understanding of the fluid dynamics of plume-lithosphere interactions and resulting effects on topography, gravity, surface heat flux, and seismic velocity variations. In this spirit, the uplift of the seafloor by plumes was modeled by Emerman and Turcotte (1983), Yuen and Fleitout (1985), and Olson and Nam (1986). The most successful attempt at matching the observations with a dynamic model was made by Olson and Nam (1986), who explored the fluid dynamics of chemically buoyant spheres ascending and stagnating under moderately viscous lithospheres. By scaling their results to the mantle, they found that a spherical diapir of about 200 km radius with about 1% density contrast could satisfy the rate of uplift, maximum uplift, and topographical profile for the Hawaiian swell. Realistic models will require thermal buoyancy, temperature dependent viscosity, and a very viscous lithosphere, so that the process of thermal erosion and lithospheric convective instability can be understood.

Massive Computations: Large-scale computations with time dependence (e.g. Machetel and Yuen, 1986), and realistic rheologies and geometries, including three dimensions, are needed to investi-
gate many of the problems discussed here. This requires access to supercomputers, which typically have architectures differing from the usual sequential scaler machines, and development of software appropriate for these architectures. Many of the computations referenced in this section were made possible through access to supercomputers with sequential but vector architectures. A recent exciting technical advance has been the development of massively parallel computer architectures, termed "hypercubes." Multidimensional continuum mechanics problems fit naturally into the hypercube architecture (e.g. Fox and Otto, 1984). Lyzenga et al (1986) have shown how finite element problems can be efficiently solved using hypercubes, with each processor in the hypercube being responsible for a region in the physical domain of the continuum. The future prospect for this type machine seems bright, as there appears to be no limit to their size and power.

It is crucial to validate new software through detailed comparison with previously established results. Schubert and Anderson (1985) presented high-resolution solutions of two-dimensional convection at high Rayleigh number. More work of this sort, with readily accessible solutions, is needed to provide a standard such that the accuracy of software can be established. Remarkably little work of this sort has been done, particularly with temperature and stress dependent rheologies and for time dependent flows.

Geochemical Observations and Convective Mixing

Important insight into mantle convection is gained from geochemical observations and interpretation. Isotopic systematics of mantle derived rocks provide constraints on time-integrated properties of the convecting system over billions of years. New observations and syntheses of large data sets furthered our understanding of the mantle. In particular, White (1985) and Zindler and Hart (1986) showed that at least four or five mantle components are needed to explain the variability in the Nd, Pb, Sr, and He isotopic systematics of oceanic basalts. The implications are immediate: even if the mantle is convectively layered, at least one of the layers must have geochemical heterogeneities within it (cf. Davies, 1984a). Combined with the well known observation that the isotopic variations have taken billions of years to develop (cf. Zindler and Hart, 1986), the new observations require that heterogeneities survive in a convecting layer for billions of years.

The length scale of mantle heterogeneity was expanded by Hart (1984) who showed that there is a global Pb-Sr isotopic anomaly (coherent through OIB and MORB) centered in the Indian Ocean; Hart argued that the anomaly has existed as a mantle entity for billions of years. Moreover, our grasp of mantle heterogeneity became much more quantitative with the assembly of the size spectrum of isotopic heterogeneity in oceanic basalts (Gurnis, 1986a; Zindler and Hart, 1986); the new size spectrum quantifies the well known observation that isotopic variability exists over all possible length scales.

New insight into the ultimate ancient source for oceanic islands came from studies of variations in the Nb/U and Ce/Pb ratios (Hofmann et al, 1986). The Nb/U ratio of all oceanic basalts was shown to be surprisingly uniform but distinct from chondrites and continental crust. This implies that the source of oceanic basalts can be neither
primordial mantle nor subducted continental crust; the hypothesis that the source of OIB arises from an internal mantle differentiation process is reinforced. Subducted ancient oceanic crust is one hypothesized OIB source which fulfills these new geochemical requirements (Hofmann et al, 1986).

In order to relate what we now know about the geochemical state of the mantle with what we know about convection, significant effort went into trying to understand the fluid dynamics of convective mixing. Geochemical observations are sensitive to the survival of reservoirs (heterogeneities); the survival time of reservoirs can be calculated with geophysical models characterized by convective velocities and length scales. Olson et al (1984a,b) studied the survival of passive heterogeneities in steady cellular flow, concluding that most of the mixing occurred by shear strains (laminar mixing), instead of by normal strains (turbulent mixing). Hoffman and McKenzie (1985) studied mixing of large-scale passive heterogeneities by an unsteady, high Rayleigh number flow, concluding that because the surface area of the heterogeneities exponentially increased with time, the heterogeneity was turbulently mixed. However, Gurnis (1986b) showed, using a much simpler model, that the stirred heterogeneity consisted of both thin streaks and blobs. The thin streaks exponentially lengthen with time because they are long and are frequently advected into stagnation points. However blobs, which are smaller than the scale of flow, are primarily stirred by laminar flow within the interior of convection cells (Gurnis, 1986b). Gurnis (1986c) and Gurnis and Davies (1986a) modeled the mixing of a thin sheet of fluid introduced at the downwelling limb of thermal convection flows constrained by generalized features of plate kinematics and found that a significant fraction of the fluid is mixed by simple shear strains. Kellogg and Turcotte (1986), on the other hand, modeled the mixing of subducted lithosphere by a sheet of fluid flowing into a stagnation point.

A problem related to geochemistry and convection is the residence time of the mantle, i.e. the average time the entire mantle (or the upper mantle if stratified) is sampled by ridge magmatism (incorporation into the oceanic lithosphere). Gurnis and Davies (1986a) numerically calculated residence times in two-dimensional thermal convection simulations. Allegre and Turcotte (1986) and Gurnis and Davies (1986a), developed analytical models for residence times and calculated mean ages for the mantle in the range 1 to 2 b.y. assuming whole-mantle convection, in good agreement with Pb-Pb and Rb-Sr isotopic systematics (c.f. Zindler and Hart, 1986). Gurnis and Davies (1986b) investigated the effect which a high viscosity lower mantle has on residence times and found that if the lower mantle were more than 100 times more viscous than the upper mantle, a significant fraction of the mantle could escape sub-ridge sampling for 4 b.y.; this raised the possibility that rare gas systematics are consistent with whole-mantle convection, contrary to the view of Allegre et al (1983), and O’Nions and Oxburgh (1983).

Small-scale convection

The existence of convection on a scale small compared to plate dimensions has been hotly debated for some time, but the issue seems to have been resolved through geophysical observations, at least in
some tectonic environments. The most direct evidence comes from seismic tomography, where Humphreys et al (1984) found a curtain of high velocity material in the mantle beneath the Transverse Ranges in Southern California, juxtaposed against low velocity material beneath the Salton Trough ~ 200 km to the south. Haxby and Weissel (1986) interpreted linear features in the gravity field determined from SEASAT altimeter data over 5 Myr old oceanic lithosphere having similar wavelengths as being due to smallscale convection resulting from instability of the cooling lithosphere. Such rapid onset of convection at short wavelengths requires a low viscosity asthenosphere of order $10^{18}$ Pa S, as might be expected due to the competing effects of temperature and pressure on rock rheology (Yuen and Fleitout, 1984, 1985; Buck, 1985; Jaupart and Parsons, 1985).

While stability analyses of the cooling lithosphere have traditionally concentrated on the growth of infinitesimal perturbations, geological processes can result in finite amplitude thermal perturbations. Examples include transform faults (Craig and McKenzie, 1986), rifts (Steckler, 1985; Buck, 1986), convergence zones (Buck and Toksoz, 1983), and interaction of plumes with the lithosphere (Yuen and Fleitout, 1985). Such finite amplitude perturbations make the lithosphere much more susceptible to convective instability. A common practice in searching for small-scale convection using the geoid has been to high-pass filter the geoid, throwing away all terms with spherical harmonic degree less than 11. Renkin and Sandwell (1986) pointed out that the regular features of wavelengths ~ 3500 km in the residual geoid are mathematical artifacts from ringing caused by application of a sharp edged filter, not physical manifestations of small-scale convection, as previously claimed.

Thermal Evolution of Earth and Planets

Since sub-solidus convection is the dominant heat removal process for the earth and other solid planets, an understanding of mantle convection is necessary to understand their thermal evolution. In previous years, this topic was the subject of active study using “parameterized” convection models, but now the validity of some assumptions are being questioned.

Richter (1984a,b) presented a parameterized convection model including continent-ocean differences and pointed out that apparently discrepant results of previous studies were due mainly to different estimates for relevant parameters. Christensen (1985b) pointed out how crucial the choice of reference viscosity in the Rayleigh number is in the parameterization of convective heat loss. We interpret his study as showing that if most of the mechanical dissipation associated with plate motions occurs in the lithosphere, rather than in the remainder of the mantle, the rate of convection, and hence the thermal evolution of the planet, can be insensitive to interior viscosity and temperature, depending only on the effective viscosity of the lithosphere. This would be the case if average stresses associated with plate boundaries are of order kilobars.

The thermal history of planets is related to important geological problems such as the generation of magnetic fields and planetary differentiation. Stevenson et al (1983) studied the relation between magnetism and the thermal evolution of the terrestrial planets with particular emphasis on the existence or absence of magnetic fields
on Mercury, Venus, Earth and Mars and the different paths the thermal evolution of these planets may have taken. Gurnis and Davies (1986d) studied the relation between continental crust formation and the thermal history of the mantle and concluded that the overall features of the crustal age distribution are consistent with a smoothly evolving, but decaying, thermal history.

An exciting observational constraint on the thermal evolution of the mantle was the discovery by Richardson et al (1984) of 3 Byr old diamonds from a depth of 150km beneath the South African craton, requiring cold mantle to have existed at depth at the same time that komatiites indicate mantle temperatures several hundred degrees in excess of those today (e.g. Glikson, 1983). Explaining such long-lived continental roots in the presence of hot mantle presents a challenge to mantle dynamicists.

Extraction of melt from the mantle

Percolation of melt along grain boundaries is an important process transferring mass and heat in the interior, acting as the initial stage of melt segregation leading to igneous activity near the surface. For rocks of mantle composition, partial melts form an interconnected set of tubules along grain edge triple junctions (e.g. Jurewicz and Jurewicz, 1986; von Bargen and Waff, 1986). In order for melt to move relative to the surrounding crystalline matrix, continuity requires either more melt to be produced to take its place or the matrix to collapse, analogous to squeezing a sponge. Much effort has been directed towards quantitative modelling of the latter process, coupling Darcy’s law to the equations describing compaction of the matrix (e.g. McKenzie, 1984; Richter and McKenzie, 1984). Depending upon the material properties matrix deformation is thought to be important on scales of meters to kilometers. Under certain circumstances, e.g. a sudden increase in rate of supply of melt to the base of a layer, solitary waves of high porosity and melt content result, christened “magmons” by Scott and Stevenson (1984). Analogies to these one-dimensional solitary waves can be demonstrated in the laboratory using viscous conduits (Scott et al, 1986; Olson and Christensen, 1986; Whitehead and Helfrich, 1986). The solitary wave behavior has been demonstrated in two dimensions (Scott and Stevenson, 1986), although the waves do not strictly preserve their form upon interaction and are thus not true solitons (Barcilon and Richter, 1986). The challenge in this field is now to relate the elegant models to geological or geophysical observations (e.g. McKenzie, 1985).

The density contrast between melt and mantle is important in governing the rate and even the direction of melt migration. Experiments reported by Rigden et al (1984) give added support to the hypothesis that at high pressures, the density contrast between melt and residual changes sign. The dynamics of convection in such a system have yet to be addressed.

When melt accumulates near the surface, diapiric motion may become the important transport process. The temperature dependence of rock rheology then becomes important in determining how fast a diapir rises and how far it can go before solidifying, as discussed previously. Multiple diapirs seem to be required to heat a conduit to the surface. Alternatively, magma-driven propagation of cracks may
be another important mechanism for transport of magma through the lithosphere (Spence and Turcotte, 1985).
Other papers of note include:


Is the Mantle Layered?: Christensen and Yuen (1985); Davies (1983); Jackson (1983); Jeanloz (1986); Kato (1986); Kenyon and Turcotte (1983); Knittle et al (1986); and Lange and Ahrens (1984).

D" and the Core-Mantle Boundary: Bloxham and Gubbins (1985); Cormier (1985); Cormier and Choy (1986); Doornbos (1983); Doornbos et al (1986); Gire et al (1986); Gubbins and Richards (1986); Knittle and Jeanloz (1986); Lay (1986b); McFadden and Merrill (1984); Menke (1986); Stacey and Loper (1983); and Zhang and Lay (1984).

Fluid Dynamics of Convection: Boss (1983); Boss and Sacks (1984, 1986); Boss et al (1984); Busse (1983); Busse and Frick (1985); Carrigan (1985); Davies (1986); Fowler (1984, 1985c); Frick et al (1983); Honda (1986); Houseman (1983a,b); Jarvis (1984); Kaula (1983); Officer and Drake (1983); Oliver and Booker (1983); and Zebib et al (1983).

Realistic Rheologies: Gurnis and Davies (1986d); Londe and Davies (1985); and Morris and Canright (1984).


Small-scale Convection: Fleitout and Yuen (1984a,b); Spohn and Schubert (1983); and Stark and Forsyth (1983).


Extraction of Melt from the Mantle: Cooper and Kohlstedt (1984, 1986); Emerman and Turcotte (1984); Fowler (1985a); Jurewicz and Watson (1985); Nickel and Green (1985); Ribe (1985a,b; 1986); Richter (1985b); Stevenson (1986); and Whitehead et al (1984).

Plate motions and Driving Mechanism: Carlson (1983); Carlson et al (1983); Davis and Solomon (1985); Gordon (1983); Jurdy and Gordon (1984); Richardson and Cox (1984); Schult and Gordon (1984); and Wiens and Stein (1985).


Acknowledgements

Critical reviews by D. L. Anderson, P. Olson and D. R. Scott are much appreciated. This work was sponsored by a grant from Ametek Corporation and NASA grant NAG5-315. Contribution number 4453, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125.
References


Bowin, C., G. Thompson, and J. G. Schilling, Residual geoid anomalies in Atlantic Ocean Basin: Relationship to mantle plumes, *J.*


Carlson, R. W., Plate motions, boundary forces, and horizontal temperature gradients: Implications for the driving mechanism, Tectonophysics, 99, 149-164, 1983.


Davies, G. F., Subduction zone stresses; constraints from mechanics and from topographic and geoid anomalies, Tectonophysics, 99, 85-98, 1983b.


Griffiths, R. W., Dynamics of mantle thermals with constant buoyancy or anomalous internal heating, Earth Planet. Sci. Lett., 78, 435-446, 1986b.


Kenyon, P. M. and D. L. Turcotte, Convection in a two-layer mantle with a strongly temperature-dependent viscosity, J. Geophys. Res.,


in length of day, EOS, Trans. AGU, 67, 908, 1986.
Spohn, T. and G. Schubert, Convective thinning of the lithosphere;
a mechanism for rifting and mid-plate volcanism on Earth, Venus
Stacey, F. D. and D. E. Loper, The thermal boundary layer interpre-
tation of D'' and its role as a plume source, Phys. Earth Planet.
Int., 33, 45-55, 1983.
Stacey, F. D. and D. E. Loper, Thermal histories of the core and
Stark, M. and D. W. Forsyth, The geoid, small-scale convection,
and differential travel time anomalies of shear waves in the central
Stark, P. B. and C. Frohlich, The depths of the deepest earthquakes,
Stevenson, D. J., Planetary magnetic fields, Rep. Prog. Phys., 46,
555-620, 1983.
Stevenson, D. J., On the role of surface tension in the migration of
Stevenson, D. J., T. Spohn, and G. Schubert, Magnetism and thermal
Tanimoto, T., The Backus-Gilbert approach to three-dimensional
structure in the upper mantle: II. SH and SV velocity, Geophys. J.
Tanimoto, T. and D. L. Anderson, Mapping convection in the mantle,
Tanimoto, T. and D. L. Anderson, Lateral heterogeneity and az-
imuthal anisotropy of the upper mantle: Love and Rayleigh waves,
Turcotte, D. L. and J. C. Pflugrath, Thermal structure of the accreting
Vassiliou, M. S., B. H. Hager, and A. Raefsky, The distribution of
earthquakes with depth and stress in subducting slabs, J. Geody.
1, 11-28, 1984.
von Bargen, N. and H. S. Waff, Permeabilities, interfacial areas
and curvatures of partially molten systems: Results of numerical
computations of equilibrium microstructures, J. Geophys. Res., 91,
9261-9276, 1986.
Walck, M. C., The P-wave upper mantle structure beneath an active
Walck, M. C., The upper mantle beneath the north-east Pacific rim:
A comparison with the Gulf of California, Geophys. J. R. Astron.
Ward, S. N., Small-scale mantle flows and induced lithospheric stress
Watts, A. B., D. P. McKenzie, B. E. Parsons, and M. Roufosse, The
relationship between gravity and bathymetry in the Pacific Ocean,