27 January 1984, Volume 223, Number 4634 **SCIENCE**

The Earth as a Planet: Paradigms and Paradoxes

Don L. Anderson

Sit down before fact as a little child, be prepared to give up every preconceived notion, follow humbly wherever and to whatever abysses nature leads, or you shall learn nothing.—From the *Life and Letters of Thomas Henry Huxley*

Most investigators concerned with global problems tend to work with a set of standard assumptions about how the interior of the earth works. These assumptions can be summarized in the following paradigm.

Summary. The independent growth of the various branches of the earth sciences in the past two decades has led to a divergence of geophysical, geochemical, geological, and planetological models for the composition and evolution of a terrestrial planet. Evidence for differentiation and volcanism on small planets and a magma ocean on the moon contrasts with hypotheses for a mostly primitive, still undifferentiated, and homogeneous terrestrial mantle. In comparison with the moon, the earth has an extraordinarily thin crust. The geoid, which should reflect convection in the mantle, is apparently unrelated to the current distribution of continents and oceanic ridges. If the earth is deformable, the whole mantle should wander relative to the axis of rotation, but the implications of this are seldom discussed. The proposal of a mantle rich in olivine violates expectations based on evidence from extraterrestrial sources. These and other paradoxes force a reexamination of some long-held assumptions.

Convection in the mantle is steady (*1*) and is driven by thermal expansion resulting from self-heating or heat from the core. Plate tectonics, or something like it, including deep subduction, has operated throughout most of geologic time. Continents are passive elements in the system. Apparent polar wandering is always dominated by plate motions rather than motion of the whole mantle relative to the pole. Basalts are partial melts of an olivine-rich garnet-peridotite mantle. Midocean ridge basalts, or tholeiites, are generated in the partially molten low-velocity zone above a depth of 200 km in the oceanic mantle and undergo little crystal fractionation or contamination before eruption. Hot spots and mantle plumes originate in a "primitive" lower mantle. Alkali-rich magmas represent small degrees of partial melting deep in the mantle. Tholeiites, the most abundant basalt type, represent large degrees of partial melting of a shallow source. Olivine and its high-pressure forms are the dominant minerals of the mantle. Discontinuities in the mantle result from phase changes in olivine. The upper and lower mantles have the same bulk chemistry. The crust is the main repository of incompatible elements.

There are now reasons to suspect every assumption stated above. Indeed the paradoxes associated with these assumptions suggest an alternative scenario such as the following.

Mantle convection is episodic and is driven by lateral temperature gradients caused by continental insulation (*2*). Most of the buoyancy differential that drives mantle convection is provided by partial melting and the basalt-eclogite phase change rather than thermal expansion. Overridden oceanic lithosphere cools the upper mantle, generating geoid lows and regions of high seismic velocities. Active subduction and stagnant continents generate geoid highs, which reorient the mantle relative to the spin axis.

Continents move away from thermal highs and come to rest over cold parts of the mantle; sea-floor spreading and subduction slowdown or cease until the subcontinental mantle warms up. The whole mantle shifts relative to the axis of rotation to accommodate the changing geoid. The outer shell of the earth is stressed to failure by the migration of the equatorial bulge. Episodes of rifting, continental drift, and subduction follow periods of rapid polar wandering, which follow periods of continental stagnation.

As a result of early terrestrial differentiation the mantle is chemically stratified; some layers are refractory residues remaining after melt extraction, some are cumulates, and some are rich in residual fluids. The lower mantle is depleted in the radioactive and crust-forming elements. Radioactive elements are concentrated in the upper mantle as well as in the crust. A thick, buoyant crust is impossible on the earth because of pressure-induced phase changes. The effect of pressure on the melting point is such that melting can only occur in the upper mantle. Diapirs originate in thermal boundary layers at chemical discontinuities in the upper mantle.

Tholeiites evolve from picrites, which represent large degrees of partial melting in a region of the mantle that is depleted in large-ion lithophile elements (*3*), rich in garnet, and compositionally distinct from the shallow mantle and the lower mantle. These magmas provide the bulk of the material that erupts at midocean ridges and hot spots, and they undergo varying degrees of crystal fractionation and contamination before eruption. The characteristic trace element and isotopic

The author is a professor of geophysics and director of the Seismological Laboratory, California Institute of Technology, Pasadena 91125.

signature of plume or hot-spot magmas is acquired in the low-velocity zone of the upper mantle. Olivine is the dominant mineral only in the shallow mantle. The bulk of the mantle is pyroxene and garnet and their high-pressure equivalents.

In this article I will explore some of these alternative ways of looking at the earth. A brief tour of the inner solar system will show that extraterrestrial concepts can now be brought down to earth, perhaps to displace indigenous ideas.

The Earth from an Extraterrestrial Perspective

Current theories suggest that the early histories of the moon and the earth were quite different. The moon was very hot early in its history and is a highly differentiated body. The early lunar scenario has giant plagioclase "rockbergs" floating on a global magma ocean. The freezing of this ocean established the igneous stratigraphy and resulted in strong upward concentration of incompatible trace elements, culminating in KREEP (a potassium-, rare earth element–, and phosphorus-rich lunar rock type) (*4*), the final dregs of the crystallizing magma ocean, and a depleted lower mantle. Some lunar basalts appear to be remelted from cumulate layers that settled out of the original magma ocean. Pyroxene may be the dominant mineral in the lunar mantle.

The bulk of the earth's mantle is usually assumed to be an olivine-rich but "primitive" material which, by heating up, provides melts to the various volcanic centers. But this leaves unexplained phenomena such as the disappearance of the geologic record for the time that the moon was a violent cauldron; the absence of an extensive, ancient, anorthositic crust; the small amount of terrestrial crust in comparison with other bodies; the high Mg/Si ratio of the earth relative to cosmic abundances; and the lack of differentiation in the bulk of the mantle. Since the earth has cooled extensively in the past 4.5×10^9 years and is still near or above the solidus at oceanic ridges, island arcs, continental rifts, and hot spots, one can view its mantle as a slowly cooling, crystallizing system, whereas the moon, which is a much smaller body, completed its evolution much earlier. Some magmas on the earth may represent residual melts, rather than initial melts of a previously solid source region, and others may be the products of the remelting of cumulate layers, as on the moon. If the earth's mantle really has a high olivine content, then we have a major cosmochemical paradox to explain.

Small bodies such as the moon and Mars, although volcanically active in their early history, now appear to be tectonically dead. This is presumed to be the result of rapid cooling and development of a lithosphere too thick and buoyant to break and subduct. Venus, about the same size as the earth, was expected to be a tectonic twin of the earth, but it also lacks the obvious signs of plate tectonics such as long linear ridges and trenches. If anything, because of the high surface temperatures, the lithosphere on Venus should be thinner, weaker, and more mobile than that on the earth. However, low pressures, as found on small planets, and high temperatures, as on Venus, both favor the development of a thick, buoyant crust. If the earth were smaller, hotter, much younger, or much older, conditions apparently would not be appropriate for plate tectonics.

The accretional energy of a terrestrial planet is high enough to melt or partially melt infalling silicate material. A young, hot planet is coated by a basaltic scum, and the deep interior is depleted in the components with low melting points. The thickness of the crust depends on the depth at which the geothermal gradient intersects either the melting point or the basalt-eclogite transition. When a planet is thermally young, with a high thermal gradient or a hot atmosphere, the crust is underlain by a partial melt zone, and overthickened basaltic crust simply remelts. As the planet cools, the temperatures in the upper mantle drop, and the lower part of the original basaltic crust converts to dense eclogite. Therefore, on a planet the size of the earth with a similar thermal gradient, a thick crust is impossible. Cooling makes the original crust unstable, and a massive overturn can be expected. This may be why there is no record of the first 700 million years of the earth's geologic history. On a small planet, the basalt-eclogite boundary is pushed to great depth because of the lower pressures, and it is possible for a thick low-density crust to form. Since it is difficult to break and subduct a thick buoyant plate there is no terrestrial-style tectonics, even if the interior is convecting. On a hot planet, such as Venus, the basalt-eclogite transition is also deep, possibly below a partial melt layer, and again no deep subduction or terrestrial type of plate tectonics can occur. On such a planet, buoyant basaltic plates float around the surface much as pack ice does in the polar oceans. Venus may provide a good analog for the hot Precambrian earth. Heat is removed from such a planet by conduction through the surface thermal boundary layer and isolated plumes. On the present earth, the upper mantle is cooled mainly by the subduction of cool oceanic lithosphere; plumes occur in regions where little of the thick cold lithosphere has subducted in the last 100 million years.

It has often been suggested that life originated on the earth because of a coincidence between the narrow temperature interval over which water is liquid and the temperature extremes that actually occur on the earth. The earth apparently is also exceptional in having active plate tectonics. If the carbon dioxide in the atmosphere of Venus could turn into limestone, the surface temperatures and those in the upper mantle would drop. The basalt-eclogite phase change would migrate to shallow depths, causing the lower part of the crust to become unstable. Thus there is the interesting possibility that plate tectonics may exist on the earth because limestone-generating life evolved here.

The closest terrestrial analog to the lunar rocks called KREEP is the rare rock type kimberlite, which explodes from depths as great as 200 km. Kimberlite has also been interpreted as a residual melt (*5*), although it is more commonly assumed to be the result of small degrees of partial melting of a garnet lherzolite (*6*). The complementary nature of the patterns of large-ion lithophile elements in kimberlite and in midocean-ridge basalts is as striking as that of the patterns in lunar anorthosites and mare basalts, the cornerstone of the lunar magma ocean hypothesis. The main difference between kimberlite and KREEP is that the former has been in contact with eclogite, a dense high-pressure aluminous rock that is stable in the earth's upper mantle, whereas the latter has been in contact with plagioclase, a light low-pressure aluminous mineral that is stable in the lunar upper mantle. If kimberlite and eclogite are the analogs of lunar KREEP and anorthosite rather than terrestrial curiosities, a terrestrial magma ocean is a distinct possibility. Although kimberlites are rare, most basalts are intermediate between kimberlite and midocean ridge basalts in their large ion lithophile content and isotopic chemistry, suggesting that such magmas are blends.

The lunar crust represents 10 percent of the planet; the earth's crust is less than 0.5 percent of the planet. The anorthositic highlands and thick crust on the moon are evidence of early and widespread differentiation. The absence of such an early anorthositic crust on the earth has been taken as evidence that extensive melting and crystal separation did not occur. The high pressures in the earth, however, mean that the early crystallizing aluminous crystals are garnet and clinopyroxene, which sink. Therefore, a buried eclogite cumulate layer is the terrestrial equivalent of the lunar crust, and the early thermal history of the earth and the moon might not be so different. This would explain the earth's thin crust—most of the "crust" is buried and always has been.

The thickness of the continental crust and the maximum thickness of the oceanic lithosphere are roughly the same as the depth at which aluminous rocks convert to eclogite. Delamination and replacement of dense eclogite by the lighter olivine-rich underlying peridotite may therefore determine the thickness of the terrestrial crust and the composition and buoyancy of the continental lithosphere. The concentration of olivine in the shallow mantle would explain the bias of terrestrial petrologists toward an olivine-rich mantle.

The largest known positive gravity anomaly on any planet is associated with the Tharsis volcanic province on Mars. Both geologic and gravity data suggest that the positive mass anomaly associated with the Tharsis volcanoes reoriented the planet with respect to the spin axis, placing the Tharsis region on the equator (*7*). There is also evidence that magmatism associated with large impacts reoriented the moon (*8*). The largest mass anomaly on the earth is centered over New Guinea, and it is also almost precisely on the equator. The long wavelength part of the geoid correlates well with subduction zones and hot-spot provinces (*9*), and these in turn appear to control the orientation of the mantle relative to the spin axis. Thus, we have the possibility of a feedback relation between geologic processes and the rotational dynamics of a planet. Volcanism and convergence cause mass excesses to be placed near the surface. These reorient the planet, causing large stresses that initiate rifting and faulting, which in turn affect volcanism and subduction. Curiously, earth scientists have been more reluctant to accept the inevitability of true polar wandering than to accept continental drift, even though the physics of the former is better understood.

Our nearest neighbors in space have given us alternative views of how a planet might operate. Global reorientation may be responsible for initiating rifting and subduction and periods of rapid apparent polar wandering. Large-scale melting, upon accretion, followed by crystal settling and migration of residual fluids similar in composition to kimberlite is a plausible scenario for the formation of isolated geochemical reservoirs. A thick, buoyant, nonsubducting crust is likely on a small planet or a hot planet such as Venus or the early earth. The absence of a terrestrial crust of lunar proportions is not paradoxical. Such a crust is impossible because of pressure-induced phase changes.

Global Petrology

The isotopic ratios of basalts are inconsistent with a homogeneous, primitive, or well-stirred mantle. The nature of the inhomogeneity, however, is in dispute. In an accreting planet the temperature of the upper mantle is above the solidus (*10*), and the planet is zone refined as it grows, with melts and incompatible elements concentrated near the surface and dense residual crystals depleted in large-ion lithophile elements concentrated toward the interior. Even if differentiation and gravitational separation do not occur on a planet during accretion, the relative slopes of the adiabat and the melting point ensure that melting in the upper mantle will take place in the rising limbs of convection cells and that residual, depleted, refractory material returning to the lower mantle will be far below the melting point at the pressures found in the lower mantle. Thus, either during accretion or by convection of the mantle after accretion, the entire mantle is processed through the melting zone in the upper mantle. Freezing of the upper mantle leaves, on a planet the size of the moon, a thick crust rich in plagioclase and, on one the size of the earth, a dense aluminous cumulate layer at depth.

Modern midocean ridge basalts are extremely depleted in the large ion lithophile elements such as rubidium, strontium, uranium, and light rare earth elements. The values of isotopic ratios such as $^{87}Sr/^{86}Sr$ show that the source of midocean ridge basalts must have been depleted in the large ion lithophile elements for at least 10^9 years (*11*). A cumulate layer formed by crystal settling from a crystallizing magma ocean would become depleted as the late-stage residual fluids were expelled.

Differentiation of the mantle separates phases rich in Al_2O_3 such as basalt and eclogite from refractory phases poor in Al_2O_3 such as ultramafic rocks. Basalts are light and tend to rise; eclogites tend to sink. There is disagreement about the relative portions of the basalt-eclogite fraction of the mantle and whether it is dispersed or is mainly in an eclogite layer somewhere at depth. Most estimates of bulk mantle chemistry give basalt fractions of 20 to 30 percent. Did the mantle yield up its basalt fraction during accretion and the "missing" 700 million years of early geological history, or has primitive mantle survived to the present with the ability to yield basaltic magmas, for the first time, by 20 to 30 percent partial melting? And to what depth will subducted basalt or eclogite cumulates sink into the mantle?

There are three possible fates for cold subducted lithosphere. It could end up in the lower mantle. It could bottom out at the 650-km discontinuity and form a layer at the base of the transition zone (*12, 13*). The third alternative is that the top part of the subducted plate warms up, becomes buoyant, and either settles under the continental lithosphere or mixes with the shallow mantle, contributing to enrichment of this region in large ion lithophile elements. If old oceanic lithosphere stays in the upper mantle, it should generate geoid lows and areas with fast seismic velocities since it serves to cool the mantle. The part of the lithosphere with Al_2O_3-poor peridotite stays in the shallow mantle because of its buoyancy. If the basalt-eclogite portion of the mantle is to stay in the upper mantle, it must be less dense than the lower mantle. This depends on the stability field of garnetite, a high-pressure form of eclogite. Garnet is stable to great depth in silicates rich in Al_2O_3, such as eclogite, but eventually collapses to a dense perovskite structure. The garnet to perovskite transformation, however, is not complete in such material until it is at a depth of about 750 km (*14*). Furthermore, perovskite that is rich in Al_2O_3 has a relatively low density (*14*). One concludes that eclogite will not be able to penetrate the 650-km discontinuity and should pile up at this depth.

The broad stability field of garnet in material rich in Al_2O_3 is responsible for the high density of eclogite, in comparison with peridotite, in the upper mantle. The high pressure needed for the transformation of garnet to perovskite in Al_2O_3-rich material and the low density of CaO- and Al_2O_3-rich perovskite, in comparison to that rich in $(Mg,Fe)SiO_3$, prevents it from sinking into the lower mantle. The sharpness of the 650-km discontinuity and absence of earthquakes below this depth suggest that this is the boundary between eclogite and depleted lower mantle. Differentiation of the mantle appears to be irreversible.

In the conventional view of mantle chemistry, olivine or its high-pressure forms dominate, and the basaltic fraction is dispersed throughout the mantle. In cosmochemical models, pyroxenes and garnets are the dominant minerals, and the lower mantle is mainly perovskite. Garnet and perovskite have distinct elastic properties, and it is possible to use seismic techniques to prospect for buried eclogite layers and to test whether the transition region and the lower mantle are similar in composition to the olivine-rich shallow mantle. The long thermal time constant of subducted lithosphere means that it should show up in global gravity and seismic velocity maps. Parts of the Pacific plate that subducted 50 to 100 million years ago, for example, should now be under eastern North America, Brazil, and the western Atlantic Ocean. These regions are, in fact, in geoid lows and have high mantle velocities.

Fine Structure of the Upper Mantle

High resolution seismic profiling has recently made it possible to address several important questions about the dynamics, petrology, and geochemistry of the mantle. These questions include the deep structure of continents, the depth of the midocean ridge reservoir, the chemical stratification of the mantle, and the composition of the deep mantle. From velocity profiles for various tectonic regions (Fig. 1), several points can be noted. In the mantle under the shield, velocity is faster than in other regions to a depth of about 150 km, and then there is a velocity decrease. The velocity and velocity gradients in the shield lithosphere imply cold temperatures and a high thermal gradient, features characteristic of a conductive thermal boundary layer. The shield lithosphere therefore acts as a cold insulating lid on top of the convecting mantle. The low velocities under the East Pacific Rise extend to 400 km. This is an important new constraint on the origin of midocean ridges, the location of the depleted reservoir, and the source of the oceanic crust. The high ratio of the velocity of compressional waves (V_P) to that of shear waves (V_S) between depths of 90 and 150 km suggests that melting is most extensive in this region, melting that presumably results from adiabatic ascent of material between 200 and 400 km.

The differences in seismic velocities between the mantle at the East Pacific Rise and that elsewhere are too great to be explained by changes in chemistry or temperature alone, at least in the upper 200 km. Partial melting, a decrease in garnet content, and high temperatures may all be involved. These effects serve to decrease density, facilitating the rise of material from below 200 km to the surface to form new oceanic crust and lithosphere.

Seismologists have recently been forced to invoke anisotropy and anelastic dispersion in order to fit the increasingly precise seismic data. When these effects are taken into account, the velocities in the shallow mantle are substantially higher than had been thought and, in most regions, can be explained without invoking partial melting. Furthermore, subsolidus mechanisms have been discovered, in particular dislocation relaxation, that can lower the seismic velocities and viscosity in the shallow mantle (*15*). The interpretation that the low-velocity zone is the result of partial melting led naturally to the idea that it was the source region for the voluminous midocean-ridge basalts. The low velocities associated with the East Pacific Rise, however, suggest a much deeper source. Those associated with the Yellowstone plume (*16*) can be traced to about 250 km. Kimberlites, the plume-type magmas most extreme in large-ion lithophile elements and isotope chemistry, appear to originate near 200 km.

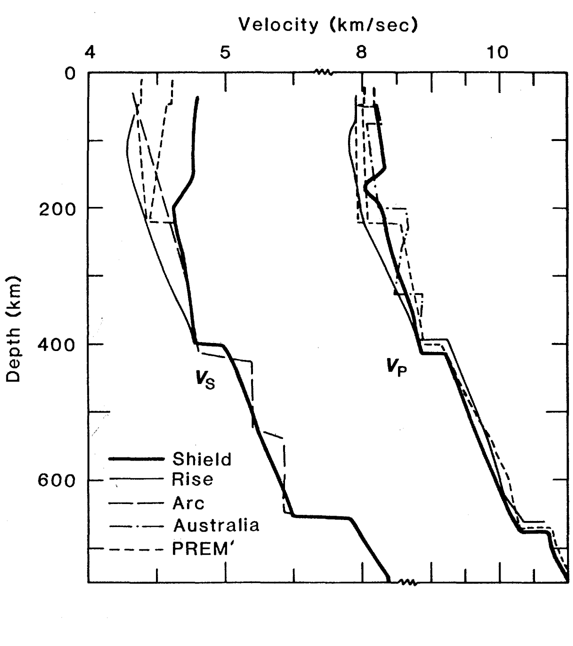

Fig. 1. Compressional (V_P) and shear (V_S) velocities as a function of depth for various tectonic provinces (*35–38*). The shield structures have velocities that are faster than those in other regions to depths of 150 to 220 km. The complications in this depth range probably result from a change in chemistry and the associated thermal boundary layer. The rise model is for the northern part of the East Pacific Rise and is slow to a depth of 400 km, suggesting a deep origin for midocean-ridge basalts. The PREM′ model is a modification of the average earth model PREM, which is anisotropic above 220 km; the two short dash curves for PREM′ represent horizontal and vertical propagation. The high gradients between 400 and 650 km may be due to the diopside-majorite transition.

The Transition Region

In the classical mantle models of Harold Jeffreys and Beno Gutenberg the velocity gradients between 400 and 800 km were too high to be the result of self-compression; hence, it was called the transition region and was interpreted by Francis Birch as a region of phase changes. This region was later found to contain two major seismic discontinuities, one near 400 km and one near 650 km (*17*), which were initially attributed to the olivine-spinel and spinel-post-spinel phase changes, respectively, in an olivine-rich mantle (*18*). These phase changes occur over depth intervals of about 20 km and therefore result in diffuse seismic boundaries rather than sharp discontinuities (*18, 19*). It was subsequently found that the 650-km discontinuity is a good reflector of seismic energy, requiring that its width be less than 4 km (*20*) and that the large increase in elastic properties was not consistent with any phase change in olivine (*21*). The spinel-postspinel transformation therefore is not an adequate explanation for the 650-km discontinuity. There appears to be no phase change in a chemically homogeneous mantle that has the requisite properties. Furthermore, the seismic velocities in the lower mantle are more consistent with a high-pressure form of pyroxene rather than one of olivine (*22*).

The velocity gradients between 400 and 650 km are higher than expected for a homogeneous self-compressed region. This region may represent the gradual conversion of diopside and jadeite to Al_2O_3-poor garnet. In the presence of Al_2O_3-rich garnet, diopside is stable to much higher pressures than calcium-poor pyroxenes (*19*). The region between 650 and 750 km also has a high gradient and is probably also a mixed phase region.

The average seismic properties of the mantle between the base of the crust and about 220 km are consistent with an olivine- and orthopyroxene-rich aggregate such as peridotite or pyrolite (*6*). The composition of the mantle below 220 km is controversial. Figure 2 shows estimates of density, compressional velocity, and bulk modulus along a 1400°C adiabat for pyrolite, eclogite, β-spinel, γ-spinel, garnet, (Mg,Fe)O, majorite, and $MgSiO_3$ (perovskite). Majorite is a high-pressure Al_2O_3-poor garnet form of pyroxene; β-spinel, γ-spinel, and perovskite plus (Mg,Fe)O are successively higher pressure forms of olivine. The mantle between 220 and 400 km is consistent with eclogite, a rock composed mainly of clinopyroxene (diopside plus jadeite) and garnet, or a peridotite with more garnet than pyrolite. Between 400 and 500 km, pyrolite should be mainly β-spinel, garnet, and majorite. Between 550 and 650 km the stable phases in pyrolite are γ-spinel and majorite. These assemblages do not satisfy the seismic data for this region. In this interval eclogite transforms from a mixture of clinopyroxene plus aluminous garnet to a garnet solid solution (garnet plus majorite). This reaction occurs over a broad pressure interval.

The transition region therefore appears to be mainly garnetite, rather than olivine and its high-pressure forms. The high gradient between 400 and 650 km implies that the phase change is spread between 140 and 230 kilobars. Phase changes in $(Mg,Fe)_2SiO_4$ are not so spread out, and those that involve $(Mg,Fe)_3Al_2Si_3O_{12} \cdot (Mg,Fe)_4Si_4O_{12}$ garnetite occur at lower pressure (*3, 19*). The transformation of $CaMgSi_2O_6$ plus $(Ca,Mg)_{1.5}Al_2Si_3O_{12}$ (primary components of eclogite) to garnetite is not complete until 200 kbar at 1200°C (*19*). The system $CaSiO_3 \cdot nAl_2O_3$ also has a broad transition field between garnet solid solution and perovskite that covers the pressure range of the transition region (*23*). In either case a CaO- and Al_2O_3-rich transition region is implied, which is consistent with an eclogitic chemistry but not with a pyrolitic chemistry. A small amount of olivine and orthopyroxene may also be present in the transition region, but the dominant minerals appear to be garnet and clinopyroxene.

The lower mantle has seismic properties consistent with $(Mg,Fe)SiO_3$ in the perovskite structure, which would make it chemically distinct from both the transition region and the shallow mantle. An olivine-rich mantle would have appreciable (Mg,Fe)O, low compressional wave velocities, and low bulk modulus, K. A lower mantle of pyrolite can therefore be ruled out, as can an explanation for the 670-km discontinuity that relied solely on phase changes. The phase change γ-spinel to perovskite plus (Mg,Fe)O would lead to a decrease in K and compressional wave velocity at 670 km.

It is usually assumed that olivine is the most important mineral in the mantle since it is the primary constituent of mantle peridotites exposed on land or brought up in magmas and kimberlites. Eclogite xenoliths are common in some kimberlite pipes but generally are subordinate to olivine-rich assemblages. By far, the most common materials that emerge from the mantle, however, are the tholeiitic basalts from midocean ridges, oceanic islands, island arcs, and continental rifts. These are chemically equivalent to clinopyroxene and garnet, the primary eclogite minerals. A zone-refined earth, with the basaltic fraction concentrated upward, would have an upper mantle that is primarily basaltic or eclogitic and therefore rich in Al_2O_3 and CaO. Furthermore, this material, when solid, is denser than peridotite and is gravitationally stable at the base of the upper mantle. As discussed earlier, the low seismic velocities associated with the East Pacific Rise can be traced to a depth of 400 km. There are thus several reasons for suggesting that the olivine-rich shallow mantle is underlain by material rich in clinopyroxene and garnet, as appropriate for a chemically stratified mantle.

By dropping the standard assumption that the transition region and lower mantle are identical in chemistry to the upper mantle, we can explain the otherwise paradoxical seismic properties of the mantle, the bulk chemistry of the mantle, and the isolation of geochemical reservoirs. Olivine appears to be a minority phase in the mantle, as it is in cosmic abundances, and is concentrated in the shallow mantle.

Earth Tomography

Three-dimensional mapping of the earth's interior is progressing rapidly with techniques similar to medical tomography being used for imaging with seismic body waves and surface waves. Surface waves from large earthquakes orbit the earth many times and are used to map velocity variations in the upper mantle on a global basis (*24*). Because different wavelengths of the surface waves sample to different depths, the

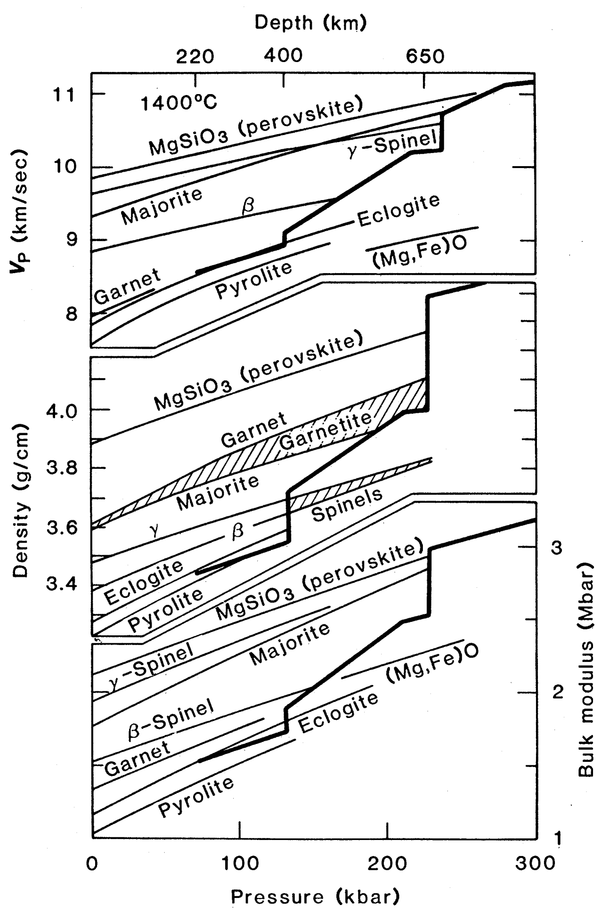

Fig. 2. Compressional velocity (V_P), density, and bulk modulus (K) of upper mantle (heavy line) for the earth model PREM compared with values calculated for various minerals and mineral assemblages along a 1400°C adiabat. Olivine is in the β-spinel structure near 400 km, in the γ-spinel structure at 650 km, and in the perovskite plus (Mg,Fe)O structure below 650 km. An olivine-rich mantle can therefore be ruled out over this depth range. An eclogite mantle should lie between garnet or eclogite and majorite (Al_2O_3-poor garnet) between 400 and 670 km, and this is permitted by the data. The lower mantle is best fit by perovskite, with only minor (Mg,Fe)O.

depth extent of anomalous structures can be determined. Figure 3 shows a map of phase velocities of 250-second Rayleigh waves, which are most sensitive to mantle velocities at depths between 200 and 400 km.

The fastest velocities are recorded from regions in the western Pacific, parts of the western and southern Atlantic, and the eastern Indian Ocean. These are probably the coldest regions of the upper mantle. The slowest velocities are recorded for western North America, the East Pacific Rise, Southeast Asia, the North Atlantic, the Indian Ocean triple junction, and the African rift, Red Sea, and Gulf of Aden regions. All these areas are experiencing extensional tectonics and must be deep-seated features.

The fastest velocities at long wavelength periods are recorded at convergence zones (*24–26*), where the upper mantle has been replaced by cold, downwelling material. In material beneath island arcs below about 200 km velocities are, on average, faster than elsewhere, including continental shields (*25*).

Melting as a Geodynamic Force

Melting is one of the more important driving forces in mantle convection. The partial melting process that generates basalts converts garnet and clinopyroxene to basaltic melt, with a decrease in density of about 10 percent. It would take a 3000° increase in temperature to equal this volume change by ordinary thermal expansion. Melting of eclogite converts it from the densest material in the upper mantle to a material that is gravitationally stable near the surface. Basalt intruded or subducted below a depth of 50 km converts to eclogite, giving it a negative buoyancy with respect to the surrounding mantle. The low velocities under the East Pacific Rise appear to be associated with the rise of a broad diapiric structure from the transition region.

On a gross scale the earth is compositionally stratified, with the crust, mantle, and core being the most obviously chemically distinct shells. A chemically layered, gravitationally stable mantle would be the natural end state of the earth's evolution. The presence of a crust and a core and the long-lived separation of the major magma reservoirs indicate that differentiation or separation is the dominant large-scale process, whereas convection can be expected to homogenize the layers between chemical discontinuities.

A peculiarity in the relative melting points and densities of mantle minerals precludes a permanent and static stratification of the upper mantle, at least in the present thermal regime. Eclogite is the densest assemblage in the upper mantle because garnet is the densest mineral under upper-mantle conditions. However, with heating or partial melting garnet is eliminated, and eclogite is completely molten at temperatures of less than 100° above the onset of melting. An eclogite layer is stable at depth as long as it is cold but, as it heats up and crosses the solidus, about 1800°C at a depth of 200 km (*27*), density decreases and, when about 10 percent has melted, it is less dense than the overlying peridotite. A

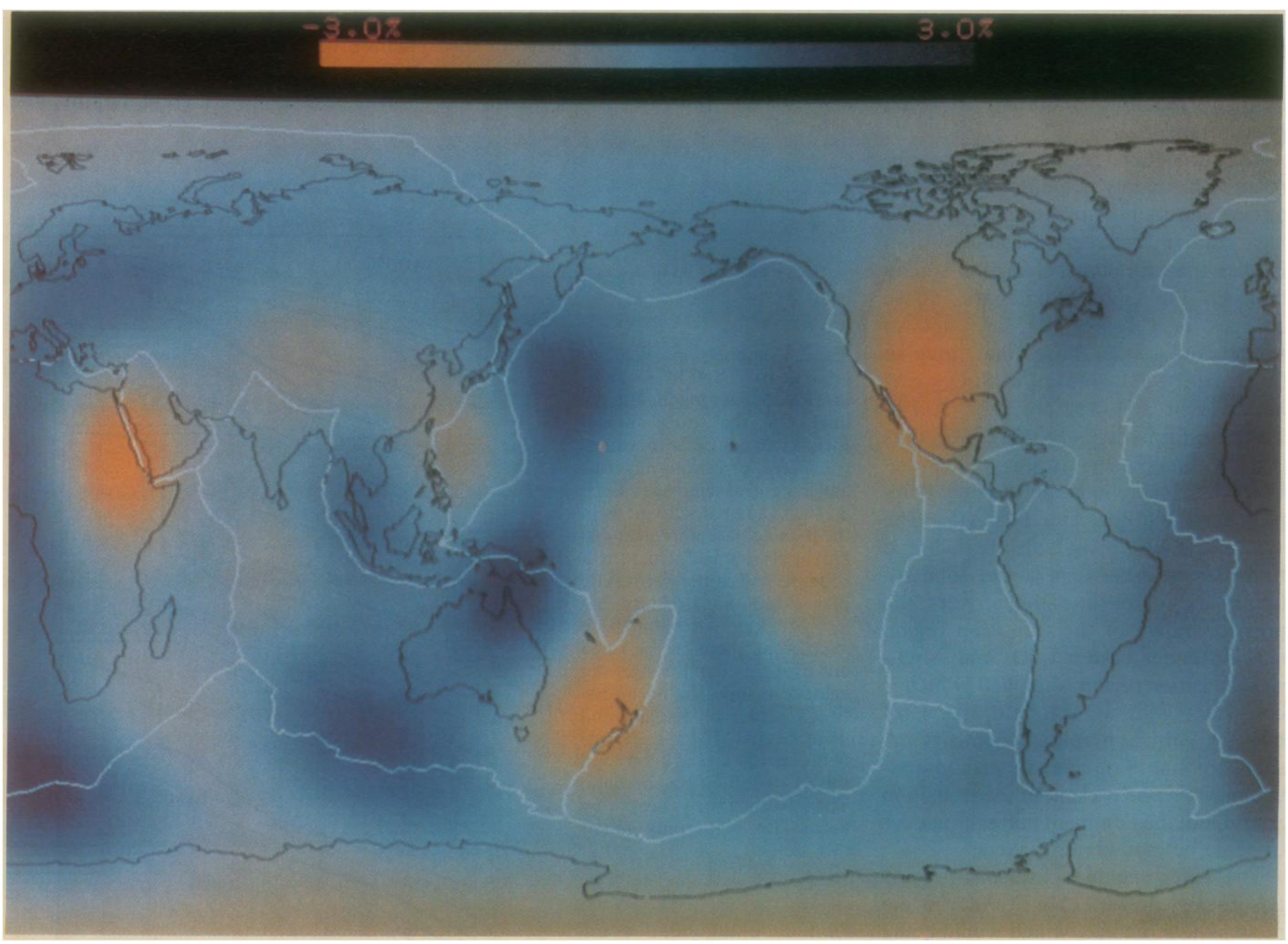

Fig. 3. Map showing the distribution of Rayleigh wave phase velocities at a period of 250 seconds. Odd and even order spherical harmonics up to degree and order 8 are used in the synthesis [modified from Nakanishi and Anderson (*25*)] and Tanimoto (*39*). Orange represents slow areas, and blue fast ones.

buoyancy-induced instability allows partially molten diapirs to ascend adiabatically with increased melting to the shallow mantle, where melt separation forms new oceanic crust. The temperature drop due to adiabatic decompression and latent heat of melting during the ascent is about 400°, which brings the diapir down to temperatures inferred for magmas in the shallow mantle. The 400° temperature increase from the shallow mantle to the depth of origin of mantle diapirs is about what is expected across a thermal boundary layer of a chemical discontinuity. Because of the rapid increase of melting temperature with depth, this chemical discontinuity must be not much deeper than 250 km; otherwise partial melting would not occur, and the deeper reservoir could not be tapped. Mantle geotherms are shown in Fig. 4 for two distributions of radioactive elements.

The oceanic lithosphere, formed by melts from the depleted reservoir, loses its buoyancy as it cools and eventually returns to the mantle, aided by the basalt-eclogite phase change.

The need for two isotopically distinct geochemical reservoirs that separated early in the history of the earth and have remained isolated is unequivocal, but this leads to a variety of paradoxes. Both reservoirs are providing melts to the surface, often at the same place, yet the reservoirs are isolated. If the lower mantle is one reservoir, there should be a large increase in temperature across the 650-km discontinuity, resulting in decreased seismic velocity gradients, high attenuation, and a large drop in viscosity, all of which are contrary to geophysical observations. Because of the effect of pressure on melting point, it is unlikely that melting is possible in the lower mantle, even when the effects of thermal boundary layers are taken into account. However, if most of the radioactivity in the mantle is concentrated in two upper-mantle reservoirs, the temperature increase at 650 km will be small, since little heat is escaping from the lower mantle and the temperature rise across a shallow chemical interface can bring temperatures to the melting point (Fig. 4).

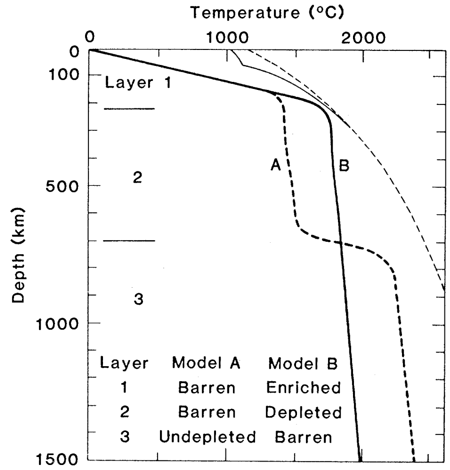

Fig. 4. Schematic diagram of mantle geotherms for two types of layered models. Model A has a very depleted, or barren, upper mantle and a primitive or undepleted lower mantle (high uranium, thorium, and potassium contents). This model has a thermal boundary between the upper and lower mantles and a hot mantle which, nevertheless, is far below the inferred melting point (light solid and dashed curves). Model B has the same amount of radioactivity, but it is distributed between two chemically distinct upper mantle reservoirs. This model approaches the melting point at the top of the depleted layer (layer 2) and has lower temperatures and higher viscosities in the lower mantle. A perturbation of geotherm B by, for example, continental insulation or a convectively induced uplift of the chemical boundary, will cause melting in layer 2. Geotherm A is modified from (*40*).

A Geoid with Memory

The geoid does not reflect the presence of continents or ocean ridges. The major geoid highs are associated with subduction and convergence regions such as Tonga-Fiji, New Guinea, Borneo, the Philippines, and Chile. The most pronounced geoid lows occur in a discontinuous polar band extending across the Canadian shield, the western Atlantic Ocean, Brazil, Antarctica, and the northern Indian Ocean to Siberia. They occur in various kinds of tectonic provinces. The third type of global geoid anomaly, the equatorial and antipodal highs associated with the two largest plates, the African and the Pacific plates, include most of the world's hot-spot provinces. One can speculate that continents which are embedded in geoid lows or which have overridden large amounts of old oceanic lithosphere (much of North America, South America, Australia, India, and Antarctica) may have completed their journey. On the other hand, western North America, eastern Africa, and Southwest Asia are underlain by substantial seismic velocity and thermal anomalies and sit astride pronounced gradients in the geoid. These regions may fragment and lose slivers to nearby geoid lows.

The rotation axis has apparently wandered about 8° in the past 60 million years and 20° in the past 200 million years (*28*). During these periods there have been major changes in the configurations of continents and subduction zones. These apparently had minor effects on the principal moments of inertia, suggesting a relatively stable and slowly evolving geoid. Even a slowly changing geoid, however, can cause a rapid shift in the whole mantle relative to the spin axis if the moment of inertia along this axis becomes less than some other axis of inertia. This may have happened sometime between 450 and 200 million years ago when Gondwana experienced a large latitude shift.

On the familiar Mercator projection the continents appear to be more or less randomly disposed over the surface of the globe. On other projections, however, there is a high degree of symmetry (*29*). The centers of the largest (African and Pacific) plates are antipodal, and the other smaller plates, containing most of the large shield areas, occupy a relatively narrow polar band. The geoid lows in Siberia, the Indian Ocean, Antarctica, the Caribbean, and the Canadian shield also fall in this band. Long wavelength geoid highs are centered on the large antipodal plates with hot spots, and it appears that shield areas have emigrated from these thermal and geoid highs (*30*). The polar band of discontinuous geoid lows correlates with areas of Cretaceous subduction (*31*) and subsequent subsidence and areas in the upper mantle where seismic velocities are fast. Isostatically compensated cold mantle generates geoid lows. The geoid lows found in the wake of the Americas, India, and Australia exhibit fast surface wave velocities, indicating that the anomalies are in the upper mantle.

Density inhomogeneities in the mantle grow and subside, depending on the locations of continents and subduction zones. The resulting geoid highs reorient the mantle relative to the spin axis. Whenever there was a major continental assemblage in the polar region surrounded by subduction as was the case during the Devonian through the Carboniferous, the stage was set for a major episode of true polar wandering.

The outer layers of the mantle, including the brittle lithosphere, do not fit properly on a reoriented earth (*32*). Membrane stresses generated as the rotational bulge shifts may be responsible for the breakup and dispersal of Pangaea as it moves toward the equator. In this scenario, true polar wandering and continental drift are intimately related. A long period of continental stability allows thermal and geoid anomalies to develop. A shift of the axis of rotation causes plates to split, and the horizontal temperature gradient causes continents to drift away from the thermal anomalies that they caused. The continents drift toward cold parts of the mantle and, in fact, make the mantle cold as they override oceanic lithosphere.

Polar wandering can occur on two distinct time scales. In a slowly evolving mantle the rotation axis continuously adjusts to changes in the moments of inertia (*33*). This will continue to be the case as long as the major axis of inertia remains close to the rotation axis. If one of the other axes becomes larger, the rotation vector swings quickly to the new major axis. The generation and decay of thermal perturbations in the mantle are relatively gradual, and continuous small-scale polar wandering can be expected. The interchange of moments of inertia, however, occurs more quickly, and a large-scale 90° shift can occur on a time scale limited only by the relaxation time of the rotational bulge. The rate of polar wandering at present is much greater than the average rate of relative plate motion, and it would have been faster still during an interchange event. The relative stability of the rotation axis for the past 200 million years (*28*) suggests that the geoid highs related to hot spots have existed for at least this long. On the other hand, the rapid polar wandering that started 500 million years ago may indicate that the Atlantic-African geoid high was forming under Gondwana at that time and had become the principal axis of inertia. With this mechanism a polar continental assemblage can be physically rotated to the equator as the earth tumbles.

The southern continents all underwent a large northward displacement beginning sometime in the Permian or Carboniferous (280 million years ago) and continuing to the Triassic (190 million years ago). During this time the southern periphery of Gondwana was a convergence zone (*34*), and a spreading center is inferred along the northern boundary. One would expect that this configuration would be consistent with a stationary or a southern migration of Gondwana, unless a geoid high centered on or near Africa was rotating the whole assemblage toward the equator. The areas of very low surface wave velocities in northeast Africa and the western Indian Ocean may be the former site of Gondwana.

Thus, expanding the paradigm of continental drift and plate tectonics to include continental insulation and true polar wandering may explain the paradoxes of synchronous global tectonic and magmatic activity, rapid breakup and dispersal of continents following long periods of continental stability, periods of static pole positions separated by periods of rapid polar wandering, sudden changes in the paths of the wandering poles, the migration of rifting and subduction, initiation of melting, the symmetry of ridges and fracture zones with respect to the rotation axis, and correlation of tectonic activity and polar wandering with magnetic reversals. Tumbling of the mantle presumably affects convection in the core and orientation of the inner core and offers a link between tectonic and magnetic field variations.

Conclusions

Seismic data show that the upper mantle is extremely heterogeneous, and the velocities have a good correlation with surface tectonics. Continental shields are relatively shallow features, but regions of slow velocity, such as that under the East African rift, the Red Sea, and the Gulf of Aden, the region under western North America and the East Pacific Rise, are deep seated and extend to the transition region. Density inhomogeneities in the upper mantle—the result of the presence or absence of subducted material—may control the orientation of the spin axis. The transition region of the upper mantle is rich in garnet and appears to be the source of the oceanic lithosphere. The regions of the upper mantle that are densest and have the fastest velocities may be those regions where cold oceanic lithosphere has been overridden by drifting continents during the past 100 million years. Such continents might appear to have deep roots.

The most plausible interpretation of the high gradient of seismic velocity in the transition region is that there is a continuous phase change in material rich in CaO-Al_2O_3 such as eclogite. Olivine is a minority phase in the transition region and lower mantle, as suggested by cosmic abundances. The large density changes associated with phase changes and melting in the basalt-eclogite system may drive convection and be responsible for the chemical stratification of the mantle and the long-term isolation of geochemical reservoirs.

A paradigm is the set of rules and assumptions that guide the workers in a given field. As facts accumulate, paradoxes multiply to the point where they topple the paradigm. The standard way to eliminate a paradox, aside from ignoring it, is to change the rules and assumptions, thereby incorporating the paradox into a new paradigm. Thus, today's paradoxes are tomorrow's paradigms.

References and Notes

1. There is abundant evidence that convection in the mantle is not a steady-state phenomenon, but almost all theoretical and laboratory work aimed at understanding mantle convection is based on the assumption that it is. Continental insulation, for example, leads to unsteady convection [J. Elder, *Nature (London)* **214**, 657 (1967)].
2. The lithosphere, which is thicker and colder under shields than elsewhere, represents a chemical thermal boundary layer through which heat is removed by conduction rather than convection. The heat flow through shields is approximately that expected from the radioactivity of the crust and lithosphere, with little or no contribution from the deeper mantle. Continental shields and platforms therefore act to insulate the mantle, whereas subducted slabs cool it.
3. The light rare earth elements are generally enriched in melts and depleted in residual crystals, relative to the source material.
4. P. Warren and J. Wasson, *Rev. Geophys. Space Phys.* **17**, 73 (1979). KREEP is a lunar rock type rich in K, rare earth elements, and P [see various authors in *Basaltic Volcanism Study Project* (Pergamon, New York, 1981)].
5. I. MacGregor, *Mineral. Soc. Am. Spec. Pap. 3* (1970), p. 51; D. L. Anderson, in *Third International Kimberlite Conference*, J. Kornprobst and M. Kornprobst, Eds. (Elsevier, Amsterdam, in press).
6. Lherzolite is a rock of the peridotite family composed mainly of olivine, $(Mg,Fe)_2SiO_4$, and two types of pyroxene. Pyrolite is a hypothetical peridotite with 57 percent olivine [A. Ringwood, *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975)]. Pyrolite is low in CaO and Al_2O_3 in comparison with eclogite.
7. J. Melosh, *Earth Planet. Sci. Lett.* **26**, 353 (1975); *Icarus* **44**, 745 (1980).
8. ———, *Earth Planet. Sci. Lett.* **25**, 322 (1975).
9. S. Crough and D. Jurdy, *ibid.* **48**, 15 (1980); B. Hager and D. Jurdy, personal communication.
10. W. Kaula, *J. Geophys. Res.* **84**, 999 (1979).
11. S. Hart, *Philos. Trans. R. Soc. London* **268**, 573 (1971).
12. The 650-km discontinuity is known to be an efficient seismic reflector [R. Adams, *Bull. Seismol. Soc. Am.* **58**, 1933 (1968); E. R. Engdahl and E. A. Flinn, *Science* **163**, 177 (1969); J. Whitcomb and D. L. Anderson, *J. Geophys. Res.* **75**, 5713 (1970)]. The 650-km discontinuity is also referred to as the 670-km discontinuity.
13. P. Sobel [thesis, University of Minnesota, Minneapolis (1978)] suggested that the 650-km discontinuity may be two seismic reflectors separated by 40 km.
14. R. Jeanloz and A. Thompson, *Rev. Geophys. Space Phys.* **21**, 51 (1983); K. Weng, H. Mao, P. Bell, *Carnegie Inst. Washington Publ.* **81**, 273 (1982).
15. J. B. Minster and D. L. Anderson, *Philos. Trans. R. Soc. London* **299**, 319 (1981).
16. H. Iyer, *Tectonophysics* **56**, 165 (1979).
17. D. L. Anderson and M. N. Toksoz, *J. Geophys. Res.* **68**, 3483 (1963); M. Niazi and D. L. Anderson, *ibid.* **70**, 4633 (1965); L. Johnson, *ibid.* **72**, 6309 (1967).
18. D. L. Anderson, *Science* **157**, 1165 (1967).
19. M. Akaogi and S. Akimoto, *Phys. Earth Planet. Inter.* **19**, 31 (1979).
20. P. Richards, *Z. Geophys.* **38**, 517 (1972).
21. D. L. Anderson, *Geophys. Res. Lett.* **3**, 347 (1976).
22. ———, *Mineral. Soc. Am. Spec. Pap.* **3**, 85 (1970); R. Butler and D. L. Anderson, *Phys. Earth Planet. Inter.* **17**, 147 (1978); J. Watt and T. J. Ahrens, *J. Geophys. Res.* **87**, 5631 (1982).
23. L. Liu, *Earth Planet. Sci. Lett.* **43**, 331 (1979).
24. I. Nakanishi and D. L. Anderson, *Bull. Seismol. Soc. Am.* **72**, 1185 (1982).
25. ———, *J. Geophys. Res.*, in press.
26. A. Souriau and M. Souriau, *Geophys. J. R. Astron. Soc.* **73**, 533 (1983).
27. H. S. Yoder, Jr., *Generation of Basaltic Magma* (National Academy of Sciences, Washington, D.C., 1976).
28. R. Gordon, *Geophys. J. R. Astron. Soc.* **70**, 129 (1982); J. Morgan, *The Sea* (Wiley, New York, 1981), vol. 7; C. G. A. Harrison and T. Lindh, *Nature (London)* **300**, 251 (1982); R. Gordon and C. Cape, *Earth Planet. Sci. Lett.* **55**, 37 (1981).
29. E. Kanasewich, J. Harskov, M. Evans, *Can. J. Earth Sci.* **15**, 919 (1978).
30. D. L. Anderson, *Nature (London)* **297**, 391 (1982).
31. C. Chase and D. Sprowl, *Earth Planet. Sci. Lett.* **62**, 314 (1983).
32. F. Vening-Meinesz, *Trans. Am. Geophys. Union* **28**, 1 (1947).
33. P. Goldreich and A. Toomre, *J. Geophys. Res.*

74, 2555 (1969); D. Fisher, *ibid.* **79**, 4041 (1974).
34. K. G. Cox, *Nature (London)* **274**, 47 (1978).
35. S. Grand and D. Helmberger, *J. Geophys. Res.*, in press; M. Walck, *ibid.*, in press; J. Given and D. Helmberger, *ibid.* **85**, 7183 (1980).
36. A. Dziewonski and D. L. Anderson, *Phys. Earth Planet. Inter.* **25**, 297 (1981). Model PREM′ is from Regan and Anderson (in preparation).
37. A. Hales and K. Muirhead, J. Rynn, *Tectonophysics* **63**, 309 (1980).
38. Y. Fukao, *Geophys. J. R. Astron. Soc.* **50**, 621 (1977).
39. T. Tanimoto, unpublished data.
40. F. Richter and D. McKenzie, *J. Geophys. Res.* **86**, 6133 (1981).
41. I thank all participants at the 1982 workshop on the lithosphere sponsored by the National Research Council and Geodynamics Committee held in Austin, Texas, in March 1982. I thank particularly J. Maxwell and C. Drake, but they do not necessarily endorse the views in this article. I thank S. Grand, M. Walck, D. Helmberger, B. Hager, T. Tanimoto, J. Bass, and I. Nakanishi for permission to use results in advance of publication. Figure 3 was prepared by R. Clayton and B. Hager from observational results of I. Nakanishi and T. Tanimoto. Supported by NSF grant EAR811-5236 and NASA Geodynamics grant NSG-7610. Contribution No. 3921, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena 91125.

Localization, Interactions, and the Metal-Insulator Transition

R. C. Dynes and P. A. Lee

Any material can be characterized as a metal or an insulator according to the following definition: upon extrapolation to absolute zero temperature, if the conductivity remains finite, it is a metal, and if the conductivity goes to zero, it is an insulator. The zero-temperature conductivity of a metal can be understood in terms of the scattering of the conduction electrons by impurities. Furthermore, as the temperature is raised, the traditional picture in terms of Boltzmann transport theory states that thermal excitation of various inelastic processes enhances the scattering rate, so that the conductivity decreases. Phase space arguments show that the conductivity can be described by

$$\sigma = \sigma_0 - AT^n \quad (1)$$

where T is temperature and the power $n = 2$ if the scattering is due to electron-electron collision and is generally larger than 2 if other scattering processes such as electron-phonon scattering dominate.

Recent theoretical developments, together with experiments on a variety of systems, have shown that almost all aspects of Eq. 1 are wrong as far as the asymptotic low-temperature behavior is concerned. The power n is found to be much less than 2 (instead it is typically 1/2) and the sign of A is often negative, so that the conductivity of a metal increases with increasing temperature. The new understanding of the metal also has important implications for the understanding of the insulator-to-metal transition, which is the process by which an insulator is transformed into a metal by changing some material parameter such as the concentration of dopants in a semiconductor. The recent advances are based on improved understanding of two aspects of the problem: Anderson localization and the effects of electron-electron interactions in a disordered medium.

Summary. Recent advances in our understanding of electronic conduction have pointed up deficiencies in traditional thinking. For a metal at a sufficiently low temperature, it is known both theoretically and experimentally that the conventional picture in terms of the Boltzmann theory breaks down. Improved understanding of both electron localization and the effects of electron-electron interactions in a disordered medium has led to experimentally verifiable predictions. These effects have an important influence on the nature of the metal-insulator transition.

Anderson Localization

The concept of Anderson (*1*) localization deals with the nature of a one-electron wave function in a disordered medium. If the disorder is weak the wave function is extended; that is, it is like a plane wave except that its phase becomes randomized on a length scale defined as the mean free path l. In 1958, Anderson showed that if the disorder is strong the wave function may change its nature completely and become localized; that is, the wave function envelope decays exponentially from a center. The decay length ζ is the localization length and may become much longer than the mean free path. When the wave function at the Fermi energy becomes localized, we have an insulator. If the disorder is gradually reduced, the localization length increases until at some point the wave function becomes extended and an insulator-to-metal transition occurs.

Mott (*2*) proposed that at this transition, the conductivity jumps to a finite value σ_{min}. His reasoning was based on an extrapolation to the strong disorder region of the usual Boltzmann formula for the conductivity of a metal

$$\sigma = \frac{e^2}{3\pi^2\hbar} k_F^2 l \quad (2)$$

where e is the electron charge, $\hbar$ is Planck's constant divided by 2π, and k_F^{-1} is the de Broglie wavelength of the electron. It is reasonable to suppose that localization sets in when the mean free path l becomes of order of k_F^{-1}, because if l becomes any shorter, the phase is so random that a plane wave description no longer makes sense. Putting the so-called Ioffe-Regel criterion (*3*) ($k_F l \approx 1$) into Eq. 2, we obtain Mott's estimate of σ_{min} (up to a numerical factor)

$$\sigma_{min} \approx \frac{e^2}{3\pi^2\hbar} k_F \quad (3)$$

It is interesting to observe that $e^2/\hbar = (2.44 \times 10^{-4})\ (\text{ohm})^{-1}$ has the dimensions of a conductance, so that Eq. 3 can be interpreted as the condition that a microscopic sample of size k_F^{-3} has a conductance of $e^2/\hbar$. This point of view is even more transparent in two dimensions (2D), where conductivity has the same dimension as conductance and Eq. 3 becomes

$$\sigma_{min}^{(2D)} \approx \frac{e^2}{2\pi\hbar} \quad (4)$$

R. C. Dynes is director of the Chemical Physics Research Laboratory at Bell Laboratories, Murray Hill, New Jersey 07974. At the time this work was done P. A. Lee was a member of the technical staff at Bell Laboratories. His present address is Department of Physics, Massachusetts Institute of Technology, Cambridge 02139.