IMPACTS AND THE EARLY ENVIRONMENT AND EVOLUTION OF THE TERRESTRIAL PLANETS

H. J. MELOSH, A. M. VICKERY, and W. B. TONKS
University of Arizona

Early models for the accretion of the planets generally assumed that the planets grew gently by the steady addition of small grains swept up by the growing planetary embryos from the solar nebula. More recent models of the formation of the solar system, however, show that the population of impactors bombarding the growing planets increased in size and relative velocity with time, so that accretionary impacts tended to become progressively more violent until the last of the giant planetesimals were destroyed in collisions with planets. Accretion then tapered off during the period of heavy bombardment and continues at a much diminished rate today. The recognition of the importance of large individual impacts as well as of the cumulative effects of smaller impacts has led to a revolution in our approach to the modeling of the processes of planetary formation and differentiation, including the origin of the Moon and the origin and evolution of atmospheres. Once the planetary embryos reached a critical size, impact velocities became high enough to liberate volatiles due to shock heating, and proto-atmospheres could begin to form. Larger impacts resulted in planetary heating as the time between impacts became smaller than the time for cooling by radiation, an effect that would have been enhanced by the presence of an atmosphere. Very large impacts may have melted significant fractions of the planets, leading to rapid core formation. If the Moon was created in a giant impact, as is now widely believed, it is likely that melting would be widespread, at least in the hemisphere of the Earth in which the impact occurred and possibly throughout the entire planet. A magma ocean would form on the Earth, but because of the high gravity, convection would be sufficiently vigorous to prevent significant fractional crystallization, as is inferred to have occurred on the Moon. The effects of such an impact on a pre-existing atmosphere, and its implications for atmospheric evolution, have not yet been explored. Impacts during heavy bombardment probably resulted in some net atmospheric loss from the Earth and Venus, but the effect would have been much more pronounced on Mars, consistent with its very thin atmosphere. Such impacts may fractionate a planet's volatile inventory by sweeping away atmospheric gases but leaving condensible phases (such as water) little changed.

I. INTRODUCTION

Over the last two decades knowledge about the physics of high-speed impact processes has grown enormously. At the same time the role of impact and related processes in the formation and early evolution of the terrestrial planets has been increasingly well-defined. The most recent developments in this field are the recognition of the importance of "giant impacts"--collisions between

embryo planets and bodies of about 10% of their mass—in the last stages of accretion, and the realization that impact processes may have played a major role in the growth and retention of planetary atmospheres. In this chapter we review the nature and implications of these new ideas as they currently stand, although the reader should be warned that these are both areas of very active research at the present time, so that new developments may make this review obsolete in a short time.

The recognition of the importance of giant impacts grew mainly out of recent research on the Moon's origin. The idea that the Moon was born in a gigantic collision between the proto-Earth and a Mars-size protoplanet has continued to gain adherents since it was first proposed by Hartmann and Davis (1975) and by Cameron and Ward (1976). This theory received a great deal of attention at the 1984 Conference on the Origin of the Moon in Kona, Hawaii (Hartmann et al. 1986), and has subsequently been the subject of a number of reviews (Boss 1986*c*; Stevenson 1987; Newsom and Taylor 1989). It is probably not an exaggeration to claim that it has become the current consensus theory of the Moon's origin. Although a great deal of work has been, and is being, done on the details of this scenario, most work up until recently concentrated on the Moon, with little consideration given to the more general effects of such large impacts on the Earth or other planets. It is clear that such an event would have profound effects on a growing planet: by the time a planet has reached the size of Mars, impacts with objects half its diameter (i.e., about 10% of its mass) are capable of melting either the entire planet or at least one hemisphere, as the shock wave from the impact traverses the planet. In this scenario, core formation is quick and inevitable; metallic iron particles separate rapidly from the silicate melt, forming a pool of iron at the bottom of the molten volume. This pool may be large enough to displace even strong underlying silicate rocks and form a proto-core (Tonks and Melosh 1991). Subsequent large impact and melting events will add more material to the core in a batchwise fashion. Vigorous convection in the melted regions (magma oceans) causes the global magma sea to cool rapidly (cooling to 50% crystallization requires only a few thousands to tens of thousands of years) and prevents crystal-liquid density segregation when the Rayleigh number is high enough to ensure turbulent convection.

At a later stage of planetary evolution, the smaller impacts during late heavy bombardment may have played an important role in bringing volatiles to the planet, in stripping away the original gaseous atmospheres of the planets and in segregating condensible substances, such as water, from volatile ones, such as CO_2. A large meteorite or comet that strikes the surface of a planet possessing an atmosphere can interact with the atmosphere in three ways: during its initial passage through it, by means of high-speed ejecta thrown out from the growing crater, and by the expanding high-energy ejecta plume that is produced at sufficiently high impact velocities. The leading edge of this plume may have a velocity many times greater than the impact velocity. In sufficiently large impacts on Earth (meteorites >300 m in diameter), the

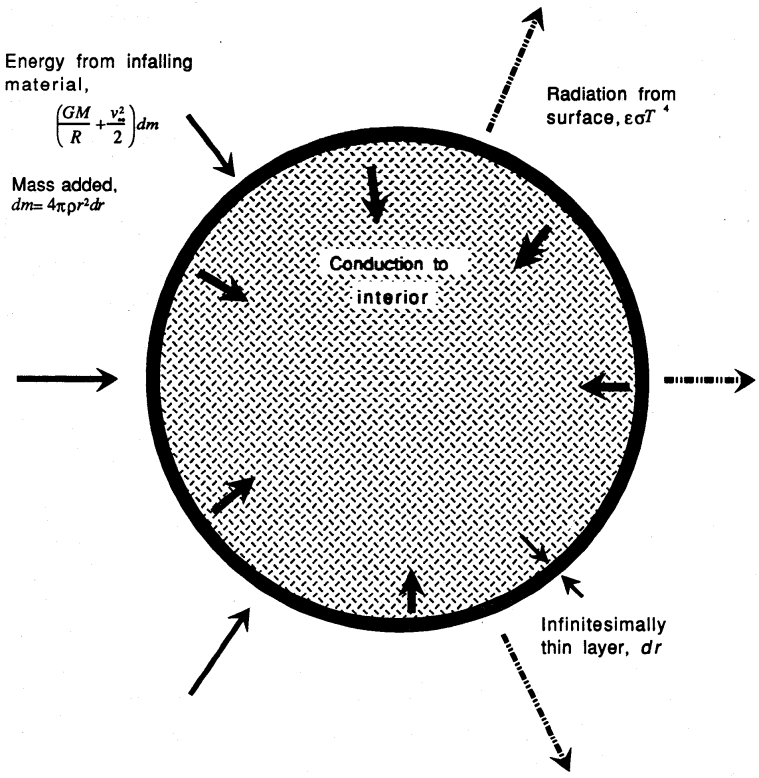

Figure 1. The "physicist's model" of planetary accretion. This model, which was current between about 1950 and 1975, assumes that the gravitational energy of infalling material is balanced primarily by thermal radiation from the surface with a small additional contribution from conduction. This model is valid only if the planet grows from planetesimals smaller than about 10 km in diameter.

plume may have enough momentum to blow aside the atmosphere and vent vaporized rock and other debris directly into space. In still larger impacts (>10 km diameter), the overlying atmopheric gases may be accelerated to greater than escape velocity so that atmospheric erosion occurs. Mars may have lost most of its primordial atmosphere in this way during the era of heavy bombardment.

II. ENERGY DEPOSITION BY IMPACTS ON GROWING PLANETS

A. Accretion of Infinitesimally Small Impactors

The history of accretion models of the Earth and planets mirrors a gradually increasing appreciation for the importance of large impacts. Early models (Hanks and Anderson 1969; Mizutani et al. 1972; Urey 1952) of the Earth's accretion assumed that material is added by the deposition of infinitesimally thin, laterally homogeneous layers. The accreted material adds a net energy per unit mass consisting of gravitational binding energy and initial kinetic

energy of the mass, $GM/R + v_\infty^2/2$, where G is Newton's gravitational constant, M is the planet's mass when it has grown to radius R, and v_∞ is the velocity of encounter between the growing planet and the infalling material at great distance from the planet. Some of this energy is lost by radiation, and the rest goes into heating the planet—initially into heating the surface layer, then gradually into heating the interior of the planet by conduction. The basic equation describing the equilibrium between infall energy, radiation and conduction is

$$\rho\left(\frac{GM(t)}{R(t)}+\frac{v_\infty^2}{2}\right)\frac{dR(t)}{dt}=\epsilon\sigma[T^4(R,t)-T_a^4]$$

$$+\rho c_P T(R,t)\frac{dR(t)}{dt}+k\left(\frac{\partial T}{\partial r}\right)_{r=R} \quad (1)$$

where $T(R, t)$ is the temperature at the surface of the growing planet, T_a is the effective temperature of the atmosphere, σ is the Stefan-Boltzman constant, ϵ is the emissivity (generally set equal to 1), c_P is heat capacity and k is the thermal conductivity. Radius r is the position *within* the growing planet of total radius $R(t)$. The second term on the right side accounts for the heat content of the hot added material and should include latent heat if melting occurs. The thermal state of a growing planet is determined by Eq. (1) as soon as the growth rate, $dR(t)/dt$, is specified.

The studies based on Eq. (1) had great difficulty explaining why any of the planets should be differentiated. Thermal radiation is so efficient at removing energy that the planets would have had to grow extremely rapidly for temperatures to reach the melting point anywhere in their interiors. Thus, Hanks and Anderson (1969) require the Earth to grow in 10^5 to 10^6 yr for melting to occur, while Mizutani et al. (1972) find the Moon must accrete within 1,000 yr if melting is to occur in its outer portions. These times are far shorter than the 10^7 to 10^8 yr time scale derived from standard planetesimal accretion models (Wetherill 1980*a*). This implies that the Earth and Moon accreted cold, and the Earth later differentiated when enough radiogenic heat had accumulated to cause internal melting. This model also implies that the Moon should not have differentiated, a prediction that was quickly proved wrong when the Apollo missions returned Moon rocks to Earth for analysis.

B. Accretion of Small Planetesimals

A solution to this quandary was first suggested by Safronov (1972,1978) and more recently was elaborated by Kaula (1979*b*,1980). They realized that the idealization of the gradual addition of infinitesimally thin global layers is not a valid representation of the actual process of accretion, which involves the impact of individual planetesimals onto the planet's surface. Each impact heats the target rocks directly beneath the impact site and causes the deposition of an ejecta blanket of finite thickness and localized extent, as shown in Fig. 2. Under these conditions radiation is able to cool only a thin layer on the top of

the ejecta blanket before yet another ejecta sheet is deposited on top of it. A large fraction of the initial energy of the infalling material is thus retained in the overlapping ejecta blankets.

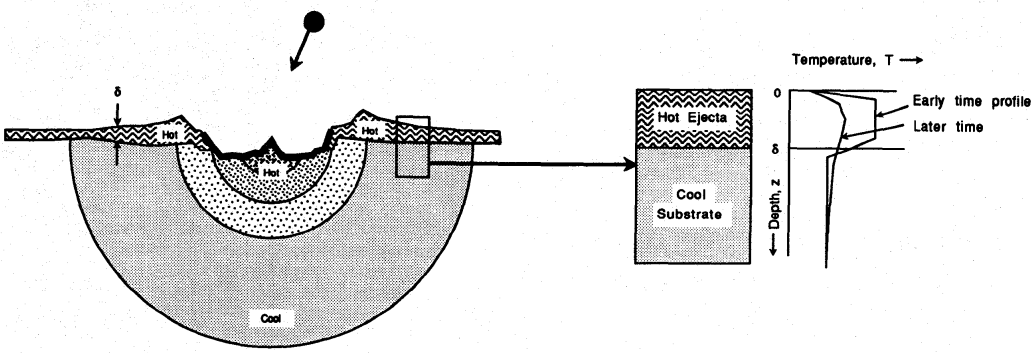

Figure 2. Heat deposition in the vicinity of a large impact. Shock waves deposit heat directly in the rocks beneath the crater, the amount of heating decreasing with increasing distance from the impact site, illustrated by the schematic temperature contours. The heavy black line inside the crater indicates a thick layer of melted target rocks. Such layers are observed inside large fresh lunar craters such as Copernicus and Tycho. The ejecta also contain a considerable amount of heat and thickly blanket the terrain to a distance of about 1 crater diameter from the rim. The temperature in the ejecta blanket (inset) declines as a result of thermal radiation from the surface and thermal conduction into the cooler substrate.

It is easy to estimate the thickness δ of an ejecta sheet for which heat retention outweighs radiative loss. If the planet grows at a rate dR/dt, the average time interval Δt between deposition of ejecta sheets of thickness δ at any given site is simply $\delta/(dR/dt)$. But the conductive cooling time of an ejecta sheet is of order δ^2/k, where k is the ejecta's thermal diffusivity. This cooling time equals the average time between deposition events for $\delta_0 = k/(dR/dt)$. Layers thinner than δ_0 radiate all of their heat before the next ejecta deposition event, whereas thicker layers do not have time to cool between events. Making the conservative assumption that the Earth grew over a period of about 100 Myr and that $\kappa \approx 10^{-6}$ m^2 s^{-1}, yields a crossover thickness δ_0 of about 3 km. The maximum ejecta thickness of a fresh crater is about 0.02 of the crater diameter, which itself is about 10 times larger than the projectile diameter (Melosh 1989), so the crossover condition is met by planetesimals about 15 km in diameter. Moreover, much of a projectile's energy is deposited more deeply beneath the crater floor than in the ejecta blanket, so that it is plausible that most of the heat added by 10 km or larger diameter planetesimals is retained by the growing planet rather than lost by radiation.

The details of the overall heat deposition process are complex and depend sensitively on impact mechanics. Thus, of the projectile's initial kinetic energy, about 30% is initially partitioned into the kinetic energy of the ejecta

and most of the remainder is directly deposited as heat in the target rocks (Melosh 1989). The kinetic energy of the ejecta is converted into heat when it comes to rest on the surface after mixing with a variable amount of preexisting surface material, resulting in a mixed sheet of hot rock debris thickly covering the surface within one or two crater diameters of the crater's original rim. The rocks beneath the crater (and the expelled ejecta) are directly heated by the shock from the impact which, depending upon the impact velocity, may melt or vaporize large quantities of material. The amount of material thus heated depends mainly upon the projectile's size and the square of its velocity (Melosh 1989). On the other hand, a large impact cools the target planet by raising deeply buried materials closer to the surface, where their heat may be more readily conducted to the surface and ultimately lost by radiation. Some of the initial heat of the ejecta may be lost by radiation during its ballistic flight from the impact crater to its site of deposition, while at high impact velocities a portion of the ejecta (especially that in the vapor plume) may travel at velocities greater than escape velocity and thus leave the planet entirely. The net thermal effect of an impact is thus a sum of gains and losses which tend to offset one another and hence make accurate estimation of the net heat deposition difficult. Crude consideration of these processes indicates that large impacts are more efficient at depositing energy than small impacts, and that for kilometer-size planetesimals the radiation term in Eq. (1) is almost entirely negligible.

Kaula's (1980) solution to the uncertainty in estimating the net heat deposition of impacts is to lump all the poorly known processes into a single numerical factor h and write Eq. (1) in a form that neglects both radiation and conduction. Thus, rearranging Eq. (1),

$$T(R,t) = \frac{h}{c_P}\left(\frac{GM(t)}{R(t)} + \frac{v_\infty^2}{2}\right). \tag{2}$$

Note that with the neglect of radiation and conductive heat loss the accretion rate $dR(t)/dt$ drops out, so that the interiors of planets growing over the 10^8 yr accretion time scale may easily reach temperatures high enough to initiate melting and differentiation. The unknown dimensionless factor h must lie somewhere between 0 (no net burial of heat) and 1 (all kinetic energy of infalling matter is retained as heat). In the absence of better information h is generally assumed to be about 0.5 for kilometer-size or larger planetesimals.

In Safronov's (1972) theory, the random velocity component v_∞ of approaching planetesimals at any time t is proportional to the escape velocity $v_{esc} = (2GM/R)^{1/2}$ of the planet growing in their neighborhood, so the entire right-hand side of Eq. (2) is proportional to GM/R times factors of the order of unity. Because M is proportional to $R^3(t)$, the surface temperature T of a growing planet at time t is proportional to $R^2(t)$. This temperature is "locked in" at radius $r = R(t)$ as more material accumulates on the former surface of the planet, establishing an internal temperature distribution of the form

$T(r) \sim r^2$. This relation holds until T approaches the melting temperature in the outer portion of the planet, at which time convection begins and the mantle temperature remains near the melting point (Kaula 1980). Melting in the Earth begins when it has reached about 10% of its final mass, or about half its final diameter. A core is presumed to form at about this time, further heating the mantle and stirring it so that the original $T \sim r^2$ thermal structure is wiped out in a manner similar to that described by Stevenson (1981).

Subsequent to core formation the mantle temperature must have been closer to the solidus than the liquidus, because convection in a completely liquid mantle is so vigorous that its cooling time is only about 10^4 yr (Tonks and Melosh 1990), far shorter than the time scale for Earth's growth from planetesimals. Once more than about 50% of the mantle material crystallizes, however, convection is regulated by the high viscosity of solid-state creep in the crystalline fraction, thus greatly lowering the cooling rates. Craters formed by impacts subsequent to core formation would therefore have excavated either the conductive boundary layer or, if sufficiently large, the semimolten convecting mantle. The extent to which the deposition of impact energy is altered in a planet with a hot mantle has not yet been thoroughly investigated (although Minear [1980] studied the disrupting effect of impacts on the boundary layer of a cooling lunar magma ocean, concluding that impacts decrease the cooling time), nor has the related problem of convection in the presence of rapid accumulation of thick ejecta sheets been studied.

Although Eq. (2) is more realistic than Eq. (1), it is nevertheless limited by the implicit assumption that the impacting planetesimals are small in comparison to the growing planetary embryo. Impact energy is still assumed to be added in thin (although not infinitesimally thin) shells that are, on average, uniformly distributed over the planet's surface (Fig. 3). Thermal radiation losses can be neglected because most of the infalling planetesimals' energy is buried below the surface, but the overall pattern of energy deposition is much the same.

C. Accretion of Large Planetesimals

In recent years, however, it has seemed increasingly likely that the distribution of planetesimal sizes follows a rough power law extending from the smallest sizes up to objects half the size of the largest planetary embryo (Hartmann and Davis 1975; Wetherill 1985), at least during the later stages of accretion. This power distribution is expected to be of the form $N_{\text{cum}}(D) = CD^{-b}$, where N_{cum} is the number of planetesimals with diameters greater than or equal to D, C is a constant, and the power b is frequently observed to be close to 2 in numerical simulations of accretion processes (Greenberg et al. 1978), impact fragmentation experiments (Fujiwara 1986) and in the size distribution of comet nuclei (Delsemme 1987). The $b = 2$ distribution has the special property that the total surface area of planetesimals in successive logarithmically increasing size intervals is constant. That is, the surface area of planetesimals with diameters between, say, D_1 and $2D_1$ is the same as

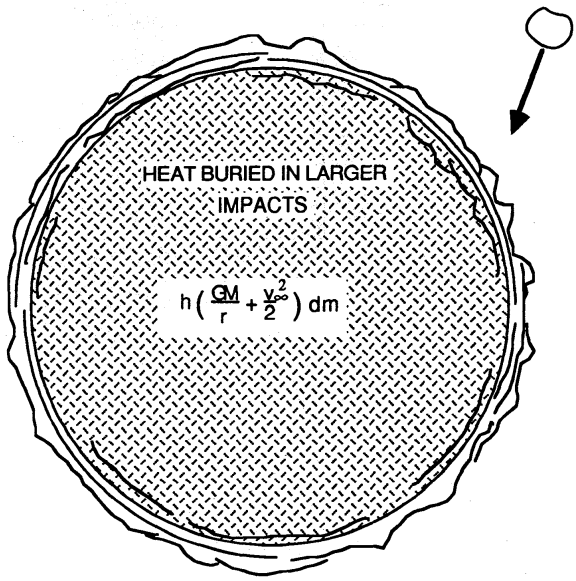

Figure 3. A more realistic model of planetary accretion. In this model, a large fraction of the planetesimal's energy is trapped in the growing planet as heat. Melting temperatures are reached by the time the Earth has grown to about 10% of its present mass. In this model the planetesimals are assumed to be much smaller than the diameter of the planet so that growth is gradual; no provision is made for the effects of very large impacts that deposit their heat catastrophically through a substantial depth of the planetary embryo's mantle.

the surface area of planetesimals with diameters between $2D_1$ and $4D_1$ (see Fig. 4). It seems natural that a distribution of this kind should arise from processes such as impact or coagulation that depend upon cross-sectional area (Chapman and Morrison 1989).

While the largest numbers of planetesimals in a $b = 2$ distribution are concentrated at the small sizes, most of the mass (and therefore energy) resides in the largest objects (see Fig. 5). Thus, although a growing planetary embryo would be constantly battered by small planetesimals, most of the overall mass and heat transfer would occur in large, rare events that bury their heat deep within the embryo's interior. If this catastrophic mode of growth was indeed important, then the assumption that the infalling planetesimals were much smaller than the growing planet fails for the most significant events, and a revised thermal analysis must take account of impacts between objects of comparable size. In the following section, we examine the effects of a single such giant impact between the proto-Earth and an object half its size.

III. EFFECT OF A GIANT IMPACT ON EARTH'S THERMAL STATE

The collision between a Mars-size protoplanet and the proto-Earth adds a truly prodigious amount of energy to the Earth over a time interval measured

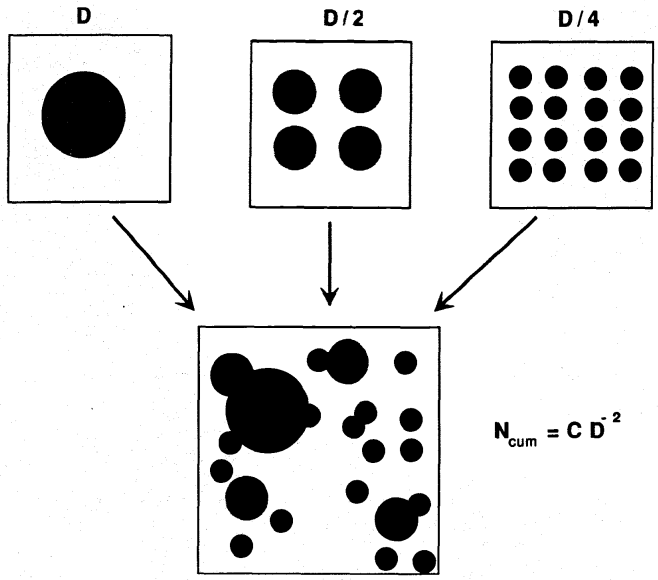

Figure 4. Schematic illustration of the distribution of planetesimal sizes in a population described by a $b = 2$ cumulative number distribution. In this distribution, the projected area of planetesimals in logarithmically decreasing size intervals is constant.

in hours. The mass m, velocity v, and impact parameter b of the projectile are constrained only by the total angular momentum L of the Earth-Moon system, 3.49×10^{34} kg m^2 s^{-1}. Because the angular momentum is a product of all three terms, $L = mvb$, the value of each individual quantity is uncertain within broad limits. The impact parameter is bounded between zero and the sum of the Earth's and the projectile's radii. The impact velocity must be at least as large as the proto-Earth's escape velocity. The overall geochemical similarity of the Earth and Moon implies that the projectile had an initial orbit close to that of Earth's, which in turn implies that the impact velocity was probably not as much as twice the escape velocity. These constraints point to an impact by a body with a mass about 10% of the Earth's mass, hence a diameter approximately half of the Earth's—about the size of Mars.

The energy released in such a collision is the sum of gravitational and kinetic energies, and was probably within the range 2×10^{31} to 5×10^{31} J. Averaged over the Earth's mass, this is about 7.5×10^6 J kg^{-1}. Several authors have used this energy in conjunction with a single value of silicate heat capacity to infer rather high average temperatures for the Earth. Thus, a silicate heat capacity of 10^3 J kg^{-1} K yields an average temperature of 7500 K for the Earth. However, such estimates neglect the relatively small latent heat of melting, $\sim 4 \times 10^5$ J kg^{-1}, the large latent heat of vaporization, $\sim 5 \times 10^6$ J kg^{-1}, and the rapid increase of heat capacity with temperature, ranging from zero at 0 K and reaching a constant maximum at the Debye temperature (about 676 K for dunite, the main constituent of the Earth's mantle). According to the ANEOS equation of state for dunite (Benz et al. 1989), an internal energy

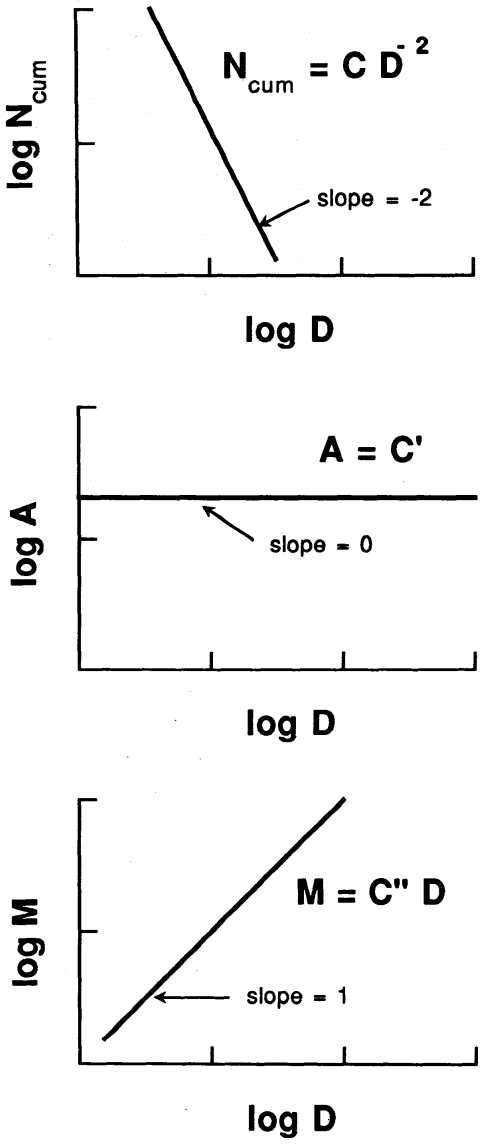

Figure 5. This figure shows dependence of various quantities on planetesimal diameter D in a population where the cumulative number N_{cum} is a function of D^{-2}. In such a population the number distribution is dominated by the smallest objects, the area (either total surface or cross sectional) is uniformly distributed, and the mass distribution is dominated by the largest objects.

of 7.5×10^6 J kg^{-1} corresponds to a temperature rise of about 3200 K for dunite near the Earth's surface and about 7000 K deep in the mantle, where vaporization does not limit the temperature rise. Nevertheless, it is clear that melting, even vaporization, should be widespread in a giant collision.

Impacts at velocities on the order of 10 km s^{-1} in dunite generate pressures on the order of 300 GPa and will thus melt and partially vaporize a few projectile masses of the target, but such melting is confined to regions within about 2 projectile radii of the impact site. More distant regions of the planet will be warmed by impact heating, but will not necessarily melt unless they

are close to melting already (see Fig. 6). Large impacts thus deposit their energy deeper than small impacts, and impacts with objects half the size of the proto-Earth can be expected to melt at least one hemisphere of the Earth's mantle right down to the core. The melt pool following a large impact will undergo further change in shape after it forms, because the melt will have a different (generally smaller) density than surrounding rocks at the same depth. Because the surrounding rocks were likely to be hot, and therefore relatively fluid, subsolidus viscous deformation subsequent to the melting event should close the initial melt-solid crater, producing a global magma ocean of nearly uniform depth overlying hot, more dense, solid mantle material. Although the time scale of this relaxation is difficult to estimate, is was probably not longer than the time scale of post-glacial relaxation in the present Earth; i.e., a few thousand years.

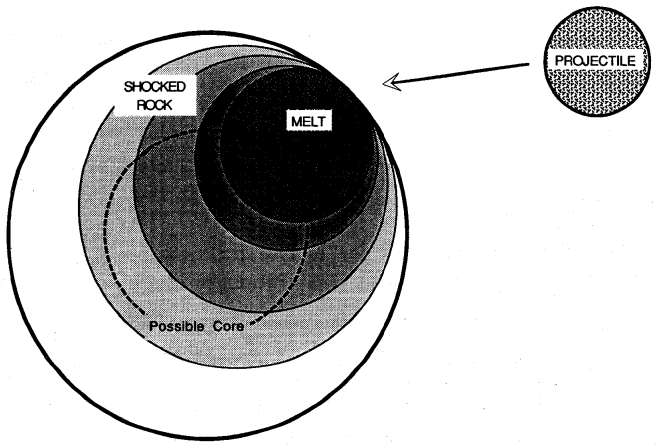

Figure 6. Schematic illustration of the pattern of heat deposition in a planet struck by a projectile of comparable size (1/4 its diameter in this figure). Adjacent to a melted region roughly twice the projectile's diameter, the shock level, and thus temperature, falls off steeply with increasing distance from the impact site.

Although the total energy available from the collision of a Mars-size projectile with the proto-Earth is impressive, the distribution of the energy within the Earth is equally important. If, as has been suggested by Stevenson (1987), this energy is mainly expended in vaporizing the projectile, the Earth may acquire a transient silicate vapor atmosphere without strongly heating the deeper mantle. This is much more likely to occur, however, if the impactor were much less dense than the target, i.e., a comet. The similarity between the compositions of the Earth and Moon, however, strongly suggest that the impactor was a silicate body instead of a comet. For this case, the simple considerations of shock wave geometry discussed above indicate that only partial vaporization occurs (although jetting [Melosh and Sonett 1986] may enhance the local production of hot vapor), and that deep melting should be

widespread in at least the hemisphere that the impact occurs.

An additional factor not previously considered here (but see also Benz and Cameron [1990]) is that the projectile might have an iron core which will sink through the Earth's mantle shortly after the collision and merge with the Earth's core, releasing its gravitational potential energy. Assuming a core equal to 30% of the projectile's mass, the energy released by sinking 3000 km through Earth's mantle is of order 3×10^{30} J, which itself will cause strong heating of the mantle through which it sinks and of the Earth's core when it arrives. After the cores have merged, this heat is applied to the bottom of the mantle, so that any portion of the mantle that escaped melting by the direct shock wave will likely be melted by this means.

To address the temperature rise in the Earth during a Moon-forming collision more exactly, H. J. Melosh performed a series of 3-dimensional numerical hydrocode computations in conjunction with M. E. Kipp of Sandia National Laboratory. These computations were designed to simulate the impact between the proto-Earth and a Mars-size protoplanet. They used the code CTH, implemented on the Cray X/MP supercomputers of the Sandia National Laboratory. This computation uses the ANEOS equations of state for dunite in the mantles and iron in the cores of the two colliding planets. The Earth has a central gravitational field, and is adjusted so that its initial temperature profile is similar to that of the present-day Earth. These models thus start out relatively cold, with mantle temperatures well below the solidus of dunite. The computations were performed at a variety of initial velocities and impact parameters, including pairs that give the Earth-Moon system its present angular momentum.

At the lowest velocity ($v_\infty = 0$), for impact parameters b of 0.88 (Fig. 7a) and 1.25 times the Earth's radius R_e, the strongest heating upon impact is confined to the hemisphere on which the projectile strikes. Shock-induced temperature rises are typically 2000 to 3000 K between the site of the impact and the Earth's core. A crater forms that extends most of the way down to the core. The gravitational energy of this excavation itself is of order 10^{30} J, which appears as heat within an hour or two as mantle material flows inward to fill the crater cavity. Unfortunately, in the high impact parameter runs the computations do not extend to long enough times for the entire projectile to merge with the Earth. This limitation is a result of the finite grid size: by the time fallback should have occurred a substantial amount of material had left the grid and could not be taken into account properly. These computations were thus stopped while some of the projectile was still falling on portions of the Earth more distant from the impact site. In these cases, more than half the mantle will be strongly heated, so the quoted results are a lower limit. Figure 7b illustrates the temperature contours for the $b = 0.88R_e$ computation 1800 s after the impact (this impact parameter and velocity correspond to the angular momentum of the present Earth-Moon system). Most of the Earth lies within contour D, that is, is at temperatures greater than 1800 K, and so is probably melted. Note that in this computation a very hot (>4200 K)

low-velocity vapor plume is expelled backwards from the impact site. This plume eventually spreads over the entire Earth, producing a transient silicate vapor atmosphere.

The results for the higher impact velocity $v_\infty = 7.8\,\mathrm{km\,s^{-1}}$ are more spectacular. For the impact parameters studied (0.59 and $1.25R_e$) the hemisphere near the impact was heated nearly uniformly by 1000 to 3000 K. The projectile's core was almost entirely vaporized and a much larger crater formed in the proto-Earth. A fast, hot vapor plume also carries several lunar masses of material out along trajectories that eventually take up elliptical orbits about the Earth. Figure 8b illustrates temperature contours for $b = 0.59R_e$, corresponding to the angular momentum of the present Earth-Moon system, at 1200 s after the impact. Again, a hot, low-velocity backwards vapor plume is formed that will eventually cover the Earth's surface.

IV. IMPLICATIONS OF GLOBAL MELTING

The three-dimensional hydrocode computations, in conjunction with the more general considerations described above, indicate that a Moon-forming impact would have had a profound effect on the Earth's thermal state. The shock produced by the impact would have heated the Earth to great depths, raising at least the hemisphere adjacent to the impact above the melting temperature. Later phenomena, such as the merger of the projectile's and proto-Earth's cores and the collapse of the mantle-deep crater created by the impact would have added comparable amounts of energy to the Earth. There seems to be no way to avoid the conclusion that a large Moon-forming impact is inevitably accompanied by widespread melting of most or all of the Earth's mantle.

Although the idea of molten planets is not new, having arisen most recently in the context of the lunar "magma ocean" (Wood 1970), modern geochemists have a difficult time accepting that the Earth once had a similarly extensive molten mantle. Kato et al. (1988) argue that fractionation of a small amount of perovskite and/or majorite garnet from a terrestrial magma ocean would drive the nearly chondritic ratios of most refractory lithophile elements substantially away from their observed chondritic values. Fractionation would also result in a primitive crust rich in Ba, K, Rb and Cs and relatively depleted in U, Th, Pb, Sr, Zr and Hf. In a similar vein, McFarlane and Drake (1990) argue that fractionation of olivine by flotation at high pressure would yield ratios of Sc/Sm, Ni/Co, and Ir/Au more than 10% above their observed chondritic values.

Ringwood (1990) argues that Kato et al.'s (1988) results can be used as a test of the giant impact hypothesis and Safronov accretion theory. Because the giant impact leads to global melting of the Earth, he infers that such an impact could not have occurred because the oldest (~4.2 Gyr) zircons and most ancient (~3.9 Gyr) crustal rocks would show no signature of the element fractionations noted above. Ringwood thus proposes that the Earth accreted

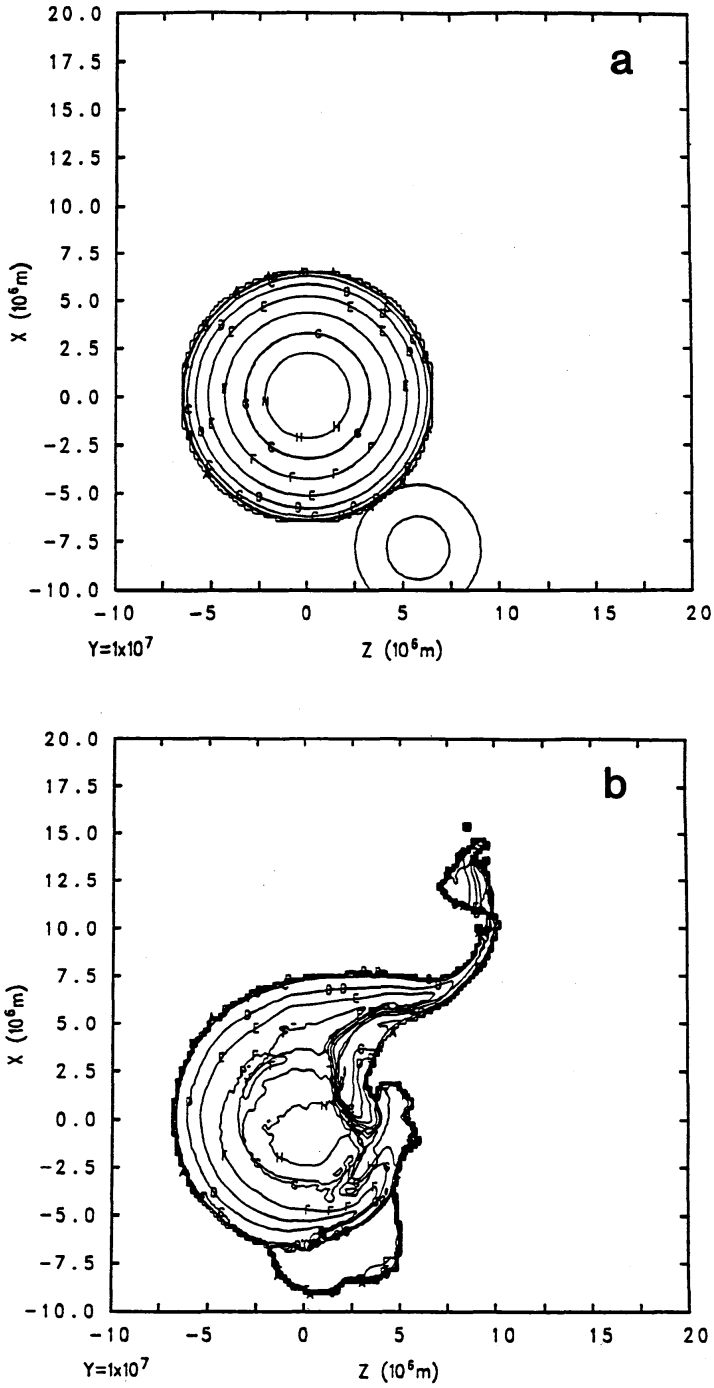

Figure 7. Temperature contours in the collision between the proto-Earth and a protoplanet half its diameter. This computation is for $v_\infty = 0$ km s^{-1} (8 km s^{-1} at contact) and an impact parameter of 0.88 R_e at contact. Plot (a) shows the initial configuration before impact in which the projectile is traveling upward (positive x-direction) and the proto-Earth is at rest. Plot (b) shows the configuration 1802 s after contact. The contour values are A = 300 K, B = 600 K, C = 1200 K, D = 1800 K, E = 2400 K, F = 3000 K, G = 3600 K and H = 4200 K. Figures 7 and 8 were computed by M. E. Kipp at Sandia National Laboratory, Albuquerque, NM using the three-dimensional hydrocode CTH. These plots are in the symmetry plane of the two colliding spheres.

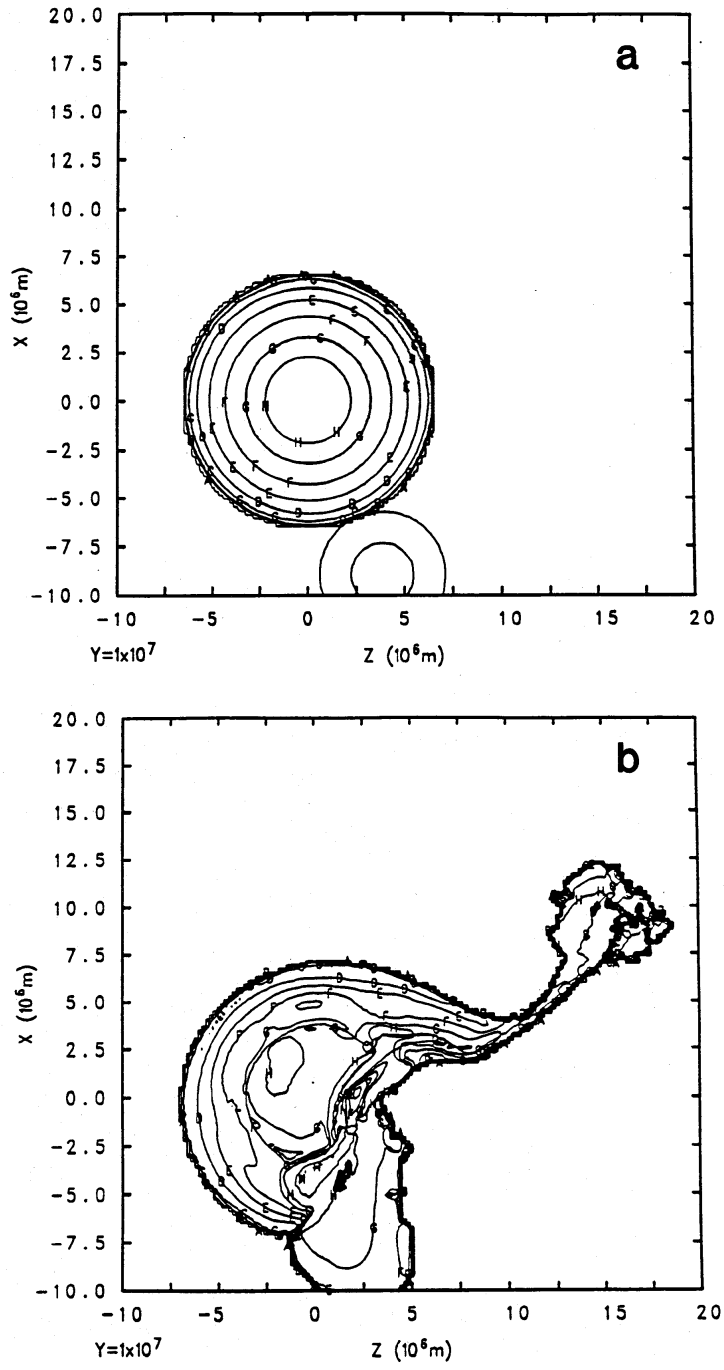

Figure 8. (a) Initial configuration and (b) temperature contours 1193 s after the impact between the proto-Earth and a protoplanet half its diameter moving at $v_\infty = 7.8$ km sec^{-1} (12 km s^{-1} at contact) and an impact parameter of 0.59 R_e at contact. Contour values are the same as in Fig. 7.

without widescale melting and suggests several conditions under which this might occur.

All of these objections make the implicit assumption that a magma ocean necessarily undergoes fractional crystallization, that is, that the crystals that

grow within the cooling magma are removed from the magmatic system by either settling or floating (because of density differences) and do not subsequently interact with the magma. However, it seems obvious that if convection is sufficiently vigorous, separation may be either inhibited or even prohibited. Thus, the reconciliation of the theoretical arguments for widespread melting in the Earth and the geochemical arguments prohibiting it may be rooted in the mechanical behavior of crystals in a convecting magma ocean.

The mode of convection in such a magma ocean, or any other naturally convecting system, is approximately characterized by the Rayleigh number Ra and Prandtl number σ (Chandrasekhar 1961), given by

$$Ra = \frac{\alpha g \Delta T D^3}{\kappa \nu} \tag{3}$$

$$\sigma = \frac{\nu}{\kappa} \tag{4}$$

where α is the thermal expansion coefficient, g the acceleration of gravity, D the depth of the convecting fluid, κ is the fluid's thermal diffusivity, ν its viscosity, and ΔT is the superadiabatic temperature gradient. Although these two parameters must be supplemented by others describing the temperature dependence of viscosity and the change of viscosity with depth in an exact model of magma ocean convection, the success of parameterized convection models suggests that the Rayleigh and Prandtl numbers define the major convective regimes to first order.

On the basis of the values of the Rayleigh and Prandtl numbers, magma ocean convection is utterly different from that of the more familiar sub-solidus convection now occurring in the planets, shown in Fig. 9. Magma ocean convection and crystal settling has more affinity to aeolian transport in a deep planetary atmosphere than to traditional studies of convection. In this range of conditions, a vigorously convecting system can be divided into three distinct regions (Kraichnan 1962): an upper conductive boundary layer, a sublayer dominated by viscous forces, and a volume of turbulent fluid whose flow is controlled by inertial forces (see Fig. 10). This inertial flow zone, which does not exist in the subsolidus convection familiar to most geophysicists, is dominated by large scale turbulent eddies that are analogous to gusty winds in a planetary atmospheric convective system. The inertial flow zone appears in the central part of a convecting fluid when the following inequality between the Rayleigh and Prandtl numbers is satisfied:

$$Ra^{1/3} \geq 35\sigma^{1/2}. \tag{5}$$

Convection in a magma ocean is expected to remain vigorous so long as the liquid can lose its heat near the free outer surface. Although a thin conductive skin may form at the immediate surface, differential convective

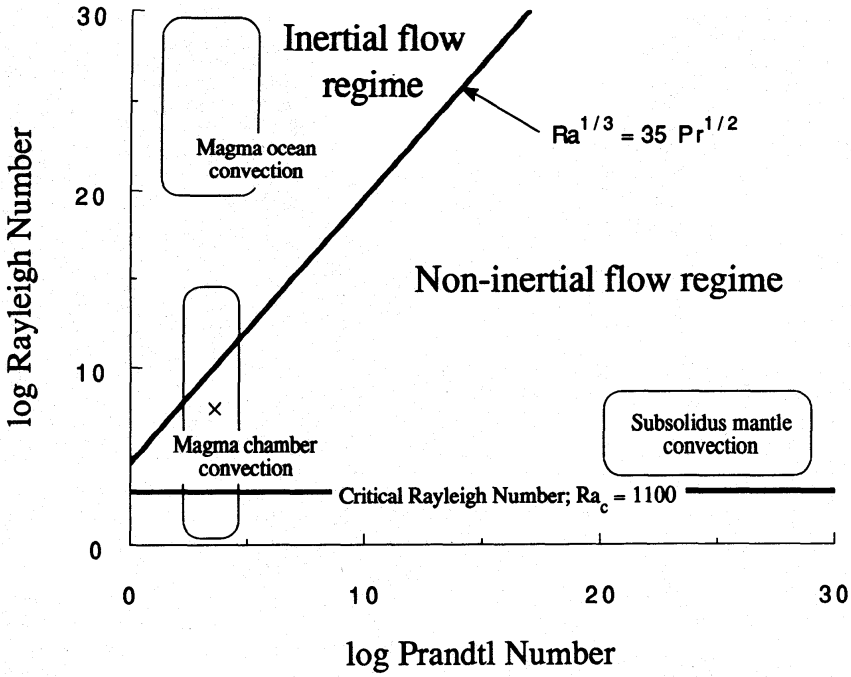

Figure 9. Rayleigh number vs Prandtl number, showing the convective regimes of common geologic processes and of planetary magma oceans. The inertial flow zone exists above and to the left of the inertial flow zone disappearance boundary and disappears to the right of the boundary. See the text and Fig. 10 for further explanation. The *x* marks where an experimental study (Martin and Nokes 1988) showed the onset of crystal suspension.

velocities are generally high enough to overcome the skin's strength and keep the heat loss near the theoretical value. This situation is quite different from that of a magma chamber, where a thick roof may strongly inhibit convection (Brandeis and Marsh 1989). The presence of an atmosphere may also have an important effect in limiting the rate at which heat is lost from a magma ocean. By blocking or limiting the rate of heat loss to space by infared radiation, an atmosphere can effectively blanket the top of the magma ocean, becoming a partial substitute for the thin upper conductive skin and controlling the cooling time scale of the planet's interior. A very thick initial atmosphere may even lead to the formation of a magma ocean without the intervention of a giant impact (Abe and Matsui 1985). The role of such atmospheric blanketing in controlling the surface temperature and rate of heat loss from a planet with a magma ocean is not well understood, and is an obvious target for future research.

A criterion for crystal suspension near the upper and lower margins of a convecting magma ocean can be derived from empirical studies of sediment transport. The Rouse number is defined as the ratio between the terminal

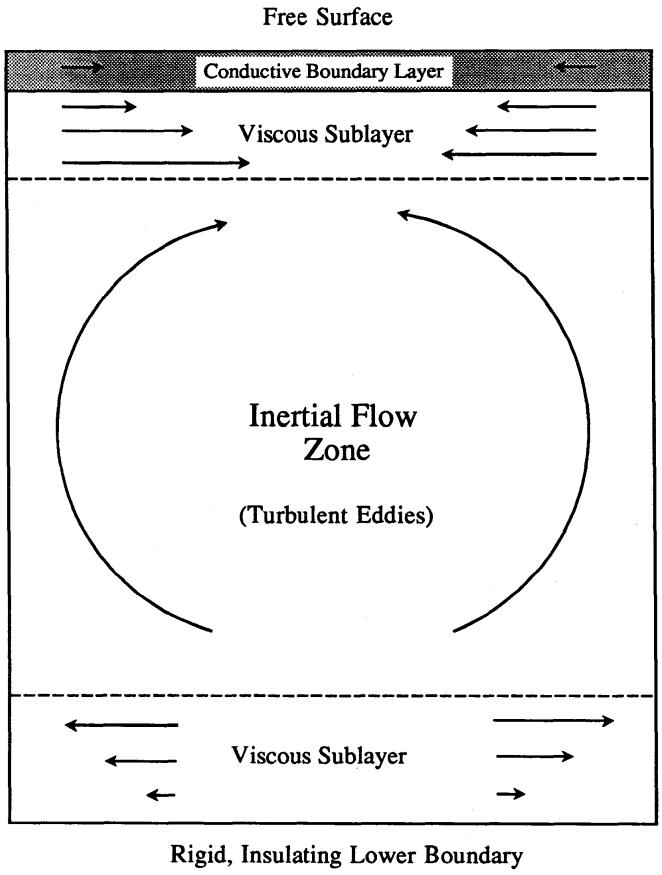

Figure 10. Diagram showing convection breaking into three distinct regions of behavior at very high Rayleigh number. The topmost region is a conductive boundary layer underlain by a viscous sublayer. In the third region, flow is dominated by inertial forces which produce large-scale turbulent eddies (the inertial flow zone).

velocity of the crystals in the fluid and the turbulent friction velocity. Settling (or flotation) takes place only for Rouse numbers greater than about 1 (Vanoni 1975). A study of the evolution of an ultramafic magma ocean with realistic magma rheology (Tonks and Melosh 1990) found that on both the Earth and Moon crystals up to 1 to 2 cm in diameter can be suspended from the onset of cooling. In the later stages of cooling, as viscosity increases, crystals of 10 cm to more than 1 m in diameter may easily be suspended. A surprising feature of the analysis is that the ability of a magma ocean to suspend crystals is only weakly dependent on gravity and depth: viscosity is the major factor determining suspension. This result, although gratifying for the Earth, raises a serious problem for the *lunar* magma ocean hypothesis, because the differentiation of an anorthositic crust on the Moon strongly suggests that crystal fractionation did occur there.

One explanation for this apparent paradox appeals to the low lunar gravity. Because pressure in the Moon increases slowly with depth compared to the Earth, the solidus and liquidus temperatures are nearly independent of

depth. In a convecting magma ocean on the Moon, adiabats lie between the solidus and liquidus at all depths, so that growing crystals circulate through the ocean for long periods of time without dissolving. Under these circumstances crystals may grow large enough to settle out, producing the observed differentiation. In contrast, on the Earth the liquidus and solidus profiles are steeper than an adiabat so crystallization only takes place over a restricted depth range. Convective velocities are so high that crystals nucleating in this zone do not have time to grow larger than a few microns and are thus incapable of settling out. In this way, even a totally molten initial Earth may have failed to differentiate by fractional crystallization.

Although magma ocean convection may be sufficiently vigorous to prevent fractionation between silicate melt and silicate crystals (whose density differs only a few hundred kg m^{-3} from that of the melt), the density difference between the silicate melt and droplets of molten iron (typically 2000–3000 kg m^{-3}) is great enough to allow iron droplets to rain out of the magma ocean, especially in the early stages when melt viscosities are low, and accumulate in a dense pool at the base of the magma ocean. Such dense pools of molten iron may, if sufficiently large, descend to the center of the growing planet, forming (or adding to) its protocore (Tonks and Melosh 1991).

V. ORIGIN AND EVOLUTION OF ATMOSPHERES

The origin and evolution of planetary atmospheres is a complex and poorly understood process. Elemental and isotopic compositions of planetary atmospheres differ from one another and from the composition of any single (or simple combination) of plausible volatile reservoirs in the early solar system (Pepin 1989). These reservoirs include solar nebular gas, solar wind, comets, and volatile-rich meteorites. Accretionary impacts have two competing effects on the development of planetary atmospheres: the release of volatiles due to shock heating and/or vaporization of the impactor and part of the target in relatively low-energy events may add to the planet's volatile inventory (Arrhenius et al. 1974; Benlow and Meadows 1977; Chyba 1990), and sufficiently large impacts may cause the loss of part of the existing volatile inventory.

During most of the accretionary process, impact velocities generally differ little from the escape velocity of the growing protoplanet because most of the collisions are between bodies in nearly matching orbits. At some stage in planetary growth, the impact velocities become high enough that the shock pressures resulting from the impact are high enough to devolatilize the impactor (and possibly part of the target). For carbonaceous chondrites, partial devolatilization can begin on the Earth and Venus when the planetary embryos reach approximately 12% of their current size, and complete devolatilization begins when the embryos are approximately 30% of their current size (Ahrens et al. 1989; Lange and Ahrens 1982). The corresponding values for Mars are roughly 23% for partial devolatilization and 56% for complete devolatilization. Devolatilization of comets presumably occurs at lower planetary radii

because comets are more volatile than silicates and because their lower density with respect to the target means that a larger fraction of the impact energy is partitioned into heating the projectile. Abe and Matsui (1985) consider thermal models of accretion in which CO_2, H_2O and fine impact ejecta are produced in accretionary impacts. They conclude that the proto-atmosphere forms a thermal blanket that traps the internal energy added by impacts at or near the planet's surface. The result of this thermal trapping is that the surface temperature becomes high enough to induce devolatilization regardless of the shock pressure produced by an impact, once the Earth (or Venus) achieves ~30% of its final radius. This modeling thus strongly suggests that substantial atmospheres/hydrospheres could develop this way.

VI. ATMOSPHERIC EROSION

Toward the end of accretion, however, collisions are rarer but much more energetic, involving large planetesimals and higher impact velocities, as discussed above. Such impacts may cause a net loss of atmosphere from a planet, so that the cumulative effect of impacts during the period of heavy bombardment might have dramatically depleted the planets' original atmospheres (Cameron 1983). The basic concept is that the fraction of the projectile's momentum that is transferred to the atmosphere may be sufficient to accelerate a portion of the atmosphere to velocities greater than the escape velocity of the planet.

A meteorite or comet may transfer momentum to the atmosphere of the target planet in three ways: first, during the initial passage of the impactor through the atmosphere, there is a direct transfer of momentum as the impactor penetrates the atmosphere, compressing and accelerating the gas in front of it. A small meteorite or comet may be broken up by, or significantly ablated during, its passage through the atmosphere, and thus deposit most or all of its energy directly in the atmosphere, mainly as shock waves and heat. This is apparently the explanation for the famous Tunguska event of 1908, in which shock waves from a meteoroid that did not quite reach the ground leveled about two thousand square kilometers of dense Siberian forest, snapping off meter-diameter trees like matchsticks. Large impactors, however, are not significantly slowed by the atmosphere: O'Keefe and Ahrens (1982) showed that in this case, the impactor delivers only a small fraction of its momentum directly to the atmosphere, and Walker (1986) showed that this momentum is distributed in such a way that no significant amount of atmosphere escapes from a planet with an escape velocity $\geq$10 km s^{-1}. This mechanism is thus negligible for the Earth and Venus, but it may have contributed to atmospheric erosion on Mars. Second, solid ejecta thrown out of the growing crater can transfer momentum to the atmosphere in a similar fashion, but again this has been shown to result in negligible atmospheric loss even for large, high-speed impacts (Melosh and Vickery 1988), even on Mars. Third, for a sufficiently energetic impact, a great deal of very highly shocked impactor (and possibly target) material expands upward and outward at high velocities, driving the

overlying atmosphere ahead of it (Melosh and Vickery 1989; Vickery and Melosh 1990; Watkins 1983). This last mechanism is by far the most important for atmospheric erosion, and we discuss it in more detail below.

A. Minimum Projectile Size

In order that a high-energy ejecta plume should form, the most basic requirement is that the projectile should be large enough to strike the target surface at high velocity and remain intact. The minimum diameter $L_{\min}$ of a projectile that reaches the surface with about half of its initial (pre-atmospheric) velocity, neglecting ablation and crushing effects, is

$$L_{\min} = P_a/(\rho_0 g \sin\theta) \tag{6}$$

where P_a is the surface atmospheric pressure, g is the acceleration of gravity at the planet's surface, ρ_0 is the projectile's density and θ is the angle of entry, measured from the horizontal. Thus, a vertically incident ($\theta = 90°$) stony projectile on Earth with its current atmospheric pressure must be larger than about 3.4 m in diameter to strike with half or more of its initial velocity, while on Venus such a projectile must exceed a diameter of 350 m. If atmospheric pressures were greater than their current values at early times in the history of a given planet, this minimum size would have been correspondingly greater. In actuality, these limits are probably much too small because they neglect ablation and atmospheric breakup, both of which tend to increase the minimum size of a projectile that can reach the surface with high velocity. Ablation is most important for meter-sized projectiles, so that a limit placed by the process of atmospheric breakup is probably more realistic.

Projectiles entering a planetary atmosphere encounter aerodynamic stresses of order $\rho(z)v_e^2$, where $\rho(z)$ is the atmospheric density at altitude z and v_e is the entry velocity. These stresses may reach many kilobars in the lower atmospheres of the Earth and Venus, and so the entering projectiles are nearly certain to be crushed. This effect is independent of the diameter of the projectile. Although the broken fragments of smaller projectiles may be scattered by these stresses, creating the familiar km-wide elliptical strewn fields, the fragments of larger projectiles do not have time to separate significantly. These fragments thus fall in nearly the same location and produce a crater almost indistinguishable from one produced by a single intact projectile. For a vertically incident projectile, the diameter L_c above which breakup is not important is given by (Melosh 1981)

$$L_c = 2(\rho_a/\rho_0)^{1/2}H \tag{7}$$

where ρ_a is the density of the atmosphere at the surface and H is the scale height. Because the surfaces of both planets are quite young, we use current values for the atmospheric parameters to calculate L_c, then use scaling relations to predict minimum crater sizes, and compare these to observations. On

Earth, L_c is about 300 m for stony projectiles, while on Venus it is 4500 m. This implies that the minimum-size hypervelocity impact crater that can form on Earth is a few km diameter (for iron projectiles, the minimum is about a kilometer—about the size of Meteor Crater, Arizona, which was clearly formed by an iron projectile that at least partially fragmented in the upper atmosphere), while the minimum size crater that can form on Venus is about 30 km in diameter, in good agreement with the observations of the Magellan spacecraft (Phillips et al. 1991). We believe therefore that this model gives a reasonably accurate estimate of the minimum size projectile that can survive atmospheric passage as a function of the density structure, and so may be applied to early (hypothetically) denser atmospheres.

B. Melt and Vapor Plume Formation

When a projectile strikes a planet's surface at high speed, a rapid but orderly sequence of events is initiated that ultimately results in a crater. The principal topic here, however, is not the crater itself but the plume of hot melt and vapor that expands into the atmosphere above it. The ultimate origin of this plume lies in the thermodynamics of shock compression and release. When the projectile first strikes the surface, a region of high pressure develops at the interface that spreads into both the projectile and surface rocks as shock waves expand away from the contact area. These waves quickly reach the sides and rear free surface of the projectile and are reflected from there as rarefaction waves that then propagate back downward, releasing the compressed projectile to low pressure. In the surface rocks the shock waves travel outward and downward, engulfing more material and weakening as they proceed.

The resulting pattern of maximum shock pressure is shown schematically in Fig. 11. The projectile and a roughly equal mass of target (surface) rocks are shocked to the highest pressure P_o in a volume comparable to that of the projectile itself (strong shock compression decreases the initial volume of most materials by a factor of about 2). For a target and projectile of the same composition, P_o can be simply estimated using the Hugoniot equations and impedence matching to give

$$P_o = \rho_o \left(C + S\frac{v_i}{2}\right)\frac{v_i}{2} \tag{8}$$

where v_i is the impact velocity, ρ_o is the density of the uncompressed material and C and S are material constants. For a typical crustal rock (granite), $\rho_o = 2630$ kg m^{-3}, $C = 3.68$ km s^{-1} and $S = 1.24$ (Kieffer and Simonds 1980). At an impact velocity of 20 km s^{-1}, this yields a maximum pressure P_o of about 425 GPa, somewhat higher than the pressure at the Earth's center. For the more complex case where the target and projectile differ in composition see Melosh (1989).

The maximum shock pressure outside this "isobaric core" (Croft 1982*a*) falls approximately as a power of distance away from the impact site

$$P = P_o\left(\frac{a}{r}\right)^n \tag{9}$$

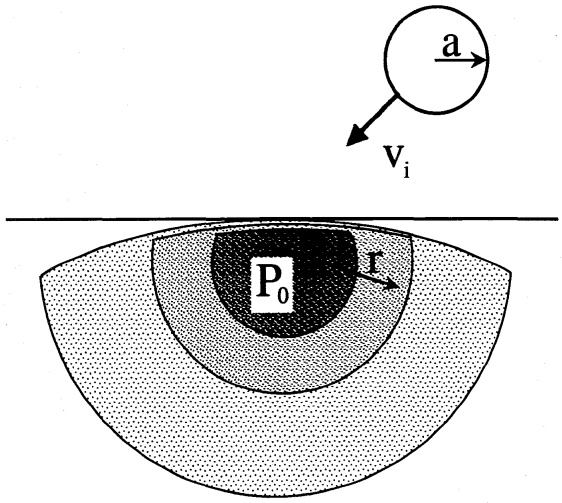

Figure 11. Schematic illustration of shock contours beneath the site of an impact. The projectile of radius a strikes the surface at a velocity v_i. A region of nearly uniform high pressure P_0 develops with a volume comparable to that of the projectile itself. Outside this region the pressure falls off with increasing distance r from the impact site. The free surface is constrained to zero pressure, so the pressure contours fall off rapidly in a narrow near-surface zone.

where a is the radius of the projectile and the power n ranges between 3 for strong shocks, declining to 2 at shock pressures lower than 100 GPa. The shock wave thus establishes a pattern of highly shocked rocks near the impact site surrounded by roughly concentric zones of successively less shocked material. The near-surface zone is a special case, where a very thin layer of target material is protected from high shock pressure by the boundary condition at the free surface. The small amount of material within this zone experiences only low shock pressures, but the pressure gradients are high, and the material is ejected at high speed. The concentric shock pressure structure is established long before the excavation flow has time to open the crater. The pattern is subsequently altered by the cratering flow, and may not be discernible except in the deformed rocks directly beneath the final crater.

The heating of shocked materials is the result of a thermodynamic cycle. Shock compression is a thermodynamically irreversible process. At a shock front, mass, momentum and energy are conserved, but not entropy. On a P-T plot, the state of the shocked material jumps discontinuously from its initial state to a final state lying on the Hugoniot curve, a locus of points each of which results from different degree of shock compression, i.e., this curve does *not* define a thermodynamic path traversed by the shocked material. The shocked material then decompresses along an adiabat, which does define a continuous path of P-T states attained by the material.

The standard P-T plot, however, is not very convenient for illustrating phase transformations induced by shock, as the adiabats are complicated curves whose paths give little insight into the ultimate state upon decompression, especially near the critical point. In this chapter, we introduce a P-S

plot (S is entropy) which, although perhaps unfamiliar, contains far more information about the ultimate fate of shocked material. One major advantage of this plot is that release adiabats are vertical lines (constant entropy), and the state of the material at 1 bar (10^{-4} GPa) is generally well known. Thus, if the entropy along the Hugoniot can be computed, the phases present upon decompression to 1 bar, at least, are well determined.

Figure 12 shows an example of a Hugoniot and the phase structure of forsterite on a P-S plot. The Hugoniot and liquid-vapor phase curve were computed from the ANEOS equation of state (Benz et al. 1989), modified slightly to correct a small error in the input parameters. The thermodynamic cycle of shock compression followed by adiabatic release does work on the material, and upon release to low pressure the shocked material may be in the form of a hot solid (for weak shocks), a liquid or a gas (at the highest pressures). Figure 12 indicates that at shock pressures less than about 500 GPa (with a corresponding particle velocity of 10 km s^{-1}), the entropy of the shocked material is less than that required to reach the critical point, and the release curve crosses the liquid/vapor phase curve from the liquid side. Thus, for most asteroidal impact velocities, silicates never vaporize completely (for this to occur the release adiabat must pass to the right of the critical point in Fig. 12), but instead come to a boil, rapidly releasing vapor bubbles and breaking the melt apart into discrete clumps as the pressure declines (Melosh and Vickery 1991). Once the material reaches the liquid/vapor phase curve, its temperature and pressure are linked by the phase relation $P\,(\mathrm{GPa}) \approx 220 \exp(-48{,}900/T)$. The relative masses of liquid and vapor (or solid and vapor below the triple point at about 10^{-9} GPa) can be determined from this diagram by the lever rule. When considering the problem of atmospheric erosion, it is not critical whether the material expands into the liquid field and breaks up into droplets or expands into the vapor field and condenses into droplets: in both the cases the plume expands adiabatically, engulfing and accelerating the ambient atmosphere. The difference may be important, however, for the problem of tektite and microspherule formation.

The specific internal energy initially deposited in the shocked material is approximately equal to one-half of the particle velocity squared. Because the particle velocity itself is about one-half of the impact velocity (for projectile and target composed of the same material), the internal energy of the shocked material per unit mass is given by $E \approx v_i^2/8$. When the impact velocity is high enough that melting or vaporization occur, this energy is available for acceleration of the material out of the growing crater. The average final velocity attained by the isentropic expansion of a fluid with initial internal energy E is given by $v_\infty = (2E)^{1/2}$, or in this case $v_\infty = v_i/2$. Thus, the plume of melt and vapor expanding away from the site of an impact eventually achieves a mean speed of about half the impact velocity. Furthermore, detailed numerical computations with the ANEOS equation of state for dunite show that the edge of the expanding plume typically moves at twice this speed, or at nearly the impact velocity itself.

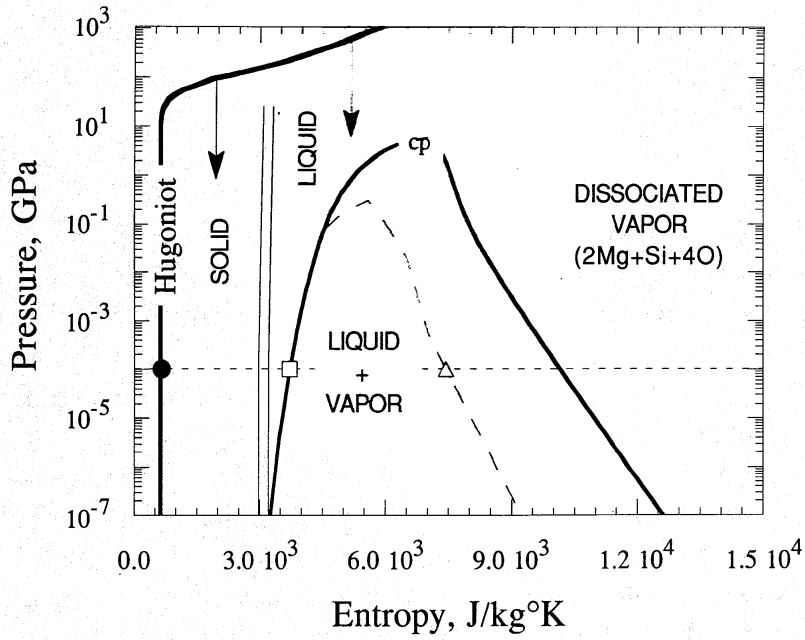

Figure 12. The thermodynamic path of shocked dunite in P-S (pressure-entropy) space. The Hugoniot curve and solid liquid-vapor phase curve are based on the ANEOS equation of state for dunite (Benz et al. 1989), with the slightly corrected parameters $ZB(11) = 2.23 \times 10^{11}$, $COT(1) = 0.571$, $COT(2) = 0.143$, $COT(3) = 0.286$. The entropy of the 1 bar (10^{-4} GPa) reference state (from the JANAF tables) is shown as a heavy black circle, the near-vertical light lines denoting incipient and complete melting are constrained at 1 bar by JANAF data, and the points for incipient (square) and complete (triangle) vaporization are from Ahrens and O'Keefe (1972). Note that although the ANEOS phase curve agrees well with the point for incipient vaporization, it lies at much too high an entropy at complete vaporization. Similarly, the critical point is at about a factor of 10 higher pressure than that derived from the hard sphere formalism (Ahrens and O'Keefe 1972). These displacements are a result of ANEOS's implicit assumption that the gas phase is a monatomic mixture of species, whereas vaporized dunite is likely to consist mainly of a mixture of MgO, SiO and O_2. The long-dashed curve is an interpolation through the hard-sphere critical point indicating the probable true nature of this phase boundary. Release adiabats are shown schematically as downward pointing arrows. Note that even for the hard-sphere critical point at $S = 5600$ J kg^{-1} K^{-1} the shock pressure must exceed about 500 GPa (with corresponding particle velocity of 10 km s^{-1}) to reach the liquid-vapor curve from the vapor side.

C. Atmospheric Erosion by Melt and Vapor Plumes

The structure of this expanding plume of hot melt and vapor is initially rather complex, as revealed by detailed numerical computations (O'Keefe and Ahrens 1982), but as time goes on, it quickly attains a roughly hemispherical shape (Basilevsky et al. 1983). The mean density of the plume declines below the melt density by the time it has expanded to about twice the projectile radius, indicating that by this time the plume is composed of droplets of melt in local thermodynamic equilibrium with vapor. The gas and melt droplets are strongly coupled, however, until much later in the expansion. As this fast-

moving dispersion of liquid droplets and vapor meets the ambient atmosphere, shock waves form at the interface that decelerate the plume while accelerating the surrounding air.

In the case of a small impact, the mass of the atmosphere may be sufficiently great that the plume's expansion is entirely halted. If this occurs over a distance small compared to the atmospheric scale height H, the hot melt and vapor form a hemispherical bubble that rises into the stratosphere under the influence of its own buoyancy, like a hot air balloon. This is the classic fireball scenario, familiar to everyone as the mushroom cloud emblematic of nuclear explosions. An upper limit on the projectile size for this scenario can be roughly derived by equating the radius of a vapor "fireball" to the atmosphere's scale height H (see Melosh 1989; Jones and Kodis 1982). Assume that the projectile and an equal mass of target begin expanding from an initial pressure given by the second Hugoniot equation, $P_i = \rho_p(v_i/2)[C + S(v_i/2)]$, where v_i is the impact velocity on the surface (the initial particle velocity is $v_i/2$ if projectile and target materials are similar), C and S are material constants in the linear shock-particle velocity equation of state, and ρ_p is the projectile density. At high velocities, only the S term is important: $P_i \approx \rho_p S v_i^2/4$. Then the radius of the smallest projectile that can generate an atmosphere-piercing plume is:

$$a_p = H\left(\frac{P_a}{S\rho_p v_i^2}\right)^{1/3\gamma} \tag{10}$$

where P_a is the surface atmospheric pressure and γ is the ratio of specific heats ($\gamma = 1.4$ for air). As $S \approx 1.3$ for most silicates (Melosh 1989, Table AII.2), $H = 10$ km on Earth, and taking $v_i = 20$ km s^{-1} this equation yields $a_p \approx 200$ m, corresponding to a transient crater about 6 km in diameter. Larger projectiles than this produce vapor plumes that cannot be stopped by the atmosphere.

Because the smallest mass of atmosphere is vertically above the crater (Fig. 13), it is expected that projectiles somewhat larger than the limit (Eq. 10) will create strongly focused vertical plumes above the impact site. The maximum velocities in these plumes may approach the impact velocity, and thus some material (mostly derived from the projectile itself) may be ejected back into interplanetary space. A simple method of estimating whether the atmospheric gases will be ejected is to compute the momentum of the expanding melt and vapor plume in a small element of solid angle extending outward from the impact site. The mass of the atmospheric gases in the same solid angle is then computed and added to the mass of the crater ejecta. Momentum conservation then determines the mean velocity of the mixture. If the mean velocity is greater than escape velocity, then the atmosphere in this element of solid angle is considered to be ejected. A minimum requirement is that the mean velocity of the melt and vapor plume must exceed escape velocity, so that the initial impact velocity must be twice the escape velocity, or about 20 km s^{-1} on the Earth. Similarly, the mass of the plume must be sufficient

to accelerate the atmosphere along with it. This puts a lower limit on the projectile size for atmospheric loss to occur, which we have found to be close to 10^{13} kg or roughly 2 km in diameter (Vickery and Melosh 1990) for silicate projectiles and Earth's current atmosphere. This lower limit would be greater for Venus and less for Mars at the present time, because it increases with increasing atmospheric mass, and decreases with increasing H/R, the ratio of the atmospheric scale height to planetary radius.

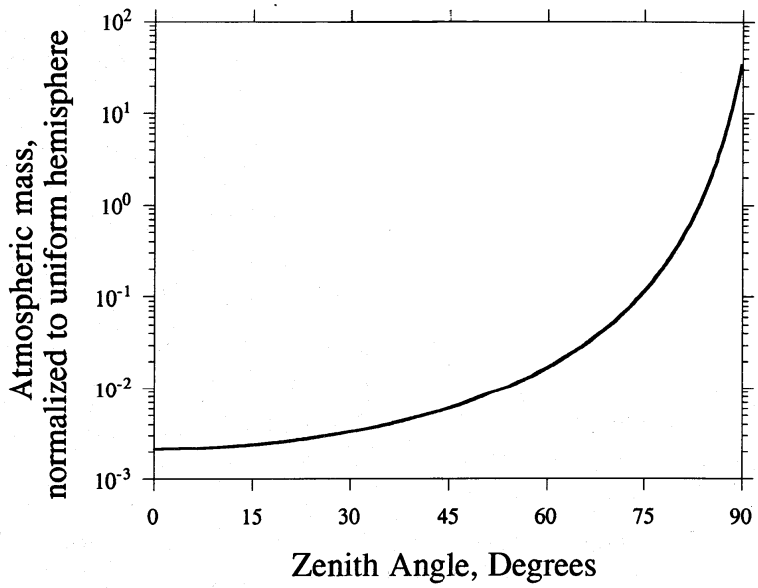

Figure 13. Mass in the Earth's atmosphere as a function of zenith angle (0° is vertical, 90° is horizontal), normalized by the mass in a uniform hemisphere of the same total mass. This gives the "coupling factor" between an expanding hemisphere of gas and the ambient atmosphere. The mass in the Earth's atmosphere is small directly overhead and strongly concentrated toward the horizon. The mass in an element of solid angle is the same as a uniform hemisphere at a zenith angle of 83°.2 (based on the U. S. Standard Atmosphere).

As the size of the projectile increases, more and more of the atmosphere is liable to be stripped away from the planet by the expanding plume of vapor mixed with melt droplets. The maximum that can be stripped is given approximately by the mass lying above the plane tangent to the planet at the impact point, $m_{tp} = 2\pi P_a H R_P/g$, where R_P is the radius of the planet, P_a is the surface atmospheric pressure and g is the surface acceleration of gravity (Melosh and Vickery 1989). This amounts to about 0.1% of the total mass of the atmosphere (the fraction is H/R_E), or 3×10^{15} kg on the Earth. Detailed numerical computations (Vickery and Melosh 1990) have shown that at impact velocities of >20 km s^{-1}, the atmospheric mass ejected is not a strong function of impact velocity and that projectiles with masses in the range of 10^{17} kg and above are capable of removing essentially all of the atmosphere

above the tangent plane on Earth. These computations are difficult, however, because most of the atmosphere's mass is concentrated at low elevations from the impact point (see Fig. 13), whereas the expanding plume of melt and vapor is more or less isotropic. The two thus do not couple well together. The coupling may be enhanced, however, by the concentration of the plume's mass at low angles, as is expected for oblique impacts (Schultz and Crawford 1987). Such low-angle processes as jetting (Melosh 1989) or ballistic curvature of lower-speed ejecta may also enhance the efficiency of atmospheric loss. To achieve accurate estimates of the process of atmospheric erosion by impacts, it is thus necessary to perform full scale numerical computations on the fast, low-angle ejecta from an impact, and to consider the radius of curvature of the target planet. Such computations have not yet been done with sufficient accuracy: this is clearly a direction that future work on atmospheric erosion by impacts must take.

Although an exact treatment of the projectile mass needed to eject the atmosphere above the tangent plane is still lacking, the computations discussed above (Vickery and Melosh 1990) show that a projectile mass $m_\star$ between 1 and 10 times m_{tp} is needed to strip the atmosphere above the tangent plane. Using this estimated threshold along with a postulated cratering flux permits the rate of atmospheric erosion by impacts to be roughly computed. The present cumulative flux (number/sec/m^2) of projectiles with mass greater than or equal to m is parameterized by $N_{\rm cum}(m) = am^{-b}$, where a and b are constants. Unfortunately, neither the over-all flux a nor the slope of the distribution b are well known. A pair of constants that predicts the correct present lunar crater distribution via the revised Schmidt-Holsapple scaling law (Schmidt and Housen 1987) is $a = 1.55 \times 10^{-23} (MKS), b = 0.47$. The slope of the crater distribution on the Martian plains is the same as the lunar distribution (Neukum and Wise 1976; Strom 1984), while the overall cratering rate on Mars at present is estimated to between 1 and 4 times the lunar rate, with a preferred mean of about 2 (BVSP 1981).

Using a constant cratering rate, the rate of mass loss from a planet's atmosphere $dM_{\rm atm}/dt$ is given by the flux $N_{\rm cum}(m_\star)$ of projectiles large enough to remove the atmosphere above the tangent plane, times the planet's surface area $4\pi R^2$, times the mass above the tangent plane m_{tp}. As $m_\star$ itself depends upon $M_{\rm atm}$, a simple differential equation for $M_{\rm atm}$ (or, equivalently, P) results whose solution, converted to a convenient and universal form, is:

$$\frac{M_{\rm atm}(t)}{M_0} = \frac{P(t)}{P_0} = \left(1 - \frac{t}{t_\star}\right)^{1/b} \tag{11}$$

where the M_0 is the present ($t = 0$) atmospheric mass, P_0 is the present surface atmospheric pressure, and $t_\star$ is the length of time required for impacts to reduce the atmospheric pressure to zero. In terms of previously defined quantities,

$$t_\star = \frac{f_{\rm crit}^{\,b}}{2\pi ab(RH)^{1-b}} \left(\frac{4\pi P_0}{g}\right)^b \tag{12}$$

where f_{crit} is the ratio of the critical projectile mass $m_\star$ to the mass above the tangent plane m_{tp}. This ratio is assumed to be a constant. As $b = 0.47$, the characteristic time $t_\star$ is not very sensitive to uncertainty in f_{crit}.

Note that the above equation for the time dependence of atmospheric mass or pressure can be used for times before the present ($t<0$), so that the atmospheric pressure at any previous era can also be computed. It is interesting to note that the atmospheric pressure declines rigorously to zero at time $t_\star$—it does not simply fall exponentially towards zero. This interesting fact is a unique characteristic of the impact erosion mechanism. As the atmospheric pressure declines, smaller projectiles are capable of removing the atmosphere above the tangent plane. But by the distribution law, there are more smaller projectiles than larger ones, so a greater fraction of the atmosphere is removed in each unit time interval. The net result is the complete stripping of the atmosphere after time $t_\star$, barring volcanic or other sources of replenishment.

The post-late-heavy-bombardment flux, along with other parameters and the assumption $f_{crit} \approx 1$, gives a value $t_\star \approx 60$ Gyr for Mars at the present time (note that $t_\star$ for Earth and Venus is longer by several orders of magnitude, so that impact erosion of their atmospheres is entirely negligible). This value for $t_\star$ predicts an atmospheric pressure on Mars at the end of late heavy bombardment (3.2 Gyr) only 1.1 times the present low pressure. This would not explain the geomorphic evidence (Baker and Partridge 1986) for running water on Mars' surface early in its geologic history. However, it is widely known (*BVSP* 1981) that the impact flux on the Moon in the first Gyr of its history was many orders of magnitude larger than at present. It is presumed (Neukum and Wise 1976) that Mars, too, shared in this era of heavy bombardment, so that the enhanced impact fluxes of this era may have produced an early period of rapid atmospheric pressure change. The heavy bombardment flux can be adequately fit by the expression

$$N_{cum}(m,t) = a\left(1 + Be^{-\lambda(t+4.6)}\right)m^{-b} \tag{13}$$

where a and b are the same as before, and $B = 1.17 \times 10^4$ and $\lambda = 6.94$ Gyr^{-1} (Chyba 1990).

The differential equation for the rate of atmospheric mass loss implied by Eq. (13) can be readily integrated, giving a simple analytic expression similar to Eq. (11) for the evolution of atmospheric pressure as a function of time t from the present:

$$\frac{M_{atm}(t)}{M_0} = \frac{P(t)}{P_0} = \left(1 - \frac{t}{t_\star} - \frac{Be^{-4.6\lambda}}{\lambda t_\star}[1 - e^{\lambda t}]\right)^{1/b}. \tag{14}$$

The atmospheric pressure of Mars predicted by this expression is shown in Fig. 14. It is clear that early in Mars' history its atmospheric pressure may have been as much as 100 times its present value, or approximately 1 bar. It is interesting to note that this pressure is near that considered necessary for Mars

to have supported free water on its surface (Cess et al. 1980; Kasting and Toon 1989). Of course, other processes besides impact erosion probably played major roles in determining Mars' atmospheric pressure, so this computation should be regarded merely as an indication of the magnitude of the effect and time interval over which impact erosion may have been important.

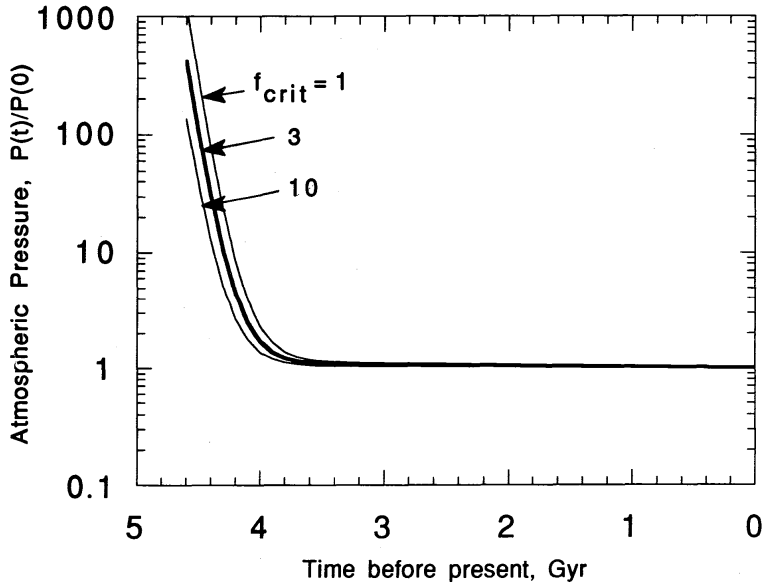

Figure 14. The history of atmospheric pressure on Mars in a model where impact erosion alone can affect the mass of the atmosphere (although this is an unrealistic assumption, it serves to highlight the effect of this process). The different curves reflect different assumptions about the ratio f_{crit} between projectile mass at the threshold of atmospheric erosion and the mass above the tangent plane. The atmospheric pressure declines rapidly through the period of heavy bombardment.

In the above computation, we neglected the flux of volatiles delivered by impacts (as well as those from mantle degassing, etc.). Whether a given impact will add or erode volatiles is a function of its impact velocity as well as its size, so that the spectrum of velocities as well as the compositions of projectiles striking a planet's surface must be known (or assumed from a plausible model). A beginning on this problem has been made by (Chyba 1990), although it is clear that a great deal of additional work must be done before this topic is fully understood.

One additional consequence of the impact erosion of atmospheres, however, is worth noting. Volatile constituents in the atmosphere above an impact site are vulnerable to stripping by the expanding plume of vapor. The mass that may be removed is $m_{tp} = 2\pi P_a H R_P / g$, which is a fraction $H/2R_P$ of the total atmospheric mass. However, if the volatile is condensed on the surface of the planet, the maximum amount expelled with the expanding vapors is roughly equal that intersected by the projectile, equal to the area of the projectile projected onto the surface of the planet times the thickness t of

the condensed volatile, $\pi a^2 t$, where a is the projectile radius. As the total amount of this material on the planet is $4\pi R_P^2 t$ (assuming a uniform distribution) the fraction of the consensed volatile removed is thus $a^2/4R_P{}^2$. At the threshold for removal by atmospheric erosion $m_\star = f_{\text{crit}} m_{tp}$, so at radius $a_\star = (3m_\star/4\pi\rho_0)^{1/3}$ we can show

$$\frac{\text{Condensibles Ejected}}{\text{Total Inventory}} = \{\left[\frac{3}{4} f_{\text{crit}} \left(\frac{\rho_a}{\rho_0}\right)\right]^{2/3} \left(\frac{H}{2R_P}\right)^{1/3}\} \frac{H}{2R_P}. \quad (15)$$

Because this ratio is generally much smaller than $H/2R_P$, it is clear that surface condensates are much less vulnerable to impact erosion than atmospheric gases. In this way, condensible volatiles may have been fractionated from noncondensible ones during accretion or heavy bombardment. For example, Mars may have acquired a substantial inventory of water (which condenses or freezes on or near the surface) while only retaining a small amount of the noncondensible gas nitrogen. The fate of CO_2 is more complex, as it can both form carbonates on the surface as well as remain uncondensed in the atmosphere.

VII. SUMMARY

The effects of "giant" impacts (i.e., impacts by bodies roughly half the diameter of the primary) on the thermal state of a growing planetary embryo is a more general problem than that of the hypothetical Moon-forming impact on the proto-Earth. If the cumulative spectrum of planetesimal sizes is close to a power law of slope -2, then most of the mass and energy added to a growing planet will be deposited by such "giant" impacts. In this case a simple pattern of temperature versus radius may never develop, as the thermal state at any given era will depend upon the time, velocity and obliquity of the last large collision. The process of core formation and differentiation in a catastrophically growing body may be qualitatively different than that suggested by current gradualist models of planetesimal growth. Global melting and the formation of "magma oceans" may have occurred repeatedly in the history of all planetary-size bodies. The implications of such catastrophic growth have yet to be worked out, but it is clear that much more work needs to be done on the effects of large impacts on protoplanets of all sizes, and the chemical and mechanical evolution of planetary-scale magma bodies.

Impacts also contribute in a fundamental way to the origin and evolution of atmospheres. Devolatilization of impactors (and of surface rocks, if they have a volatile-rich composition) during accretion may lead to the formation of relatively dense proto-atmospheres. Giant impacts, such as that which is believed to have formed the Moon, influence atmospheric development in ways that are presently unclear: although a number of authors declare that giant impacts will simply strip away any pre-existing atmosphere, the loss mechanism is sufficiently obscure that this conclusion is open to doubt.

The cumulative effect of the smaller, but much more numerous, impacts during heavy bombardment was to strip away a fraction of the primordial atmospheres; this effect would have been much more pronounced for Mars than for the Earth or Venus. Impact erosion is much more effective for atmospheric gases than for condensed liquids or ices on the surface, and may thus have helped to segregate noncondensible from condensible substances.