

DYNAMICS OF LUNAR FORMATION

Robin M. Canup

*Department of Space Studies, Southwest Research Institute, Boulder,
Colorado 80302; email: robin@boulder.swri.edu*

Key Words Moon, planetary formation, satellite formation, impacts, planetary satellite

■ **Abstract** The giant impact theory is the leading hypothesis for the origin of the Moon. This review focuses on dynamical aspects of an impact-induced lunar formation, in particular those areas that have advanced considerably in the past decade, including (a) late-stage terrestrial accretion, (b) giant impact simulations, (c) protolunar disk evolution and lunar accretion, and (d) the origin of the initial lunar inclination. In all, recent developments now provide a reasonably consistent dynamical account of the origin of the Moon through a late giant impact with Earth, and suggest that the impact-generation of satellites is likely to be a common process in late-stage solid planet formation.

1. INTRODUCTION

Although the Moon is by far the most familiar and well-studied satellite, it is a rather unusual planetary object. The low lunar density implies that the Moon is severely depleted in iron compared with Earth and other inner solar system objects. The Moon is further distinguished by its relative size compared with its parent planet, containing $\sim 1\%$ of Earth's mass. In contrast, Mercury, Venus, and Mars lack large moons, and even the largest satellites of the outer gaseous planets contain just $\sim 0.01\%$ of their planet's mass. Only in Charon, Pluto's moon, do we find another massive planetary companion. The angular momentum of the Earth-Moon system is also unusually high; if it were contained solely in Earth's rotation, it would yield an approximately 4-h day, much shorter than those of the other inner planets. Compositionally, the Moon presents mixed messages regarding its heritage, in that lunar materials both resemble and differ from terrestrial mantle material. One notable difference is that the Moon is more depleted in volatile elements, suggesting it was subject to additional thermal processing. Finally, the Moon's orbital radius has expanded more significantly over its history than any other planetary satellite as a result of tidal interactions with Earth, such that the Moon in its early stages was about 15 times closer to Earth and in an orbit that was inclined to Earth's equatorial plane by at least 10° .

Together, these traits pose a challenge to theories of lunar cosmogony. Prior to the Apollo era, three origin hypotheses dominated thinking: capture, fission, and coformation (e.g., Wood 1986). Each of these hypotheses failed to account for one or more of the major characteristics of the Earth-Moon system. Capturing an independently formed Moon into an Earth-bound orbit did not offer a natural explanation for the lunar iron depletion, and it appeared dynamically unlikely. In fission, a rapidly spinning Earth becomes rotationally unstable, causing lunar material to be flung out from its equator; this scenario required the Earth-Moon angular momentum to be several times higher than its actual value. Perhaps the most plausible of the three was coformation, which supposed that the Moon grew in Earth orbit from the sweep-up of smaller material from the circumsolar disk, as is believed to be the case for the outer planet satellites. Although coaccretion models were successful in producing satellites, they struggled to explain both the lunar iron deficiency and the Earth-Moon angular momentum, as growth via the accretion of small material produced typically slow planetary rotation rates (e.g., Dones & Tremaine 1993, Lissauer & Kary 1991) unless the material originated preferentially from the outer edges of a planet's feeding zone (e.g., Greenberg et al. 1997, Ohtsuki & Ida 1998).

In the mid-1970s, two independent groups contemporaneously proposed an alternative model. Hartmann & Davis (1975) suggested that the impact of a satellite-sized object with the early Earth ejected iron-depleted mantle material into bound orbit, from which the Moon then formed. Guided by results of planetary accretion models at the time, they considered a roughly lunar-sized impactor, and hypothesized that a moon produced through such an energetic collision would be depleted in volatile elements relative to Earth. Cameron & Ward (1976) recognized similar advantages to a lunar origin via impact, but in addition, realized that the grazing collision of a much larger impactor than that suggested by Hartmann & Davis (1975)—one roughly Mars-sized, or containing $\geq 10\%$ of Earth's mass—could account for Earth's rapid initial rotation. They also suggested that impact-induced vaporization might provide a physical mechanism to allow ejected material to avoid reimpact with Earth and be placed into bound orbit. The concepts described in these works contain the basic elements of the so-called giant impact theory of lunar origin.

In the decade subsequent to its conception, the impact hypothesis was largely ignored, and even some of its originators did not actively pursue its study. It took a topically focused conference on lunar origin in Kona, Hawaii, in 1984 to force a side-by-side appraisal of the existing theories for lunar formation; from this conference the impact theory emerged as the surprise favorite, as much as a result of problems with all the other ideas as to its own still-untested strengths.

In the past decade, nearly the entirety of work on lunar origin has focused on the impact theory across all fields of what is a highly interdisciplinary problem (e.g., the 2000 volume, *Origin of the Earth and Moon*). This review focuses on dynamical aspects of an impact-induced lunar formation, in particular those areas that have advanced considerably since the prior Annual Review on this topic (Stevenson

1987), including:

1. *Late stage terrestrial accretion.* Direct dynamical simulation of the accretion of tens to hundreds of planetary embryos into final terrestrial planets is now possible. Giant impacts capable of providing the Earth-Moon system's angular momentum appear common, occurring within a time frame generally consistent with isotopic constraints.
2. *Modeling lunar-forming impacts.* Hydrocode and equation-of-state developments, in combination with vast increases in computational speed, now allow for more physically realistic impact simulations with orders-of-magnitude higher resolution than those first performed in the mid-1980s. Whereas models in the 1990s focused on progressively larger impactors, the most recent works have returned to the smaller, \sim Mars-mass impactor suggested by Cameron & Ward (1976) as that most likely to yield the Earth-Moon system.
3. *Disk evolution and lunar accretion.* The first generation of models has been developed that track the collisional accumulation of impact-generated material into a moon or moons. These suggest the most likely accretional outcome is a single moon just exterior to Earth's Roche limit. The timing of this formation relative to the impact event itself, as well as the evolution of Roche-interior orbiting material, remains less well understood.
4. *Early lunar orbit evolution.* Goldreich (1966) first recognized that the $> 10^\circ$ initial lunar inclination implied by the current lunar orbit poses a problem for any model postulating the Moon's formation from a circumterrestrial disk, which would naturally tend to produce a satellite orbiting in Earth's equatorial plane. Solutions to the so-called inclination problem have now been proposed.

In all, these developments now provide a consistent dynamical account of the origin of the Moon through a late giant impact with Earth. For each critical element of the problem, solutions exist that appear reasonable and noncontradictory, although not all are unique. Whereas the general strengths of the impact theory lauded at the 1984 Kona conference still appear intact in the face of nearly two decades of modeling, key questions linger, particularly in the reconciliation of dynamical model predictions with certain geochemical and physical signatures of Earth and the Moon.

2. CONSTRAINTS

The dynamical and bulk compositional properties of Earth and the Moon provide key constraints with which any lunar formation model must be reconciled. Those constraints most relevant to this review are summarized below (see also Wood 1986).

2.1. Mass and Angular Momentum

The Moon, with mass $M_L = 7.35 \times 10^{25}$ g, contains 1.2% percent of Earth's mass, $M_\oplus = 5.98 \times 10^{27}$ g. The Earth-Moon angular momentum—contained in Earth's spin and the lunar orbit—totals $L_{\oplus-M} = 3.5 \times 10^{41}$ g-cm²/s. This quantity has likely decreased somewhat over the Earth-Moon system's history because of tidal (e.g., Goldreich 1966, Touma & Wisdom 1994) or resonant (Kaula & Yoder 1976) interactions with the Sun, and/or the accretion of small material onto Earth and the Moon after the Moon-forming event (Morishima & Watanabe 2001). Although the possibility of a later large impact capable of altering the system angular momentum cannot be ruled out, requiring such a secondary event to account for $L_{\oplus-M}$ makes the impact theory more restrictive. In the simplest case, the Moon-forming impact produces an Earth-Moon system with $1.0 L_{\oplus-M} < L < 1.2 L_{\oplus-M}$.

2.2. Lunar Orbital Evolution

The Moon has an orbital semimajor axis of $a \approx 60 R_\oplus$, where $R_\oplus = 6378$ km is Earth's radius. Tidal interaction has caused the lunar orbit to expand and Earth's rotation to slow, implying that the Moon formed close to a then more rapidly rotating Earth. A minimum lunar formation distance is Earth's Roche limit, $a_R = 2.456(\rho_\oplus/\rho_L)^{1/3} R_\oplus$, with $a_R = 2.90 R_\oplus$ for lunar density material; accretion interior to this would be frustrated by planetary tidal forces. A lunar-mass satellite on a circular orbit with $a = a_R$ contains an orbital angular momentum $M_L \sqrt{GM_\oplus a_R} \approx 0.18 L_{\oplus-M}$, suggesting that a successful impact needs to partition at least $\sim 20\%$ of its angular momentum into orbiting protolunar material.

2.3. Lunar Iron Depletion

The $\rho_L = 3.34$ g/cm³ lunar density is so low that it can only be accounted for by a depletion of typically abundant high-density iron. There are two potential reservoirs for lunar iron: a core containing metallic Fe and the silicate mantle/crust containing FeO. Seismic and gravitational analyses suggest the presence of a small lunar core containing 0.01 to 0.03 M_L (e.g., Hood & Zuber 2000). Lunar composition models predict 8 to 10% Fe by mass in the silicate Moon (e.g., Jones & Delano 1989, Jones & Palme 2000, Wood 1986), similar to that measured in lower lunar crustal materials exposed in the South Pole-Aitken basin (~ 7 to 8% by weight; Lucey et al. 1995). Although there are considerable uncertainties, the overall lunar iron mass fraction is likely to be no greater than 10% so that a successful impact must yield an iron depletion of at least a factor of three in protolunar material versus that of bulk terrestrial composition ($\sim 30\%$ iron by mass).

2.4. Lunar Magma Ocean

The anorthositic composition of the lunar crust suggests that low-density silicate mineral classes floated to the top of an upper molten layer on the Moon of several hundred kilometers in depth (e.g., Snyder et al. 2000, Walker & Hays 1997, Wood

et al. 1970). Sufficient heat to account for this so-called lunar magma ocean must be provided as part of a lunar origin scenario. However, a lack of observed stress features on the lunar surface provides a limit on the degree to which an initially molten Moon could have contracted in radius as it cooled. A standard interpretation, which has been re-examined of late (Pritchard & Stevenson 2000), is that the Moon's interior was initially cold (i.e., subsolidus) beneath the magma ocean, leading to a contraction in radius of ≤ 1 km over the Moon's history (e.g., Solomon 1986).

2.5. Comparative Compositions of Earth and the Moon

There are a variety of compositional relationships between Earth's mantle and the Moon; most imply similarities rather than differences. Terrestrial and lunar oxygen isotope abundances lie on the same fractionation line, which is distinct from those of Mars and most sampled meteorites (e.g., Weichert et al. 2001). The standard interpretation is that differences in O-isotope composition resulted from compositional zoning in the original protoplanetary disk; the implication is then that the Moon and Earth formed from material that originated within a common radial provenance. Similarly, Earth and the Moon display a similar chromium isotopic composition, also suggestive of a common origin provenance (e.g., Shukolyukov & Lugmair 2000). The depletion pattern of siderophile elements, or "iron-loving" elements, has similarities in both objects, save an additional depletion in the Moon believed associated with the formation of its small core (e.g., Righter 2002). This suggests that the Moon formed either from terrestrial mantle material or from the mantle of a differentiated planet-sized object. Key differences between the Moon and Earth's mantle include a relative depletion in lunar material of volatiles with vaporization temperatures < 1300 K, and a higher lunar FeO content (e.g., Jones & Palme 2000).

3. THE IMPACT PHASE OF TERRESTRIAL ACCRETION

In the so-called planetesimal hypothesis, the terrestrial planets grew from initially small particles in the protoplanetary disk through collisional accumulation, or accretion. It is believed that the final stages of such growth involved tens to perhaps hundreds of planetary embryos in the terrestrial zone accreting into a final few planets over a timescale of 10^7 to 10^8 years. In the final stage, mutual gravitational interactions among the embryos led to eccentricity excitation, orbit crossing, and large collisions; it is during this phase that the Moon-forming impact is believed to have occurred.

3.1. Late Stage Accretion

Models of the accretion of a system of planetary embryos into terrestrial planets were pioneered by Wetherill (e.g., 1985, 1992), who utilized a statistical Monte

Carlo approach for tracking embryo orbits. Such models demonstrated that the growing Earth was likely to undergo large impacts by objects containing up to \sim the mass of Mars.

Late stage accretion models have been revolutionized by the advent of symplectic mapping methods for the Kepler problem (Wisdom & Holman 1991), which allow for direct N-body integrations for $>10^8$ orbits while maintaining energy conservation. Expansions to the Wisdom-Holman map (WHM) and alternative implementations (e.g., Chambers 1999; Duncan, Levison & Lee 1998) allow the dynamical and collisional evolution of a system of embryos to be tracked explicitly, including effects of mutual resonances and correlated encounters that could only be approximated with prior statistical methods.

Chambers & Wetherill (1998) applied a modified WHM method to initial systems containing a few tens of embryos, whose evolution was tracked for $\sim 10^8$ years with the assumption that all collisions resulted in accretion. Their final systems were not vastly dissimilar from the terrestrial planets. They tended, however, to contain too few planets, and their time-averaged eccentricities and inclinations were higher than those of Earth or Venus. These two characteristics are closely related, as more eccentric orbits require wider radial spacings for dynamical stability, implying a smaller number of final planets. Similarly high planetary e 's and I 's were also obtained by Agnor, Canup & Levison (1999) using a different numerical method and initial conditions.

A relevant diagnostic quantity is the angular momentum deficit (AMD), which is the difference between the component of an object's orbital angular momentum normal to the reference plane and that of a circular, noninclined orbit with the same semimajor axis (e.g., Laskar 1997). A system of N planets has

$$\text{AMD} \equiv \sum_{k=1}^N m_k \sqrt{G M_S a_k} \left(1 - \sqrt{(1 - e_k^2)} \cos I_k \right), \quad (1)$$

where m_k , a_k , e_k , and I_k are the planet masses, semimajor axes, eccentricities, and inclinations, and M_S is the mass of the central star. The planetary systems produced by late-stage N-body simulations have AMD values typically a few to 20 times higher than that of the terrestrial planets (Canup & Agnor 2000, Chambers 2001). This is likely a result of simplifications made in most of the models to date, namely ignoring the influence of potential coexisting small objects or a remnant gas nebula, as both would act generally to decrease eccentricities and inclinations through dynamical friction (e.g., Stewart & Wetherill 1988) and/or density wave damping (Agnor & Ward 2002). Recent simulations including more initial objects (Chambers 2001) or a small remnant of the solar nebula (Kominami & Ida 2002) find systems with closer to Earth-like orbits and a greater likelihood of four final planets, although this remains an open issue.

Given current models, what is the expected likelihood and timing of giant impacts? Agnor, Canup & Levison (1999) analyzed the collisions in their terrestrial accretion simulations and found an average of one to two impacts with angular

momenta $\gtrsim L_{\oplus-M}$ —occurring predominantly in the 10^7 - to 10^8 -year time interval—with an average impact velocity just over the mutual escape velocity of the colliding objects. Earth analogs in these simulations experienced their largest collision at an average time of 29 Myr (ranging from 1.4 to 95 Myr, with a median of 17 Myr), and the end of their accretion at an average time of 46 Myr (ranging from 3.1 to 108 Myr, with a median of 40 Myr). Final planetary obliquities were consistent with random impact orientation. Similar results were found in simulations by Chambers (2001), who considered a larger number of initial embryos (~ 150), and found median accretion times for Earth-analogs to obtain 50% and 90% of their final mass of ~ 20 and 54 Myr, respectively.

Giant impacts therefore appear to have been common in the final stages of terrestrial accretion, consistent with earlier results of Wetherill. Earth-sized planets require an average of 20 to 50 Myr to complete the bulk of their accretion, with a significant spread in this timescale between individual simulations. These accretion times could potentially be influenced by model deficiencies believed associated with overly large final planetary system AMDs. For example, simulations by Kominami & Ida (2002) that include a gas disk containing $\sim 10^{-4}$ to 10^{-3} times that of the minimum mass nebula yield appropriately small final planet eccentricities, but predict an earlier era of large impacts, e.g., between 1 and 10 Myr.

3.2. Timing Constraints from the Hf-W Chronometer

Ideally, theoretical models are reconciled with physical evidence from planetary objects themselves. One key breakthrough in this regard has been the introduction of the use of the hafnium-tungsten chronometer for dating planetary core formation (e.g., review by Halliday, Lee & Jacobsen 2000). Radioactive ^{182}Hf decays to ^{182}W with a half-life of $\tau_{1/2} = 9$ Myr; both Hf and W are refractory, and so unlike volatile elements, their abundances are expected to be in solar system proportions in planetary objects as a whole. A critical distinction between them is that hafnium is lithophile (“silicate-loving,” i.e., tending to be concentrated in oxygen-containing compounds such as silicates), whereas tungsten is siderophile (“iron-loving,” or tending to enter metallic phases). During core formation, whatever tungsten is present in the mantle at that time (including radiogenic ^{182}W as well as nonradiogenic W-isotopes, e.g., ^{183}W or ^{184}W) is largely removed from the mantle and incorporated into the core, while hafnium remains in the mantle. Thus the mantle of a differentiated object will have an Hf/W ratio larger than that of bulk solar system composition; the latter is inferred from the composition of primitive chondritic meteorites.

The Hf/W and W-isotope compositions of a differentiated object provide timing constraints on core formation, with the derived age implying an average age for the core. The Hf/W chronometer can also constrain the timing of the lunar-forming impact, if this event represented the last major episode of core-mantle equilibration in Earth. The tungsten composition of chondrites includes both nonradiogenic isotopes and ^{182}W produced by the decay of primordial ^{182}Hf , and the chondritic

W-isotope composition provides a reference value believed indicative of early solar system abundances. If a planet's core formed on a timescale $\leq 5\tau_{1/2}$, its mantle, compared with chondrites, would contain excess ^{182}W (relative to other W isotopes) produced by decay of ^{182}Hf after core formation. If the core formed later when ^{182}Hf was essentially extinct, all isotopes of W would have been depleted equally by incorporation into the core, leaving the mantle with a chondritic W-isotope composition.

Early measurements showed that Earth's mantle and chondrites had identical W isotopic compositions, implying protracted accretion and the last significant episode of metal equilibration with the terrestrial mantle occurring > 50 Myr after the birth of the solar system (e.g., Halliday 2000, Lee & Halliday 1995). However, Hf-W timescales depend on both the initial solar system $^{182}\text{Hf}/^{180}\text{Hf}$ abundance and the chondritic W-isotope composition, and both of these quantities have been recently revised (Kleine et al. 2002, Yin et al. 2002). The key new result from these works is that the silicate Earth is radiogenic (i.e., contains an excess of ^{182}W) compared with the revised chondritic W isotopic composition; this implies that terrestrial core formation occurred on a timescale of $< 5\tau_{1/2}$ of ^{182}Hf .

Assuming continuous terrestrial core formation and that accreting material equilibrates isotopically with the mantle, the revised parameters imply Earth's accretion and core formation were completed in 10 to 30 Myr (Kleine et al. 2002, Yin et al. 2002). The lower end of this range is the time for Earth to accrete 63% of its final mass assuming an exponentially decaying accretion rate, whereas the longer time applies to a model in which a final event (e.g., the Moon-forming impact) leads to core-mantle equilibration with no subsequent accretion (Yin et al. 2002).

In the simplest interpretation, a Moon formed by an impact at the end of Earth's formation would have an Hf-W age similar to that of Earth, while a shorter lunar formation time would imply that the impact occurred earlier in Earth's accretion. Initial work that found radiogenic ^{182}W in lunar samples in excess of that in Earth (Lee et al. 1997) appeared in some interpretations to suggest that the Moon was somewhat older than Earth (e.g., Halliday, Lee & Jacobsen 2000; Jacobsen 1999). However, it was later shown that a major portion of the lunar ^{182}W excess could be explained through non-Hf decay processes (e.g., cosmic ray bombardment, Lee et al. 2002; Leya, Wieler & Halliday 2000). Recent works assuming an approximately terrestrial Hf/W ratio derive a lunar formation time ~ 25 to 30 Myr after the formation of the solar system, consistent with a lunar-forming impact near the end of Earth's accretion (Kleine et al. 2002, Yin et al. 2002).

If the equilibration of accreting material is incomplete, the above Hf-W timings can be lengthened (Halliday 2004), and this issue is an active area of research. But, it is encouraging that the Hf-W timings and the late-stage dynamical models both yield estimates in the 10 to 50 Myr time interval for planetary accretion, giant impacts, and the final episodes of core formation. This agreement provides significant weight to the existence of a late, impact-dominated accretionary phase. If the terrestrial planets had grown to their final sizes through the accumulation of

small material and runaway growth (e.g., Weidenschilling et al. 1997), their final accretion times would likely have been much shorter, $\sim 10^6$ years or less.

4. IMPACT SIMULATIONS

What type of impact can produce the Earth-Moon system, and what are the expected consequences for Earth and protolunar material? The impact theory proposes that the same impact that created the Moon was also the primary source of the Earth-Moon system angular momentum, $L_{\oplus-M}$. Although there could have been earlier or later impacts that contributed significantly, this is the least complex case and a reasonable starting assumption.

The angular momentum delivered by an impactor of mass $M_{imp} \equiv \gamma M_T$ is

$$L_{imp} = b M_T^{5/3} f(\gamma) \sqrt{\frac{2G}{(4\pi\rho/3)^{1/3}}} \left(\frac{v_{imp}}{v_{esc}} \right) \\ \approx 1.3 L_{\oplus-M} b \left(\frac{M_T}{M_{\oplus}} \right)^{5/3} \left(\frac{\gamma}{0.1} \right) \left(\frac{v_{imp}}{v_{esc}} \right), \quad (2)$$

where $b \equiv \sin \xi$ is the impact parameter normalized to the sum of the impactor and target radii (R_{imp} and R_{tar}), ξ is the angle between the surface normal and the impact trajectory (so that a grazing impact has $b = 1$ and $\xi = 90^\circ$), M_T is the total colliding mass (impactor + target), γ is the impactor-to-total mass ratio, $f(\gamma) \equiv \gamma(1-\gamma)\sqrt{\gamma^{1/3} + (1-\gamma)^{1/3}}$, ρ is the density of the colliding objects, and (v_{imp}/v_{esc}) is the ratio of the impact velocity to the mutual escape velocity, with $v_{esc} \equiv \sqrt{2GM_T/(R_{imp} + R_{tar})}$. Equation 2 implies a minimum $M_{imp} \sim 0.08M_{\oplus}$ required to yield $L_{\oplus-M}$ with a single grazing impact with $M_T \approx M_{\oplus}$ and $v_{imp} \approx v_{esc}$. Larger impact speeds would result for objects with an appreciable relative velocity at large separation, v_{∞} , with $v_{imp}^2 \equiv v_{esc}^2 + v_{\infty}^2$.

Thus, even for the simplest case in which a single impact supplies the Earth-Moon system angular momentum, a multidimensional parameter space in b , M_T , γ , and v_{imp} of candidate impacts is implicated that could all provide $L_{\oplus-M}$. The goal is then to determine what subset could also produce an appropriately massive and iron-depleted Moon.

In order for impact-ejected material to achieve bound Earth orbit, some modification to standard ballistic trajectories must occur since, in the latter case, ejecta launched from the planet's surface either reimpacts or escapes (e.g., Cameron & Ward 1976, Stevenson 1987). Two nonballistic effects are gravitational torques (resulting from either mutual interactions among ejected material or interaction with a nonspherical distortion of the target planet) and pressure gradients associated with vaporization. These effects become important for large impacts: the first when the impactor is a significant fraction of the target's mass, and the second when the specific impact energy exceeds the latent heat of vaporization for rock, $L_v \sim 10^{11}$ erg/g, i.e., for $v_{imp} \gtrsim 5$ km/s.

For a lunar-forming impact, $v_{imp} \gtrsim 10$ km/s ($v_{esc, \oplus} = 11.2$ km/s), and both torques and vaporization could be important, necessitating a full hydrodynamic approach including an equation of state appropriate to describe the thermodynamic response of material subjected to very high impact energies and pressure, as well as an explicit treatment of self-gravity.

4.1. Method

Models of potential Moon-forming impacts have primarily utilized smooth particle hydrodynamics, or SPH (e.g., Benz, Slattery & Cameron 1986; Lucy 1977; Monaghan 1992). SPH requires no underlying grid and so is well suited to treat deforming systems evolving within mostly empty space; its Lagrangian structure also allows for easy tracking of material and compositional histories. In SPH, an object is represented by a great number of spherical overlapping particles, each containing a quantity of mass of a given composition, whose three-dimensional spatial distribution is specified by a density weighting function (the kernel) and the characteristic radial scale of the particle, the smoothing length, h . In modern applications, the smoothing length of each particle is adjusted so as to maintain an overlap with a minimum number of other particles. For the impacts of interest here, the evolution of each particle's kinematic (position and velocity) and state (internal energy, density) variables are evolved due to (a) gravity, (b) compressional heating and expansional cooling, and (c) shock dissipation; material strength and radiative processes are ignored.

The equation of state relates a particle's specific internal energy and local density to a pressure as a function of input material constants, with density typically determined by a sum of contributions from the particle itself and its overlapping "neighbor" particles. Equations of state range from simple analytic approximations (e.g., Tillotson 1962) to complex semianalytical methods affording a more rigorous thermodynamic treatment [e.g., Analytic Equation of State (ANEOS) (Thompson & Lauson 1972)].

A critical element in the accuracy of SPH results is numerical resolution. SPH is an interpolative method, and so is only truly meaningful in regions containing many overlapping particles. In the first works modeling giant impacts (Section 4.2, below), $N = 3000$ particle simulations described a lunar mass (and thus the typical amount of orbiting material) by only a few tens of particles. In addition, an early version of SPH was utilized in which particle smoothing lengths were fixed, so that ejected particles could become isolated from one another. Once this occurred, particles were no longer subject to surface pressure forces, and instead they evolved purely gravitationally (Cameron 1997).

More recent simulations (Section 4.3, below) have up to $N = 10^5$ particles and time-varying smoothing lengths, which ensures that even low-density regions do not suffer from particle isolation. Larger numbers of particles allow in particular for an improved description of the orbiting iron, which for a total lunar iron mass $< 0.1 M_L$ contains $< 0.1\%$ of the total mass in a collision, and therefore a

similarly small proportion of SPH particles. Resolving the orbiting iron was nearly impossible with $N = 3000$ simulations, and remains challenging even with modern resolutions. While the factor of ~ 40 increase in N achieved in the past decade is quite substantial, it implies a linear resolution increase of only about a factor of 3.4, as for a 3-D simulation, $h \propto N^{1/3}$. With $N = 10^5$, initial particles in the protoearth mantle have $h \sim 300$ km, and processes occurring on smaller scales are therefore still not resolved.

4.2. Low-Resolution Models (1986–1991)

The use of SPH in modeling lunar forming impacts was pioneered by Willy Benz, Alastair Cameron and colleagues (Benz, Slattery & Cameron 1986, 1987; Benz, Cameron & Melosh 1989; Cameron & Benz 1991). Benz, Slattery & Cameron (1986) provided the first true hydrodynamical demonstration of the emplacement of orbiting material through a giant impact, assuming pure granite compositions. Benz, Slattery & Cameron (1987) performed nine impact simulations in which the colliding objects were assigned an initial core-mantle structure, involving several different impactor sizes ($0.09 \leq \gamma \leq 0.2$) and $M_T \approx M_\oplus$. Each collision was tracked for ~ 10 to 25 h; at this point the orbiting mass—defined as those particles having periapses above the surface of the protoearth but lacking sufficient energy to escape—ranged from $\sim 0.1 M_L$ to $2 M_L$ for impacts with $1.06 \leq L_{\text{imp}}/L_{\oplus-M} \leq 1.35$. The orbiting material originated overwhelmingly (84 to 92% by mass) from the impactor. It was concluded that low-mass impactors (with $\gamma < 0.12$) produced overly iron-rich disks, while impacts by somewhat larger impactors colliding with $v_{\text{imp}} \approx v_{\text{esc}}$ produced more favorable results.

For their most successful impacts, an initial off-center collision partially destroyed the impactor, a portion of which was sheared out into a long “arm” of material as it grazed past the target. Within this arm, the impactor core remained more central to the planet and recollided with the protoearth, while some of the more exterior impactor mantle material was left in bound orbit. Benz et al. proposed that gravitational torques, rather than pressure gradients, were primarily responsible for orbital injection. In one of their simulations, a single intact “clump” of impactor mantle containing $\sim 1 M_L$ was left with a stable, Roche-exterior orbit. Two possible modes of forming the Moon were then suggested: from a protolunar disk or through direct formation by the impact event.

There were a number of issues raised by the Benz, Slattery & Cameron (1987) results (e.g., discussion in Stevenson 1987). The low resolution made it difficult to interpret the meaning of an intact “moon” containing only a few dozen SPH particles. The Tillotson equation of state does not actually model phase changes or vaporization, instead it simply derives a particle’s pressure from a linear interpolation between that appropriate to a solid and that of an ideal gas. Thus, latent heat or the potential for mixed liquid-vapor phases is not accounted for, compromising the ability to distinguish the relative importance of gravitational torques versus pressure gradients. This concern was reinforced by independent simulations of the

initial stages of a giant impact using an Eulerian hydrocode with a finer resolution scale, which indicated a greater degree of vaporization (Kipp & Melosh 1987, Melosh & Kipp 1989). Today such issues continue to challenge impact models. However, it is interesting to note that many of the conclusions of Benz, Slattery & Cameron (1987) are qualitatively similar to those obtained by recent works using more sophisticated techniques to model similar types of impacts (Canup 2004, Canup & Asphaug 2001). A key distinction was the Benz et al. conclusion that small impactors produced disks that were too iron-rich. Although this was later shown to be largely an effect of poor resolution (Canup & Asphaug 2001), it influenced impact modeling for more than a decade, with works generally focusing on progressively larger impactors.

Benz, Cameron & Melosh (1989) incorporated an equation of state known as ANEOS (Thompson & Lauson 1972) into their SPH simulations in an attempt to better characterize vaporization and mixed-phase states. In ANEOS, thermodynamic quantities are derived from the Helmholtz free energy, with temperature and density as independent variables. ANEOS describes a mixed vapor-fluid state within a single SPH particle by assuming the phases are in temperature and pressure equilibrium, so that the mass fraction and pressure contribution from each phase is calculated.

Using ANEOS, Benz, Cameron & Melosh (1989) found smaller yields of orbiting material and more centralized disks compared with similar impacts modeled with Tillotson.¹ In an effort to increase the yield of orbiting material, Cameron & Benz (1991) used ANEOS to simulate collisions by larger impactors ($\gamma = 0.12, 0.14, \text{ and } 0.2$) with higher impact angular momenta of $1.13 \leq L_{\text{imp}}/L_{\oplus-M} \leq 1.93$. The larger impactors with the highest angular momenta produced disks containing up to $2.6 M_L$ for the iron-free cases.

4.3. High-Resolution Models (1997–2003)

In Cameron (1997), a now standard version of SPH was incorporated in which smoothing lengths are adjusted automatically to ensure overlap and prevent particle isolation. Collisions involving still larger impactors ($\gamma = 0.2$ to 0.5 , with the latter corresponding to the impact of like-sized objects), with impact angular momenta $1.2 \leq L_{\text{imp}}/L_{\oplus-M} \leq 2.2$ and $M_T \approx M_{\oplus}$ were modeled with increased resolution ($N = 10^4$). Again, as in Cameron & Benz (1991), the most massive disks were produced by high angular momentum impacts with $L_{\text{imp}}/L_{\oplus-M} \sim 2$.

Cameron then proposed a new candidate impact: one in which a relatively large impactor ($\gamma = 0.3$) collided with a similar impact parameter and velocity as in the most successful cases of Cameron & Benz (1991) and Cameron (1997),

¹This is not too surprising because with ANEOS the latent heat of vaporization serves as an additional energy sink, and it has a similar magnitude to both the specific impact energy per unit projectile mass (with $E_I \approx (4\pi\rho/3)^{1/3} G M_T^{2/3} / [\gamma^{1/3} + (1 - \gamma)^{1/3}]$ for $v_{\text{imp}} = v_{\text{esc}}$ impacts, or $E_I \approx 4 \times 10^{11} \text{ erg/g}$ for $\gamma = 0.12$ and $M_T \approx M_{\oplus}$), and the specific energy difference between an orbit with $a = 1.5 R_{\oplus}$ and Earth's surface, $\sim 10^{11} \text{ erg/g}$.

but with a reduced total mass (impactor + target). By reducing M_T , the impact angular momentum could be reduced to $\sim L_{\oplus-M}$; physically, this would imply a lunar-forming impact occurring much earlier in Earth's accretion.² The so-called early-Earth impact was investigated in $10^4 < N \leq 10^5$ simulations with $\gamma = 0.3$, and $0.55 M_{\oplus} \leq M_T \leq M_{\oplus}$ (reviewed in Cameron 2000, 2001). Of these, a collision involving an impactor containing \sim twice the mass of Mars and $M_T \sim 0.65 M_{\oplus}$ produced the best results: a disk containing 1.5 to 2 lunar masses for $L_{imp} \sim 1.1$ to $1.2 L_{\oplus-M}$.

Thus a large impactor ($\gamma = 0.3$) was shown with high resolution to produce an appropriate protolunar disk when either $L_{imp}/L_{\oplus-M} \sim 2$ and $M_T \sim M_{\oplus}$, or $L_{imp}/L_{\oplus-M} \sim 1$ and $M_T \sim 0.65 M_{\oplus}$, with both cases more restrictive than the original single-impact hypothesis. A "high angular momentum" impact required a significant subsequent dynamical event, e.g., another giant impact, to decrease the Earth-Moon system angular momentum to its current value. In an early-Earth impact, Earth gains $\sim 0.35 M_{\oplus}$ after the Moon forms, with the later growth involving sufficiently small and numerous impacts so that the system angular momentum is not drastically altered. One difficulty is the potential for the Moon to become contaminated by iron-rich material during the time Earth was accumulating the final $\sim 35\%$ of its mass (e.g., Stewart 2000). If the Moon accumulated an amount of material approximately proportionate to its physical cross section, Earth could only accrete $\sim 0.06 M_{\oplus}$ before the Moon gained $> 0.1 M_L$ in iron. A reduced lunar accretion efficiency relative to that of Earth (e.g., Morishima & Watanabe 2001) might mitigate this inconsistency; however, in that case it could be difficult to avoid a divergence in the O-isotope compositions of Earth and the Moon. Thus, in general, as the amount of material that must be added to Earth after the Moon-forming impact in a given impact scenario increases, difficulties in accounting for the Moon's composition also increase.

Canup, Ward & Cameron (2001) reexamined results of Cameron (2000), and identified scaling trends consistent across all of the simulations. The orbiting mass, angular momentum, and iron all generally increased with b for $0.4 < b < 0.8$. In particular, it was shown that for $\gamma = 0.3$, the maximum yield of orbiting material occurred for an impact parameter $0.7 < b < 0.8$, independent of the total colliding mass, M_T .

Utilizing this result, Canup & Asphaug (2001) predicted that the maximum orbiting yield for an impact in which both $L_{imp} \approx L_{\oplus-M}$ and $M_T \approx M_{\oplus}$ would also be achieved for $b \approx 0.7$ to 0.8 ; this gave an estimate for the optimal impactor-to-total-mass ratio of $\gamma \sim 0.1$, or an approximately Mars-sized object. Thus, Canup and Asphaug focused once again on small impactors, which had not been modeled

²An independent motivation (e.g., Cameron 2001) for such an impact at the time was the lunar ^{182}W excess, which seemed in some models to suggest that the Moon formed before the end of Earth's accretion (e.g., Halliday 2000; Halliday, Lee & Jacobsen 2000; Jacobsen 1999); such an age difference does not appear necessary given the most recent Hf-W timings (Yin et al. 2002; Section 3.2).

since Benz, Slattery & Cameron (1987). Whereas Benz et al. had ruled out small impactors as producing overly iron-rich disks (defined then by only three or four orbiting iron particles), with increased resolution Canup & Asphaug (2001) showed that many such impacts produced massive disks with less than a few percent iron by mass, together with $M_T \approx M_\oplus$ and $L_{imp} \approx L_{\oplus-M}$. For the first time with modern techniques and resolutions, a single impact at the end of Earth's accretion (the late impact case) was shown capable of producing the Earth-Moon system, without requiring substantial subsequent modification. However, Canup & Asphaug (2001) simulations relied on the simple Tillotson equation of state.

Recently, Canup (2004) simulated late impacts with ~ 100 simulations utilizing a newly revised version of ANEOS. In its standard rendition (used in previous giant impact simulations), ANEOS treated all vapor as monatomic species (e.g., Melosh & Pierazzo 1997), and thus the entropy and energy required for vaporization of molecular species—such as mantle rock—was overestimated. As a remedy to this, Melosh (2000) revised ANEOS (M-ANEOS) to allow for the formation of a diatomic molecular vapor (e.g., SiO or MgO for the nominal mantle mineral forsterite, Mg_2SiO_4).

Using M-ANEOS, Canup (2004) considered impacts with $0.11 < \gamma < 0.15$, and showed that the maximum yield of iron-depleted orbiting material occurs consistently for $b \sim 0.7$. For low impact velocities, i.e., with $(v_{imp}/v_{esc}) \leq 1.10$, potential lunar-forming candidates having this impact geometry in common were identified across many impactor sizes. Total disk masses similar to those of Benz, Cameron & Melosh (1989) were obtained for similar impacts. However, the Canup (2004) disks have characteristically higher specific angular momenta, with an average of about 70% of the disk material having equivalent circular orbits (defined as the circular orbit containing the same angular momentum as an eccentric orbiting particle) exterior to a_R for the successful $v_{imp} = v_{esc}$ cases versus 40–50% in the similar Benz, Cameron & Melosh impacts. It is not clear to what extent this is a result of the use of M-ANEOS versus variable particle-smoothing lengths or both.

What then is the overall implication for the types of impacts capable of producing the Moon and the Earth-Moon angular momentum? Figure 1 is a plot of M_T versus γ showing the phase space of impacts having $L_{imp} = 1.25 L_{\oplus-M}$, $0.70 \leq b \leq 0.75$ and $1.00 \leq (v_{imp}/v_{esc}) \leq 1.10$. On the upper left is the late impact regime (Canup 2004, Canup & Asphaug 2001), whereas on the lower right, with $M_T \sim 0.65 M_\oplus$ and $\gamma = 0.3$ is the early-Earth impact (Cameron 2000, 2001). Intermediate to these and between the curves is a continuous array of impacts that could all produce a \sim lunar-mass, iron-depleted satellite and $L_{\oplus-M}$. The late impact is distinguished by its ability also to leave an Earth-Moon system with the correct total mass.

4.4. An Example Lunar-Forming Impact

Figure 2 shows an $N = 60,000$ particle simulation from Canup (2004); color scales with temperature, with red indicating particles with $T > 6440$ K. Here $\gamma = 0.13$,

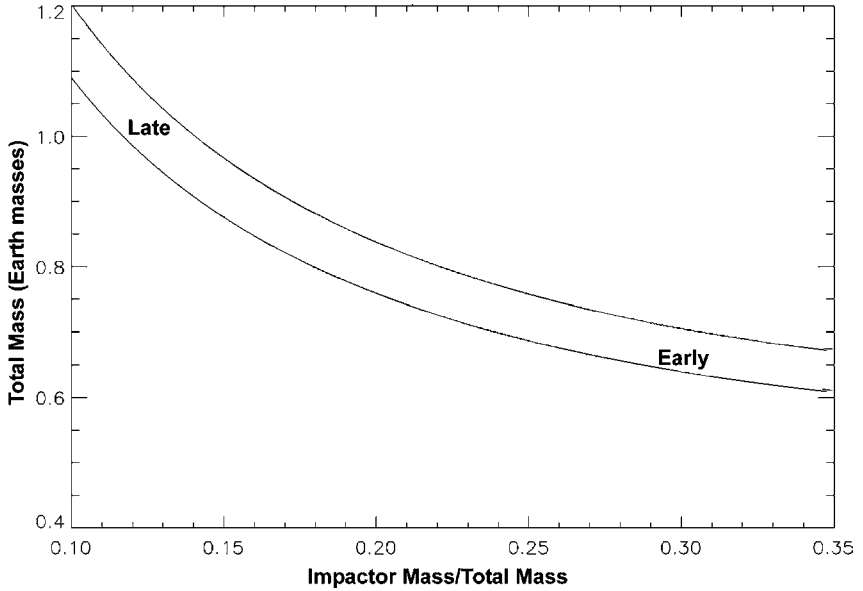


Figure 1 M_T versus γ contours for impacts having $L_{imp} = 1.25 L_{\oplus-M}$, $0.70 \leq b \leq 0.75$, and $1.00 \leq (v_{imp}/v_{esc}) \leq 1.10$. Impacts falling between the two curves are predicted to be good candidates for producing a lunar-mass, iron depleted moon. Regions corresponding to a late impact (Canup 2004, Canup & Asphaug 2001), and an early-Earth impact (Cameron 2000, 2001) are indicated.

$M_T = 1.019 M_{\oplus}$, $L_{imp} = 1.25 L_{\oplus-M}$, $b = 0.73$, and $v_{imp} = v_{esc} \approx 9.3$ km/s. Prior to the impact, both objects are differentiated into iron cores and dunite (forsterite) mantles with 30% iron by mass; initial surface temperatures are ~ 2000 K and increase adiabatically with depth.

After the initial oblique impact (Figure 2a after 6 simulated min), a portion of the impactor that avoided colliding directly with the protoearth is sheared off and continues forward ahead of the impact site (Figure 2b at 20 min). After ~ 50 min, a distorted arm of impactor material extends to a distance of ~ 3 to $3.5 R_{\oplus}$ (Figure 2c). The protoearth surface and the inner portions of the arm rotate ahead of the more distant impactor material, providing a positive torque (Figure 2d at 80 min). After approximately 2 h (Figure 2e), the most distant portions of the impactor, now at $\sim 6 R_{\oplus}$, begin to gravitationally self-contract. An arc of impactor material extends from this distant clump to Earth's surface; within this is the sheared iron core of the impactor. In the 3- to 5-h time frame, the inner portions of the orbiting material (composed primarily of the impactor's core), gravitationally contract into a semicoherent object (Figure 2f) that recollides with the planet after approximately 6 h (Figure 2g). Thus, at this point, most of the impactor's iron has been removed from orbit. The outer clump of the impactor—composed entirely of mantle

material—eventually makes a periaapse pass within the Roche limit and is sheared into a spiral arm (Figure 2*i–j* at 19 to 21 h), which finally breaks up into multiple smaller clumps (Figure 2*k* at 27 h). The last frame (Figure 2*l*) is the system at 27 h viewed on edge, shown with a higher temperature scale (with red now indicating particle temperatures in excess of 9110 K).

At the end of the impact shown in Figure 2, the bound planet-disk system has an angular momentum of $1.18 L_{\oplus-M}$, the planet contains $0.994 M_{\oplus}$, and its rotational day is ~ 4.6 h; a total of $0.41 M_L$ has escaped. The disk is described by 2200 particles, and contains $1.62 M_L$, with $0.92 M_L$ having equivalent circular orbits exterior to the Roche limit; the total orbiting angular momentum is $L_D = 0.31 L_{\oplus-M}$. Of the material with equivalent orbits exterior to a_R , 80% originated in the impactor, 24% is vapor, and 1.9% is iron, whereas for material with equivalent orbits interior to a_R , 85% is from the impactor, 22% is vapor, and 9.1% is iron. Similar vapor fractions are found by estimating the entropy produced during the initial impact shock, and then calculating an isentropic pressure release from this peak shock state (figure 4 in Stevenson 1987).

In another Canup (2004) impact simulation using $\gamma = 0.15$, $M_T = 0.95 M_{\oplus}$, and $b = 0.73$, a disk containing $1.82 M_L$ resulted including a massive single clump containing $0.6 M_L$ (and 1700 SPH particles) on a Roche-exterior orbit. This object originated from the outer portions of the arm of impactor material analogous to that seen in Figure 2*f*; about 25% of the Canup (2004) impacts produced such “moon-disk” systems with the rest resembling the final disk in Figure 2.

Where does the protolunar material originate, and how is it processed during the impact? To address this question, Canup (2004) implemented a particle tracking function that assigns particle color based upon each particle’s final state, i.e., orbiting, escaping, or in the planet. Figure 3*a* shows such a mapping of final particle states from the simulation shown in Figure 2 onto the original figures of the impactor and target. The great majority of the orbiting material originates from a common region on the leading face of the impactor that is just radially exterior to the primary impact interface. A group of escaping impactor particles just below this region is highly heated material that has jetted outward from the impact interface with $v > v_{esc}$ (e.g., Vickery & Melosh 1987). Figure 3*b* shows a similar mapping of the peak temperature change, $\Delta T \equiv T_{max} - T(0)$, experienced by each particle during the same simulation.

Comparison of Figure 3*a* with Figure 3*b* shows that the portion of the impactor that eventually composes the disk is, in general, the least heated of all of the impactor material, having for the most part avoided direct collision with the protoearth. This implies that impact geometry and gravitational torques play the primary role in orbital emplacement, with vaporization of secondary and/or indirect importance. Qualitatively similar mappings to those in Figure 3 were obtained for all of the impacts that produced massive and iron-depleted disks in Canup (2004), with 75 to 90% of the orbiting material originating from the impactor.

For the impact shown in Figure 2, 28% of the protoearth material had $\Delta T > 5000$ K and 30% had $\Delta T < 2000$ K; of the orbiting material, 18% had $\Delta T > 5000$ K

while 24% had $\Delta T < 2000$ K. Figure 4 shows final particle temperatures versus instantaneous radial position for dunite (black) versus iron (gray). In the protoearth, rock temperatures range from 2000 K to 10,000 K, with iron from the impactor reaching much higher temperatures of tens of thousands of degrees Kelvin. The orbiting silicate has temperatures in the 2500 K to 5000 K range; some of the disk iron is significantly hotter, with temperatures in excess of 10,000 K in the inner disk.

4.5. Trends in Impact Outcome

Figure 5 shows results of simulations involving varied N , b , and (v_{imp}/v_{esc}) (Canup 2004). Trends in the orbiting mass, angular momentum, and iron similar to those identified from Cameron's simulations (Canup, Ward & Cameron 2001) result. The maximum yield of iron-depleted material occurs near $b \sim 0.7$, with impacts in this "peak" having a morphology similar to that of Figure 2. For these cases, the secondary impact of the impactor core yields a reduced orbiting iron fraction. More head-on collisions yield less orbiting material because a smaller portion of the impactor shears past the target during the initial impact. For very oblique impacts ($b > 0.8$) the yields of orbiting material can be large but usually contain excessive iron.

Impacts with $(v_{imp}/v_{esc}) = 1$ to 1.10 for a (corresponding to $v_{\infty} = 0$ to 4.3 km/sec) produce broadly similar trends in outcome as a function of b , with $(v_{imp}/v_{esc}) = 1.02$ and 1.05 cases producing slightly higher orbiting yields than $(v_{imp}/v_{esc}) = 1.00$. However, for $v_{imp} \geq 1.1v_{esc}$, an increasing amount of material escapes, lowering the orbiting yield. In addition, the fraction of disk iron tends to increase with (v_{imp}/v_{esc}) for a given b , becoming excessive for $v_{imp} \geq 1.1v_{esc}$.

Requiring $(v_{imp}/v_{esc}) \leq 1.10$ for a Moon-forming impact constrains the giant impactor's precollision orbit. As a simple example, consider a circular protoearth orbit, an eccentric impactor orbit, no orbital inclinations, and an impact that occurs at the impactor's periapse (apoapse) for an exterior (interior) orbiting impactor. The impactor velocity at periapse is $v_p = \sqrt{(GM_S/a_{imp})(1+e_{imp})/(1-e_{imp})}$; requiring $a_{imp}(1-e_{imp}) \approx a_{\oplus}$ and $v_{\infty} = v_p - v_{\oplus} \leq 4$ km/s (where $v_{\oplus} = \sqrt{GM_S/a_{\oplus}} \sim 30$ km/s is Earth's orbital velocity) implies an impactor semimajor axis $a_{imp} \leq 1.4$ AU. An analogous calculation for an interior impactor gives $a_{imp} \geq 0.8$ AU. Although this dynamical constraint does not require the impactor to have formed in such a close orbit to Earth (e.g., it could have been scattered there through interactions with other planetary embryos), this would be consistent with the standard O-isotope interpretation, which implies that the impactor and the protoearth formed within the same region in the protoplanetary disk.

Whereas for a given impactor-to-total mass ratio, γ , the maximum iron-depleted orbiting yield occurs near $b \sim 0.7$, the peak orbiting mass scales approximately linearly with γ , with

$$\left. \frac{M_d}{M_T} \right|_{MAX} \approx 0.015 \left(\frac{\gamma}{0.1} \right) \quad (3)$$

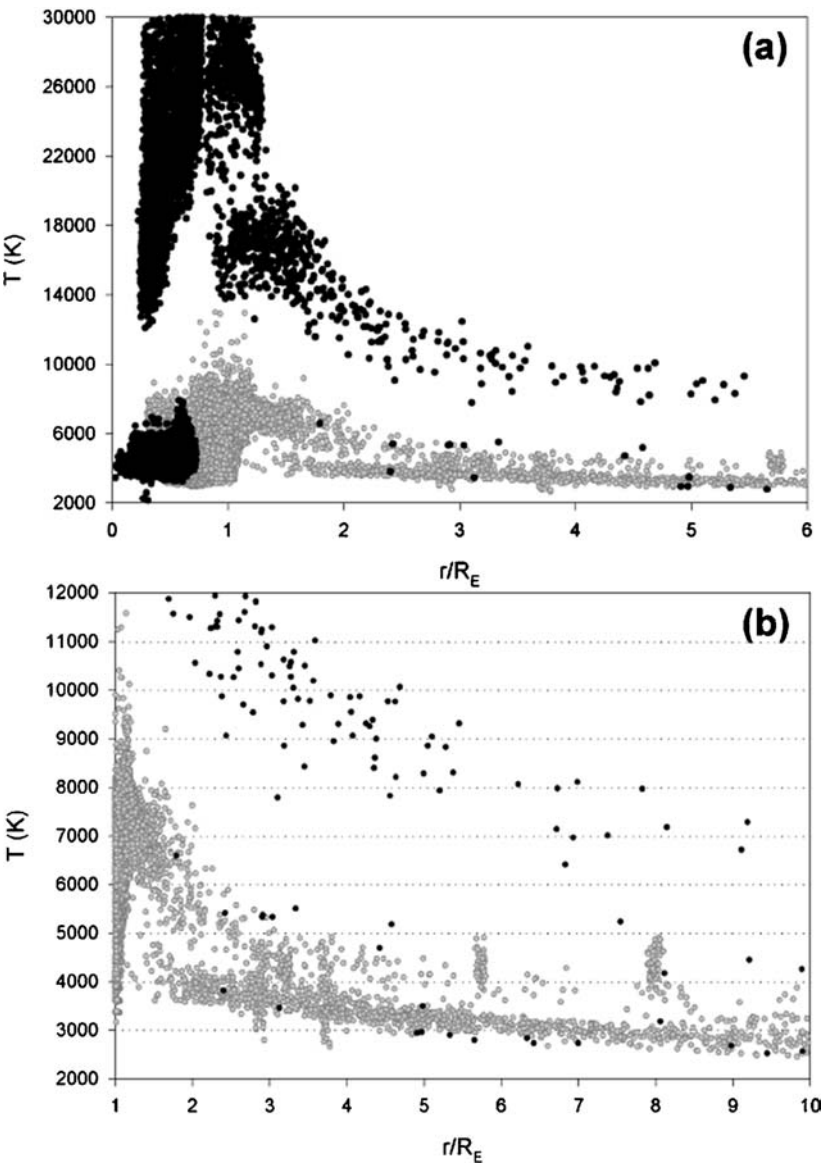


Figure 4 Temperature versus instantaneous radial position in units of Earth radii for all the particles at the end of the simulation in Figure 2. Iron particles are black; silicate particles are gray. Frame *b* shows the same data as in frame *a*, but with a lower temperature scale and with $r \geq R_\oplus$.

for $v_{imp} = v_{esc}$ (Canup 2004). Equation 3 is approximately consistent with previous simulations using ANEOS (original or M-ANEOS), including Cameron's simulations with $\gamma = 0.3$, where $(M_d/M_T)|_{MAX} \sim 0.045$, as well as simulations with $0.12 \leq \gamma \leq 0.2$ reported in Cameron & Benz (1991).

Rewriting Equation 2 with $b = 0.7$ and $v_{imp} = v_{esc}$ as $M_T/M_\oplus \approx 1.2(L_{imp}/1.25 L_{\oplus-M})^{3/5}(0.1/\gamma)^{3/5}$ and combining with Equation 3 gives an analytical estimate of the peak protolunar disk mass for a given γ and M_T ,

$$\left. \frac{M_d}{M_L} \right|_{MAX} \approx 1.46 \left(\frac{L_{imp}}{1.25 L_{\oplus-M}} \right)^{3/5} \left(\frac{\gamma}{0.1} \right)^{2/5}; \quad (4)$$

somewhat higher disk masses result for impact velocities just exceeding v_{esc} . The rather weak dependence of (M_d/M_L) on γ is consistent with impact simulations that have $L_{imp} \sim L_{\oplus-M}$ all producing similar maximum protolunar disk masses ~ 1.5 to $2 M_L$ over the $0.1 \leq \gamma \leq 0.3$ and $0.6 \leq (M_T/M_\oplus) \leq 1$ ranges.

4.6. Discussion

Results from nearly two decades of simulations investigating a multitude of giant impact scenarios have provided a substantial base of knowledge from which certain key results have emerged:

1. *Gravitational torques and impact geometry appear the dominant factors in placing material into orbit.* Even recent simulations utilizing the best available treatment for vapor production do not show a direct relationship between vaporization and orbital emplacement. The implication is that creation of impact-generated satellites can occur for gravity-dominated objects once impactors contain a significant fraction of the target mass, independent of whether impact energies reach those necessary for vaporization.
2. *Giant impacts exhibit consistent trends in dynamical outcome.* For low-velocity impacts, the fractional yield of orbiting mass, angular momentum, and iron all generally increase with impact parameters up to $b \sim 0.8$, with $b \gtrsim 0.5$ impacts generally producing significant quantities of orbiting material. The effect of preimpact spin on impact outcome has yet to be fully assessed.
3. *Properties of a lunar-forming impact.* Candidate Moon-forming impacts typically have $b \sim 0.7$ and a low impact velocity, with $(v_{imp}/v_{esc}) < 1.1$. A range of impactor sizes can produce appropriate protolunar disks as a function of the total mass at the time of the giant impact and the desired system angular momentum. For a late impact with $M_T \sim M_\oplus$ and $L_{imp} \sim L_{\oplus-M}$, $0.1 < (M_{imp}/M_T) < 0.15$ is required, corresponding to an approximately Mars mass impactor and an Earth that accretes $\leq 0.05 M_\oplus$ after the Moon-forming event.
4. *State of protolunar material.* Predicted temperatures of orbiting material are ~ 3000 – 4000 K, with 10–30% vapor by mass; the great majority (75 to 90%) originates from the impactor, and its mean orbital radius is near or just

exterior to a_R . In simulations to date, maximum disk masses range from ~ 1.5 to 2 lunar masses for $L_{\text{imp}} < 1.5 L_{\oplus-M}$. This implies that if the Moon-forming impact is primarily responsible for providing the Earth-Moon angular momentum, subsequent lunar accretion must have been a fairly efficient process.

5. *Clump versus disk?* Intact clumps containing a substantial fraction of a lunar mass have been observed in some cases across all resolutions and equations of state utilized to date. Whether the Moon's composition can be reconciled with direct formation from an impact versus formation from a disk is unclear, as is the effect of the use of SPH (which can be prone to artificial clumping, e.g., Imaeda & Inutsuka 2002) on results to date. In the majority of cases, large clumps are sheared apart as they pass within a_R (e.g., Cameron 2000), yielding a Roche-interior disk together with exterior clumps.

Despite their successes, current SPH simulations are limited in their ability to realistically track the postimpact evolution of the orbiting material. The most stringent constraint arises from the disk's resolution. Even after a postimpact system has settled into a central planet and disk (Figure 2*k*), the orbiting mass continues to decrease with time as particles exchange angular momentum and some are scattered onto the protoearth (e.g., figure 6 in Cameron 2000). With time, this behavior is increasingly influenced by the viscosity associated with disk particles interacting over a radial scale determined by their smoothing lengths, which are $h \sim O(R_{\oplus})$ for a disk containing $\sim O(10^3)$ particles. An estimate of the timescale of this effect is made by relating the SPH artificial viscosity parameter α_{SPH} (part of the code's treatment of shock dissipation) to an equivalent disk alpha-viscosity (e.g., Shakura & Sunyaev 1973) through the relation $\nu_{\text{SPH}} \sim c_s h \alpha_{\text{SPH}} / 8$, where c_s is sound speed (e.g., Nelson et al. 1998). The characteristic timescale for viscous spreading is $\tau \sim r^2 / \nu$, implying for $\nu = \nu_{\text{SPH}}$ and $\alpha_{\text{SPH}} \sim \text{unity}$

$$\tau_{\text{SPH}} \sim 10^2 \text{ hrs } (r/2R_{\oplus})^2 (0.5R_{\oplus}/h) (0.8 \text{ km/sec}/c_s). \quad (5)$$

As discussed in Section 5, τ_{SPH} is orders of magnitude shorter than any known physical viscous timescale. Spurious angular momentum transport in the disk should be minimal so long as the simulated time $\ll \tau_{\text{SPH}}$. The next generation of SPH models may help to mitigate such effects through higher disk resolutions.

5. DISK EVOLUTION AND LUNAR ACCRETION

Impact models have demonstrated the feasibility of placing ~ 1 to 2 lunar masses into Earth orbit as a result of a giant impact. But how and over what timescale can such material evolve to yield a single Moon?

5.1. Protolunar Disk Viscosity

Early considerations of a protolunar disk composed of solids suggested that the Roche-interior portion would spread rapidly (Ward & Cameron 1978). A

particulate disk would be subject to local gravitational patch instabilities; however, inside a_R such patches would be continually sheared apart, yielding enhanced collisional dissipation and an effective Roche-interior viscosity of order $v_{WC} \sim (\pi G \sigma)^2 / \Omega^3$, where σ is the disk surface density, and Ω is orbital frequency. The associated viscous spreading time is

$$\tau_{WC} = r^2 / v_{WC} = (r \Omega^2 / \pi G \sigma)^2 \Omega^{-1} \sim O(1) \text{ year } (M_L / M_d)^2 (r_d / a_R)^{3/2}, \quad (6)$$

where r_d is the disk radius; such a rapid spreading rate has been confirmed by N-body simulations (Takeda & Ida 2001). However, this description ignores the disk's thermal budget, as it assumes a disk composed of solids.

Thompson & Stevenson (1988) recognized that because of its large mass, the protolunar disk's viscosity could actually be thermally regulated. Disk material is heated by the impact event itself to a partially vapor, partially molten state, but will also continue to self-heat as it viscously spreads, with additional gravitational potential energy liberated at a rate per area of $\dot{E}_v \sim (9/4) \sigma v \Omega^2$ as the disk spreads on some timescale $\tau \sim r^2 / v$. The total energy per unit mass released during disk spreading is

$$\Delta E \sim \dot{E}_v \tau / \sigma = (9/4) (r \Omega)^2 \sim 5 \times 10^{11} \text{ erg/g } (a_R / r_d), \quad (7)$$

which exceeds the latent heat of vaporization for silicates ($L_v \sim 10^{11} \text{ erg/g}$). Thus, if a lunar-mass disk attempts to spread on a timescale that is short compared with its radiative cooling time,

$$\tau_{cool} \sim (M_d C T) / (2 \pi r_d^2 \sigma_{SB} T^4) \sim 3 \text{ years } (M_d / M_L) (a_R / r_d)^2 (2000 \text{ K} / T)^3 \quad (8)$$

(where $C \sim 10^7 \text{ erg/g} \cdot \text{K}$ is the specific heat for rock and σ_{SB} is the Stefan-Boltzmann constant), additional portions of the disk can vaporize. Vaporization acts to stabilize the disk against gravitational instability, causing disk spreading to slow as the effective viscosity decreases, until a balance between viscous dissipation and cooling is achieved. The result is a radiation-limited value for the viscosity, $v_{TS} \approx \sigma_{SB} T^4 / (\sigma \Omega^2)$, with a characteristic disk spreading time (Thompson & Stevenson 1988)

$$\tau_{TS} \sim 50 \text{ years } \left(\frac{a_R}{r_d} \right)^3 \left(\frac{T}{2000 \text{ K}} \right)^{-4} \left(\frac{M_d}{M_L} \right), \quad (9)$$

nearly two orders of magnitude slower than that in Equation 6.

5.2. Lunar Accretion

Models of accretion in an impact-generated disk were first developed in the mid-1990s (e.g., review by Kokubo, Canup & Ida 2000). A challenge in applying standard solid-body accretion models (such as those discussed in Section 3.1) to the protolunar disk is the latter's proximity to the Roche limit, where tidal effects inhibit accumulation.

Canup & Esposito (1995) developed a “particle in a box” statistical model for accretion near a_R in which the outcome of individual collisions (i.e., accretion or inelastic rebound) was determined by comparing the impact velocity to a three-body escape velocity, which included the effects of planetary tides. On average, accretion is precluded interior to $\sim 0.85a_R$, whereas exterior to $\sim 1.3a_R$, tides do not significantly limit collisional growth. In the intermediate region, the Roche zone, limited accretion can occur. Canup & Esposito (1996) applied this model to the protolunar disk, and concluded that forming a lunar-mass Moon was most likely for an initial disk containing a mass $\sim M_L$ exterior to a_R .

Ida, Canup & Stewart (1997) introduced an alternative approach, using an N-body accretion simulation to track $\sim 10^3$ initial protolunar disk objects, including tidal accretion criteria as in Canup & Esposito (1995, 1996). Figure 6 shows a simulation by Kokubo, Makino & Ida (2000) using this approach with 10^4 initial particles; resulting disk distributions are shown at times t of $0 \leq t \leq 10^3 T_K$, where $T_K = 2\pi/\Omega(a_R) \sim 7$ h. An initial disk containing $4 M_L$ with $r_d \sim a_R$ was found to flatten and viscously spread in less than a week. Once material is exterior to a_R , it rapidly accretes into moonlets that scatter inner disk material onto Earth, and after less than a year, a massive moon containing $0.85 M_L$ orbits with a semimajor axis of $a_M \approx 1.4 a_R$.

The most common outcome from the Ida, Canup & Stewart (1997) and Kokubo, Makino & Ida (2000) simulations was a single large moon (as in Figure 6) on a low eccentricity, low inclination orbit with $\langle a_M \rangle \approx 1.2$ to $1.3 a_R$. Disks that were initially more radially extended sometimes yielded two large moons. However, other models suggest that further evolution of such systems as a result of tidal and satellite-satellite interactions would eventually also yield a single moon (Canup, Levison & Stewart 1999).

Within a_R , the N-body simulations show clumping caused by instability, tidal stripping of clumps into spiral arm-like structures, and a viscosity consistent with Equation 6 (Kokubo, Makino & Ida 2000; Takeda & Ida 2001). However, from the thermal arguments given above and in Thompson & Stevenson (1988), it is clear that the viscous spreading of a \sim lunar-mass Roche-interior disk is not well-described by a particulate model, as the energy dissipated during such an evolution is sufficient to vaporize a substantial portion, if not all, of the disk’s mass. Thus, the disk’s ability to transfer much of its mass outward in less than a month as shown in Figure 6 seems physically unrealistic. In addition, impact simulations to date have never produced such massive and centrally condensed protolunar disks.

Despite such issues, results of N-body accretion simulations have revealed an important relationship between the mass and angular momentum of the initial disk and the mass of the resulting moon. Assuming that the final outcome of accretion is a single moon of mass M_M with some characteristic orbital radius a_M , a basic conservation argument can be made to estimate M_M given an initial disk mass M_D , and angular momentum, L_D . The disk material has three possible fates: collision with Earth, accretion by the Moon, or escape from the system, so that $M_D = M_{col} + M_M + M_{esc}$ and $L_D = L_{col} + L_M + L_{esc}$, where *col*, and

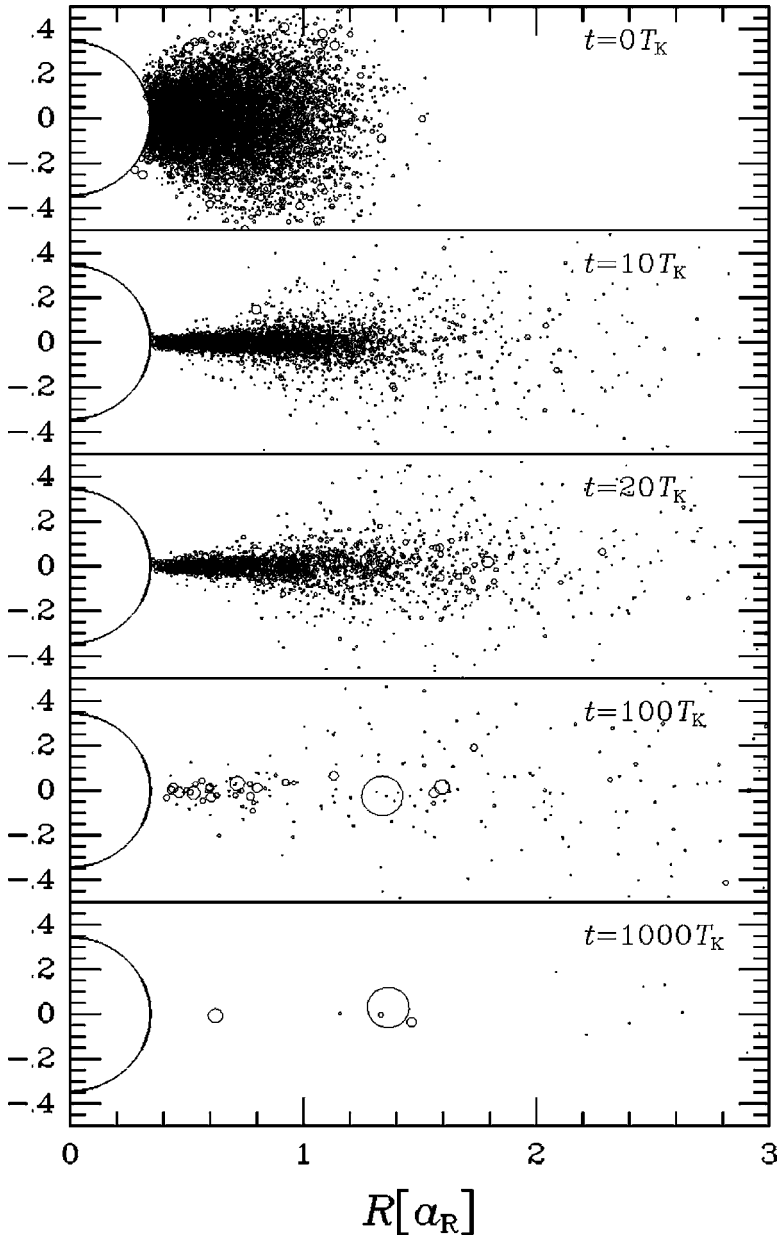


Figure 6 Time series of a protolunar disk N-body accretion simulation (from Kokubo, Makino & Ida 2000). Times are shown in units of the orbital period at the Roche limit ($T_K \sim 7$ h), distances z and $R = (x^2 + y^2)^{1/2}$ are in units of a_R ($2.9 R_\oplus$), and circles indicating disk particles are sized proportionately to the actual object's size.

esc subscripts correspond to material colliding with Earth or escaping. Given the moon's orbital angular momentum $L_M \approx M_M \sqrt{GM_\oplus a_M}$, and using typical values of L_{col} and L_{esc} from N-body simulations gives (Ida, Canup, & Stewart 1997)

$$\frac{M_M}{M_D} \approx 1.9 \left(\frac{L_D}{M_D \sqrt{GM_\oplus a_R}} \right) - 1.1 - 1.9 \left(\frac{M_{esc}}{M_D} \right). \quad (10)$$

For a variety of initial disks, simulations find values of (M_M/M_D) generally consistent with Equation 10, ranging from 0.1 to 0.55 with this fraction proportional to the initial specific angular momentum in the disk (Ida, Canup, & Stewart 1997; Kokubo, Makino, & Ida 2000). Typically, $(M_{esc}/M_D) \sim$ a few percent, although this quantity also increases for higher initial values of (L_D/M_D) . Results of impact simulations can be easily combined with Equation 10. For the orbiting disk in Figure 2 (which has a higher specific angular momentum than those considered by accretion simulations to date), Equation 10 predicts $(M_M/M_D) = 0.86$, or $M_M = 1.4 M_L$ for $(M_{esc}/M_D) = 0.05$.

Because Equation 10 results from a basic conservation argument, it will likely be appropriate over a wide range of potential disk evolution and satellite accretion timescales, so long as the final state is a moon formed near a_R . Thus, whereas the particulate simulations may overestimate the rate of the disk's evolution by neglecting the disk thermal budget, their calculated final satellite masses are probably physically reasonable. An exception to this would occur if there was an additional postimpact source of angular momentum or mass to the disk, e.g., if tidal interactions of disk material with Earth led to a positive torque on the disk on a timescale less than or comparable to the lunar accretion time. This may be possible if there are coherent structures in the disk such as spiral waves (Ward 1998), and would lead to a higher value for (M_M/M_D) than that in Equation 10.

5.3. Initial Lunar Melting

Models to date (Ida et al. 1997, Kokubo et al. 2000) suggest a lunar accretion time of only months, a period much shorter than the Moon's radiative cooling time, $M_L C / (4\pi R_L^2 \sigma_{SB} T^3) \sim 300 \text{ years } (1500 \text{ K} / T)^3$. Moreover, the specific accretional energy, $3(GM_L/R_L)/5 \sim 2 \times 10^{10} \text{ erg/g}$ (where R_L is a lunar radius), is sufficient to raise the temperature of even cooled silicate to beyond its melting point. Together these findings appear to necessitate an initially hot and molten Moon, and a heat source adequate to yield a lunar magma ocean seems all but guaranteed. The challenge lies instead in explaining how a Moon with such a hot start could have later avoided forming contraction-induced faults on its surface, as the lack of such features on the Moon is commonly interpreted as evidence that the lunar interior beneath the magma ocean was initially cold (e.g., Solomon 1986).

Pritchard & Stevenson (2000) investigated the implications of an impact origin for the Moon's initial thermal state, including the potential effects of disk cooling prior to accretion and varying lunar accretion times. They find that it takes ~ 10 years for the disk to cool significantly from the initially high temperatures

of the impact, so that if the Moon formed on this timescale or less, it inevitably would have been hot. Even a lunar accretion timescale of up to 10^2 years implies a Moon too hot to yield a ≤ 1 km change in lunar radius over the past 3.5 billion years, unless the accreting material starts with very low temperatures of ≤ 550 K (Pritchard & Stevenson 2000).

However, Pritchard & Stevenson make several arguments that the lunar radius constraint can be significantly relaxed. First, the effective Young's modulus of the lunar lithosphere (to which the predicted change in radius, ΔR , for a given lithospheric stress is inversely proportional) may be smaller than that used in previous works, e.g., owing to impact-induced cracking, perhaps allowing for a larger $\Delta R \sim 5$ km. Second, the assumption of a monotonic decrease in lunar temperature may not be correct, and could tend to lead to an overestimation of the ΔR resulting from a given initial thermal state. Arguments against an initially molten Moon based on an inferred lack of mixing and vigorous convection in the lunar interior (Turcotte & Kellogg 1986) could also be relaxed for a hot Moon if mantle convection was weak.

Pritchard & Stevenson (2000) thus conclude that a hot initial Moon cannot be ruled out with existing data and arguments, although quantifying the range of permissible initial thermal profiles remains an open problem. Future work addressing, e.g., possible geochemical constraints on the extent of lunar melting and improved modeling of the relationship between the interior evolution of the Moon and surface indicators (e.g., three-dimensional rather than one-dimensional models of radius change) may help to address this uncertainty (M. Pritchard, personal communication).

5.4. Future Work

A massive protolunar disk presents a situation in which dynamics and thermodynamics are inextricably coupled, and a self-consistent disk model must address both. There are two distinct regions of interest: inside and outside the Roche limit. Interior to a_R , transient gravitational instabilities provide an ongoing source of viscosity and dissipation, but will be limited by the disk's ability to radiate, so that the overall evolution time of the inner disk material will likely be given by the longer of the timescales in Equations 6 and 9. As such, the Roche-interior disk would be better described as a continuous, two-phase fluid (e.g., Thompson & Stevenson 1988), rather than as a collection of solid particles.

Exterior to a_R , accretion can occur rapidly once condensates are present. If much of the material that ends up in the Moon is placed into Roche-exterior orbits by the impact itself (as seems implied by the successful impacts in Canup & Asphaug 2001 and Canup 2004), similarly rapid lunar accretion times as those predicted by N-body simulations could result. A primary energy source would be the specific energy of accretion, $\sim 10^{10}$ ergs/g, which is sufficient to vaporize on the order of 10% of a lunar mass ignoring radiative cooling, and thus short accretion timescales could be accommodated by the vaporization of some small fraction

of the lunar material (e.g., Pritchard & Stevenson 2000). Given the rapidity of accretion outside a_R , and the slow evolution times inside a_R , it thus seems likely that accreting moon(s) would coexist for some time with a Roche-interior disk and potentially be affected by satellite-disk density wave interactions.

This basic picture has been explored by Ward & Canup (2000) and Canup & Ward (2000). In the latter work, a preliminary hybrid accretion model was developed in which a Roche-interior fluid disk is modeled analytically, while Roche-exterior material is tracked using an N-body simulation. As bodies in the exterior disk accrete, they gravitationally interact with the inner disk through resonant torques (e.g., the 2:1, 3:2, 4:3, etc. mean-motion resonances) that transfer angular momentum from the disk (which contracts) to the exterior moonlets (whose orbits expand) (e.g., Goldreich & Tremaine 1980). In a preliminary set of simulations, Canup & Ward (2000) found that during the Roche-exterior material's rapid accretion into a moon, the inner disk edge retreated to $\sim 2\text{--}2.4 R_\oplus$ from an initial value of $r_d = a_R = 2.9 R_\oplus$, but still contained nearly all of its initial mass. This contrasts with N-body simulations, which show rapid spreading and removal of the inner disk (e.g., Figure 6).

A related issue that merits investigation is whether the Moon's accretion could be prolonged by the slow viscous timescale in Equation 9. For example, if accretion were delayed for some time after the impact event as ejected material cooled, material from the Roche-interior disk could spread outward, and if the rate of this spreading regulated the supply of material to the growing Moon, a longer accretion timescale would result. This process could, however, be frustrated by the tendency for a massive moon to shepherd the inner disk inward. The timescale for recoil of a moon from the disk owing to resonant torques becomes comparable to its accretion time once its mass has grown to $\gtrsim O(10^{-1})M_L$ (Canup & Ward 2000; Stewart 2000). At this point, the Roche-interior disk and the outer accreting material would begin to repel one another, and further incorporation of inner disk material into the Moon could be difficult.

6. THE EARLY LUNAR ORBIT

Tracking the tidally evolved Earth-Moon-Sun system back in time provides constraints on lunar origin, and of particular importance is the Moon's current orbital inclination of $I \approx 5^\circ$ relative to the ecliptic plane. The lunar orbit normal precesses—owing both to Earth's oblateness and solar torques—about an axis that is normal to the so-called Laplacian plane. For a satellite close to a planet, the Laplacian plane is approximately the planet's equatorial plane, whereas for a more distant satellite (e.g., the current Moon) the Laplacian plane approaches the ecliptic (e.g., Goldreich 1965, 1966). Including the effect of tidal interactions, which act to decrease inclinations relative to the Laplacian plane, the Moon's current 5° inclination I , relative to the ecliptic, implies an initial lunar $I \geq 10^\circ$ relative to Earth's equatorial plane when the lunar semimajor axis, a , was within $\sim 10 R_\oplus$ (Goldreich

1966, Touma & Wisdom 1994). The origin of an initial $\sim 10^\circ$ lunar inclination is puzzling, as most dynamical processes act to decrease relative inclinations, so that a satellite accreting from a circumplanetary disk would be expected to have a very small I .

6.1. Lunar Inclination via Impact

A common proposal is that the Moon's inclination could have resulted from an impact, either as a residual of the Moon-forming event, or as a result of a later impact with Earth or the Moon by a planetesimal. Although these possibilities cannot be ruled out, all are restrictive. The first would be most likely if Earth had a significant spin prior to the giant impact, allowing for a potential offset between the postimpact spin vector of Earth and the normal to the plane initially containing the orbiting disk. However, the orbits of disk material will precess as a result of interaction with the oblate figure of Earth, and the rate of this precession is a sensitive function of orbital radius. As orbiting material differentially precesses, the disk realigns itself symmetrically about the equatorial plane, again implying a nominally low inclination moon. An exception to this could occur if a large fraction of the Moon's mass were formed as an intact object from the impact itself, as such an intact clump could retain some memory of its ejected inclination.

An impact with Earth that produced an abrupt change in the terrestrial obliquity could shift the Laplacian plane relative to the Moon's orbit. Ward (2002) pointed out that in order for such a mechanism to have been effective, the impact would have had to occur when the Moon was still sufficiently close to Earth for the Laplacian plane to be controlled by Earth, rather than by the Sun. For an impact to Earth occurring immediately after the Moon's formation, a change in the terrestrial obliquity of $\Delta\theta_\oplus \sim 10^\circ$ would suffice. But as the Moon evolves outward, a greater change in Earth's obliquity is required to render the needed offset, until this mechanism finally becomes ineffective for $a \gtrsim 20 R_\oplus$. Including a basic tidal evolution model, the impact of a ~ 1 to $4 M_L$ object with Earth could yield the required shift if the impact occurred within $\sim 10^5$ to 10^6 years of the Moon-forming impact. This is short compared with typical times between large impacts predicted by late-stage accretion models, $\sim 10^7$ to 10^8 years.

A large subsequent impact to the Moon could in principle tilt its orbit. For the case of a head-on polar impact to an initially $I \sim 0^\circ$ Moon, a change in orbital inclination of ΔI requires

$$\tan(\Delta I) = \frac{|L_{imp}|}{|L_{orb}|} \approx \frac{m_{imp} v_{imp} a}{M_L \sqrt{GM_\oplus a}}, \quad (11)$$

where m_{imp} and v_{imp} are the impactor mass and velocity, and the lunar orbit is assumed to have $e^2 \ll 1$. The ΔI needed for consistency with the Moon's orbit is a function of a , with $\Delta I \approx 12^\circ (6 R_\oplus / a)^{1/4}$. For $v_{imp} \approx v_\infty$, the required minimum

impactor mass for a head-on polar impact is

$$\frac{m_{\text{imp}}}{M_L} \gtrsim 12^\circ \left(\frac{6R_\oplus}{a} \right)^{1/4} \frac{v_{\text{orb}}}{v_\infty}, \quad (12)$$

where $v_{\text{orb}} \equiv \sqrt{GM_\oplus/a}$ is the Moon's orbital velocity. For $v_\infty \sim 6$ km/s (e.g., Stewart 2000), Equation 12 gives a minimum impactor size of $\approx 0.04 M_L$ at $a = 25 R_\oplus$, and a minimum impact energy per unit target mass of $Q \sim 7 \times 10^9$ erg/g, within a factor of approximately two of the Moon's per-unit mass binding energy, $3/5 (GM_L/R_L) \approx 1.7 \times 10^{10}$ erg/g.³ Thus, even an optimally aligned impact to the Moon capable of producing the lunar *I* could also cause significant disruption, and to the degree to which this occurred, reaccretion would tend to realign the Moon in the Laplacian plane.

6.2. Lunar Inclination via Disk Interactions

A more satisfying resolution to the “inclination problem” would be one that links the Moon's orbit to its formation from a giant impact, thus removing the need for later additional events. Ward & Canup (2000) proposed that the Moon's inclination resulted from resonant interactions between the newly formed satellite and a companion Roche-interior disk. This mechanism depends on the coexistence of the Moon for 10 to 100 years with a disk of mass ~ 0.5 –1 lunar masses.

Periodic perturbations from a satellite excite waves in an accompanying disk, and the torques resulting from gravitational interactions between the satellite and the wave patterns it generates can significantly modify the satellite's orbit. Waves in the disk are generated at locations where the ratio of the local mean motion to that of the satellite is approximately equal to a ratio of two integers. Mean motion resonances involving in-plane perturbations are known as Lindblad resonances and generate spiral density waves, whereas vertical resonances involving perturbations associated with the satellite's inclination lead to spiral bending waves (e.g., Shu 1984).

Ward & Canup point out that for a newly formed moon with $a \gtrsim a_R$, there would have initially been many resonances within a companion Roche-interior disk. As resonant torques cause the satellite to migrate outward, its mean motion resonances (which each occur at a fixed fraction of a satellite's position) also migrate, moving outward through and finally leaving the disk. Of the strongest resonances (i.e., zero- or first-order in the satellite's e or I), the most interior, and therefore the last to migrate out of the disk, is the 3:1 inner vertical resonance (IVR). It becomes the only strong resonance left in the disk once the moon has migrated outward to $a \sim 2.08 r_d$, where r_d is the radius of the disk's outer edge. At this point the 3:1 IVR

³For comparison, the specific impact energy estimated for forming the Moon's South Pole–Aitken basin is between $Q \sim 10^8$ erg/g for a head-on impact (Takata 1996) to $\sim 10^{10}$ erg/g for a very grazing impact (Schultz 1997).

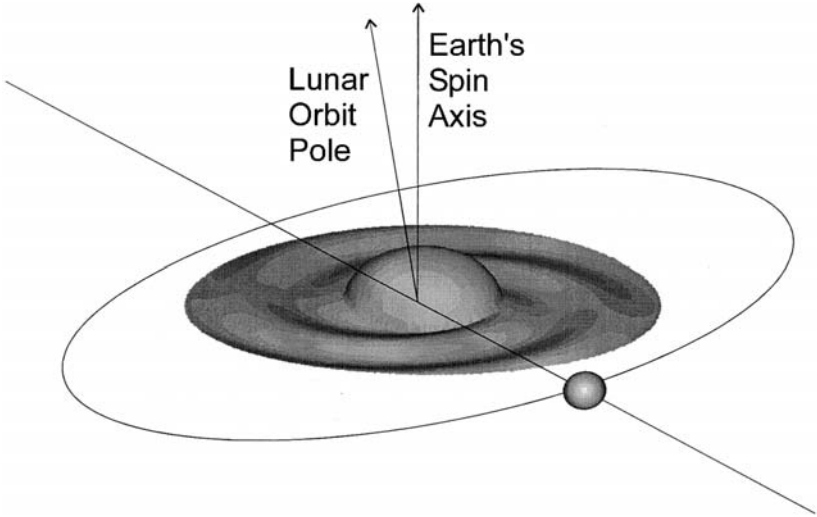


Figure 7 A schematic of the bending wave produced by the 3:1 inner vertical resonance.

increases a and I for a mass M_M satellite at the rates:

$$\frac{da}{dt} = \frac{3T_{IVR}}{M_M a \Omega}, \quad \text{and} \quad \frac{dI}{dt} = \frac{T_{IVR}}{M_M a^2 \Omega \sin I} \left(\frac{3}{2} \cos I - 1 \right), \quad (13)$$

where Ω is the moon's orbital frequency, and T_{IVR} is the torque due to the 3:1 IVR that is proportional to the disk surface density and $\sin^2 I$ (Ward & Canup 2000).

Figures 7–8 show a 3:1 bending wave and the resulting evolution of an example moon and disk from Ward & Canup. Here, the moon's orbit evolves because of the 3:1 Lindblad and vertical resonances, interaction with tides raised on Earth, and dissipation resulting from lunar tides. The disk evolves as a result of torques from the Moon and its internal viscosity, with the latter chosen to be the longer of those given in Equations 6 and 9. In Figure 8, the evolution of I , a , and σ are shown as a Moon with an initial $I = 1^\circ$ recoils from a disk containing $M_d = 0.75 M_L$; the resulting inclinations are 12.3° and 14.5° for initial disk spreading times of 37 and 50 years. The growth of I stops as either the disk is depleted or as the 3:1 IVR migrates past a_R .

The mechanism by Ward and Canup relies on a resonance that would naturally result if the Moon accreted near the Roche limit on a timescale much shorter than the viscous lifetime of a companion interior disk.⁴ Whether or not the resulting inclination is sufficient depends on the initial inner disk mass and its viscosity,

⁴The mechanism by Ward and Canup requires a period of recoil of an outer satellite from an inner disk in which the 3:1 resonance controls the evolution of both I and a ; in alternative configurations, effects of other resonances can lead to inclination damping (e.g., Borderies, Goldreich & Tremaine 1984; Ward & Hahn 1995).

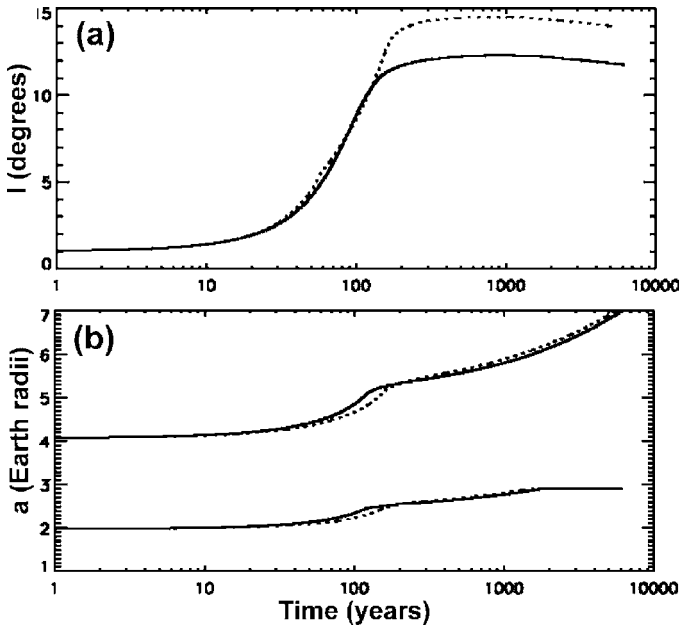


Figure 8 The evolution of the Moon's orbit as it resonantly interacts with an inner impact-generated disk (from Ward & Canup 2000). Values for (a) $I(t)$, and (b) $a_M(t)$ and $r_d(t)$, are shown as the Moon recoils away from a $M_d(t=0) = 0.75 M_L$ disk with an initial viscous spreading time of $t_{\text{spread}}(t=0) = 37$ years (solid) and 50 years (dotted).

with successful cases resulting for disks containing 0.5 to $1 M_L$, with initial viscous spreading times $\gtrsim 50$ years. These are reasonable given current impact and disk models.

6.3. Lunar Inclination via Solar Resonances

If disk torques were not sufficient to generate the Moon's inclination, a later series of dynamical events could have been responsible instead. Touma & Wisdom (1998) proposed that the lunar orbital tilt resulted through encounters with solar resonances as the Moon tidally receded from Earth (also see reviews by Peale 1999 and Touma 2000).

The first element of the model by Touma and Wisdom is the so-called evection resonance (Kaula & Yoder 1976), in which the period of precession of the lunar periape is approximately equal to Earth's orbital period around the Sun. Initially, the evection resonance would have been located well outside the lunar orbit if the Moon was accompanied by a substantial Roche-interior disk, owing to the latter's influence on the lunar precession rate. As such a disk dissipated, the evection resonance would move inward to within $\sim 10 R_\oplus$ once the disk mass has fallen to less than $\sim 0.1 M_L$.

Touma & Wisdom considered an Earth-Moon-Sun system without any disk, so that the evection resonance is located at $4.6 R_{\oplus}$. They demonstrated that a moon tidally evolving outward from an initial $a = 3.5 R_{\oplus}$ could become captured in the evection resonance, leading to a substantial increase in the lunar orbital eccentricity, e . The magnitude of this increase is a function of the assumed tidal dissipation factor of the Moon because tides raised on the Moon act to decrease e . Rates of lunar dissipation similar to or less than that of the Moon today yield up to $e \sim 0.5$ before the Moon escapes evection.

If the evection-generated e is sufficiently high, the Moon next encounters a mixed eccentricity-inclination resonance at $a \sim 6 R_{\oplus}$, dubbed the eviction by Touma and Wisdom. A first outward, noncapture passage through eviction generates a 2 to 3° inclination. If at this time the rate of lunar dissipation significantly increases—as would be expected if an initially cold Moon had been tidally heated by the high, evection produced eccentricity—the direction of the Moon's orbital migration temporarily reverses. Moving inward across the eviction resonance then leads to capture, and the excitation of a 9 to 13° inclination. As the lunar eccentricity continues to decrease owing to lunar tides, escape from the resonance occurs and the Moon again reverses direction, continuing on its outward path with its inclination preserved.

The most significant constraint of the model by Touma and Wisdom is its requirement of an initially cold Moon, so that the initial rate of tidal dissipation in the Moon is similar to its current value and a high value of eccentricity can be achieved via the evection resonance. Rapid lunar accretion and/or heating of protolunar material from the impact itself imply an initially hot Moon; a much slower accretion time of $\geq 10^3$ years would be needed for the Moon to instead accrete cold (Peale 1999, Pritchard & Stevenson 2000).

7. CONCLUSION

The most plausible explanation for the high angular momentum of the Earth-Moon system and the Moon's unusual compositional characteristics is that it is the result of an impact that occurred near the end of Earth's accretion. Taken as a whole, the past decade of dynamical models of a lunar origin via impact offer substantial support to this premise. Key dynamical findings include:

- *Large impacts are common in late-stage terrestrial accretion.* Modern numerical techniques continue to support earlier results, finding multiple potential moon-forming impacts are likely in a forming terrestrial planet system. An intriguing possibility is that such collisions might be directly observable in extrasolar systems (Stern 1994, Zhang & Sigurdsson 2003).
- *A single impact is consistent with the Earth-Moon system.* Oblique impacts by objects containing ~ 0.10 to $0.15 M_{\oplus}$ yield appropriate protolunar disks together with approximately the total mass and angular momentum of the Earth-Moon system. Such an agreement would have to be viewed as

coincidental in order to advocate alternative scenarios. The impact parameter is a key constraint, with $0.7 \leq b \leq 0.8$ for the most successful cases. For an isotropic flux of impactors, the probability of an impact with angle ξ to $(\xi + d\xi)$ (where ξ is the angle between the impact trajectory and the local surface normal) is $dP = \sin 2\xi d\xi$ (e.g., Pierazzo & Melosh 2000); the most likely value is $b = 0.7$ ($\xi = 45^\circ$), so that the probability of an impact in the optimal range to yield both a massive and iron-depleted moon is $\sim 15\%$.

- *Impact-generated material accretes into a single moon.* Orbiting material roughly centered at the Roche limit most often accumulates into a single moon. Systems of multiple, \sim lunar-sized moons are generally not dynamically stable owing to tidal orbital evolution and mutual interactions.
- *The initial lunar inclination problem can be reconciled with an impact origin.* Until recently, an apparent deficiency of the impact hypothesis has been its inability to account for the inclination of the lunar orbit (e.g., Boss & Peale 1986). This review has discussed three possible resolutions to the inclination problem: impacts, disk torques, and solar resonances. These suggest that the Moon's inclination can be reconciled with its formation from an equatorial disk, and indeed I may actually be a result of its impact origin.

In each of these areas, important issues remain outstanding and merit further work. A key future objective should be reconciliation of predictions of the dynamical models with the physical properties of the Moon and Earth. Open issues include:

- *Impact-dominated terrestrial accretion.* Can a late impact phase be reconciled with the nearly circular orbits of Earth and Venus, as well as geochemical evidence suggesting composition zoning in the protoplanetary disk?
- *Impactor origin of the Moon.* Is this a dynamical requirement, as models to date suggest? Is this consistent with the Moon's compositional similarities to Earth (e.g., Munker et al. 2003)?
- *Lunar accumulation: timescale and thermal consequences.* What happens between the occurrence of impact and the Moon's accretion, and over what timescale? Can this be reconciled with the initial disk masses predicted by impact simulations and physical constraints on the degree of initial lunar melting?
- *Impact processing of lunar material.* What are the chemical and elemental implications of an impact origin, and are these consistent with lunar geochemistry, e.g., the Moon's volatile depletions?
- *Direct formation of an intact moon.* Is this a physically realistic mode of formation, and what are its dynamical and compositional implications?

Overall, perhaps the greatest shift in thinking that has arisen from the past decade of lunar origin studies has been the realization that the impact production of satellites appears an efficient and probable event during planetary accretion. This is

somewhat ironic in that an original strength of the giant impact theory was believed to be its ability to account for the uniqueness of our Moon by invoking the occurrence of an uncommon event. Modeling instead suggests that many low-velocity impacts between similarly sized objects will leave material in bound orbit, suggesting that the very type of collisions believed necessary for the final stages of solid planet growth will also be active producers of satellites. Many such satellites would have tidally evolved to their demise aeons ago or been destroyed by later impacts, leaving us today with our current Moon—and perhaps Charon—as the surviving members of a once much-larger population of impact-produced planetary satellites.

ACKNOWLEDGMENTS

This paper has benefited from past work and discussions with William Ward, Alastair Cameron, Erik Asphaug, Jay Melosh, Betty Pierazzo, Richard Mihan, Glen Stewart, Larry Esposito, Qingzhu Yin, Kevin Righter, Dave Stevenson, and Alex Halliday. Constructive comments were provided by Bill Bottke, Luke Dones, Matthew Pritchard, Kevin Righter, William Ward, and Qingzhu Yin. The author gratefully acknowledges support from NASA's Origins of Solar Systems and NSF's Planetary Astronomy programs.

**The *Annual Review of Astronomy and Astrophysics* is online at
<http://astro.annualreviews.org>**

LITERATURE CITED

- Agnor CB, Canup RM, Levison HF. 1999. *Icarus* 142:219–37
- Agnor CB, Ward WR. 2002. *Astrophys. J.* 567: 579–86
- Benz W, Cameron AGW, Melosh HJ. 1989. *Icarus* 81:113–31
- Benz W, Slattery WL, Cameron AGW. 1986. *Icarus* 66:515–35
- Benz W, Slattery WL, Cameron AGW. 1987. *Icarus* 71:30–45
- Borderies NP, Goldreich P, Tremaine S. 1984. *Astrophys. J.* 284:429–34
- Boss AP, Peale S. 1986. See Hartmann et al. 1986, pp. 59–101
- Cameron AGW. 1997. *Icarus* 126:126–37
- Cameron AGW. 2000. See Canup & Righter 2000, pp. 133–44
- Cameron AGW. 2001. *Meteor. Planet. Sci.* 36: 9–22
- Cameron AGW, Benz W. 1991. *Icarus* 92:204–16
- Cameron AGW, Ward WR. 1976. *Lunar Planet. Sci. Conf.* 7:120–22
- Canup RM. 2004. *Icarus*. 168:433–56
- Canup RM, Agnor CB. 2000. See Canup & Righter 2000, pp. 113–29
- Canup RM, Asphaug E. 2001. *Nature* 412:708–12
- Canup RM, Esposito LW. 1995. *Icarus* 113: 331–52
- Canup RM, Esposito LW. 1996. *Icarus* 119: 427–46
- Canup RM, Levison HF, Stewart GR. 1999. *Astron. J.* 117:603–20
- Canup RM, Righter K, eds. 2000. *Origin of the Earth and Moon*. Tucson: Univ. Ariz. Press. 555 pp.
- Canup RM, Ward WR. 2000. *Lunar Planet. Sci. Conf.* 31:1916–17
- Canup RM, Ward WR, Cameron AGW. 2001. *Icarus* 150:288–96
- Chambers JE. 1999. *MNRAS* 304:793–99

- Chambers JE. 2001. *Icarus* 152:205–24
- Chambers JE, Wetherill GW. 1998. *Icarus* 136:304–27
- Dones L, Tremaine S. 1993. *Icarus* 103:67–92
- Duncan MJ, Levison HF, Lee MH. 1998. *Astron. J.* 116:2067–77
- Goldreich P. 1965. *Astron. J.* 70:5
- Goldreich P. 1966. *Rev. Geophys.* 4:411–39
- Goldreich P, Tremaine S. 1980. *Astrophys. J.* 241:425–41
- Greenberg R, Fischer M, Valsecchi B, Carusi A. 1997. *Icarus* 129:384–400
- Halliday AN. 2000. *Earth Planet. Sci. Lett.* 176:17–30
- Halliday AN. 2004. *Nature* 427:505–9
- Halliday AN, Lee DC, Jacobsen SB. 2000. See Canup & Righter 2000, pp. 45–62
- Hartmann WK, Davis DR. 1975. *Icarus* 24:504–15
- Hartmann WK, Phillips RJ, Taylor GJ, eds. 1986. *Origin of the Moon*. Houston: Lunar Planet. Inst. 781 pp.
- Hood LL, Zuber MT. 2000. See Canup & Righter 2000, pp. 397–412
- Ida S, Canup RM, Stewart GR. 1997. *Nature* 389:353–57
- Imaeda Y, Inutsuka S. 2002. *Astrophys. J.* 569:501–18
- Jacobsen SB. 1999. *Lunar Planet. Sci. Conf.* 30:1978
- Jones JH, Delano JW. 1989. *Geochim. Cosmochim. Acta* 53:513–27
- Jones JH, Palme H. 2000. See Canup & Righter 2000, pp. 197–216
- Kaula WM, Yoder CF. 1976. *Lunar Planet. Sci. Conf.* 17:440–42
- Kipp ME, Melosh HJ. 1987. *Lunar Planet. Sci. Conf.* 28:491–92
- Kleine T, Münker C, Mezger K, Palme H. 2002. *Nature* 418:952–55
- Kokubo E, Canup RM, Ida S. 2000. See Canup & Righter 2000, pp. 145–64
- Kokubo E, Makino J, Ida S. 2000. *Icarus* 148:419–36
- Kominami J, Ida S. 2002. *Icarus* 157:43–56
- Laskar J. 1997. *Astron. Astrophys.* 317:L75–78
- Lee DC, Halliday AN. 1995. *Nature* 378:771–74
- Lee DC, Halliday AN, Leya I, Wieler R, Wiechert U. 2002. *Earth Planet. Sci. Lett.* 198:267–74
- Leya I, Wieler R, Halliday AN. 2000. *Earth Planet. Sci. Lett.* 175:1–12
- Lissauer JJ, Kary DM. 1991. *Icarus* 94:126–59
- Lucey PG, Taylor GJ, Malaret E. 1995. *Science* 268:1150–53
- Lucy LB. 1977. *Astron. J.* 82:1013–24
- Melosh HJ. 2000. *Lunar Planet. Sci. Conf.* 31:1903
- Melosh HJ, Kipp ME. 1989. *Lunar Planet. Sci. Conf.* 20:685–86
- Melosh HJ, Pierazzo E. 1997. *Lunar Planet. Sci. Conf.* 28:935
- Monaghan JJ. 1992. *Annu. Rev. Astron. Astrophys.* 30:543–74
- Morishima R, Watanabe S. 2001. *Earth Planet. Space* 53:213–31
- Munker C, Pfander JA, Weyer S, Buchl A, Kleine T, Mezger K. 2003. *Science* 301:84–87
- Nelson AF, Benz W, Adams FC, Arnett D. 1998. *Astrophys. J.* 502:342–71
- Ohtsuki K, Ida S. 1998. *Icarus* 131:393–420
- Peale S. 1999. *Annu. Rev. Astron. Astrophys.* 37:533–602
- Pierazzo E, Melosh HJ. 2000. *Annu. Rev. Earth Planet. Sci.* 28:141–68
- Pritchard ME, Stevenson DJ. 2000. See Canup & Righter 2000, pp. 179–96
- Righter K. 2002. *Icarus* 158:1–13
- Shakura NI, Sunyaev RA. 1973. *Astron. Astrophys.* 24:337–55
- Shu F. 1984. In *Planetary Rings*, ed. R Greenberg, A Brahic, pp. 513–61. Tucson: Univ. Ariz. Press. 784 pp.
- Shukolyukov A, Lugmair GW. 2000. *Space Sci. Rev.* 92:225–36
- Snyder GA, Borg LE, Nyquist LE, Taylor LA. 2000. See Canup & Righter 2000, pp. 361–98
- Solomon SC. 1986. See Hartmann et al. 1986, pp. 435–52
- Stern SA. 1994. *Astron. J.* 108:2312–17
- Stevenson DJ. 1987. *Annu. Rev. Earth Planet. Sci.* 15:271–315

- Stewart GR. 2000. See Canup & Righter 2000, pp. 217–23
- Stewart GR, Wetherill GW. 1988. *Icarus*. 74:542–53
- Takeda T, Ida S. 2001. *Astrophys. J.* 560:514–33
- Tillotson JH. 1962. *Gen. At., Rep. GA-3216*, San Diego, Calif.
- Thompson C, Stevenson DJ. 1988. *Astrophys. J.* 333:452–81
- Thompson SL, Lauson HS. 1972. *Sandia Tech. Rep. SC-RR-710714*, Sandia Natl. Lab. Albuquerque, NM
- Touma J. 2000. See Canup & Righter 2000, pp. 165–78
- Touma J, Wisdom J. 1994. *Astron. J.* 108:1943–61
- Touma J, Wisdom J. 1998. *Astron. J.* 115:1653–63
- Turcotte DL, Kellogg LH. 1986. See Hartmann et al. 1986, pp. 311–29
- Vickery AM, Melosh HJ. 1987. *Lunar Planet. Sci. Conf.* 28:1042–43
- Walker D, Hays JF. 1977. *Geology* 5:425–28
- Ward WR. 1998. *Proc. Origin Earth Moon Conf., Monterey, Calif.* 957:52. Houston: Lunar Planet. Inst.
- Ward WR. 2002. *Astrophysical tides: the effects in the solar and exoplanetary systems*. IAU Colloq., Nanjing, China
- Ward WR, Cameron AGW. 1978. *Lunar Planet. Sci. Conf.* 9:1205–7
- Ward WR, Canup RM. 2000. *Nature* 403:741–43
- Ward WR, Hahn JM. 1995. *Icarus* 110:95–108
- Weidenschilling SJ, Spaute D, Davis DR, Marzari F, Ohtsuki K. 1997. *Icarus* 128:429–55
- Wetherill GW. 1985. *Science* 228:877–79
- Wetherill GW. 1992. *Icarus* 100:307–25
- Wiechert U, Halliday AN, Lee DC, Snyder GA, Taylor LA, Rumble D. 2001. *Science* 294:345–48
- Wisdom J, Holman M. 1991. *Astron. J.* 102:1528–38
- Wood JA, Dickey JS, Marvin UB, Powell BN. 1970. *Science* 167:602–4
- Wood JH. 1986. See Hartmann et al. 1986, pp. 17–56
- Yin QZ, Jacobsen SB, Yamashita K, Blichert-Toft J, Telouk P, Albarede F. 2002. *Nature* 418:949–52
- Zhang B, Sigurdsson S. 2003. *Astrophys. J.* 596:L95–98

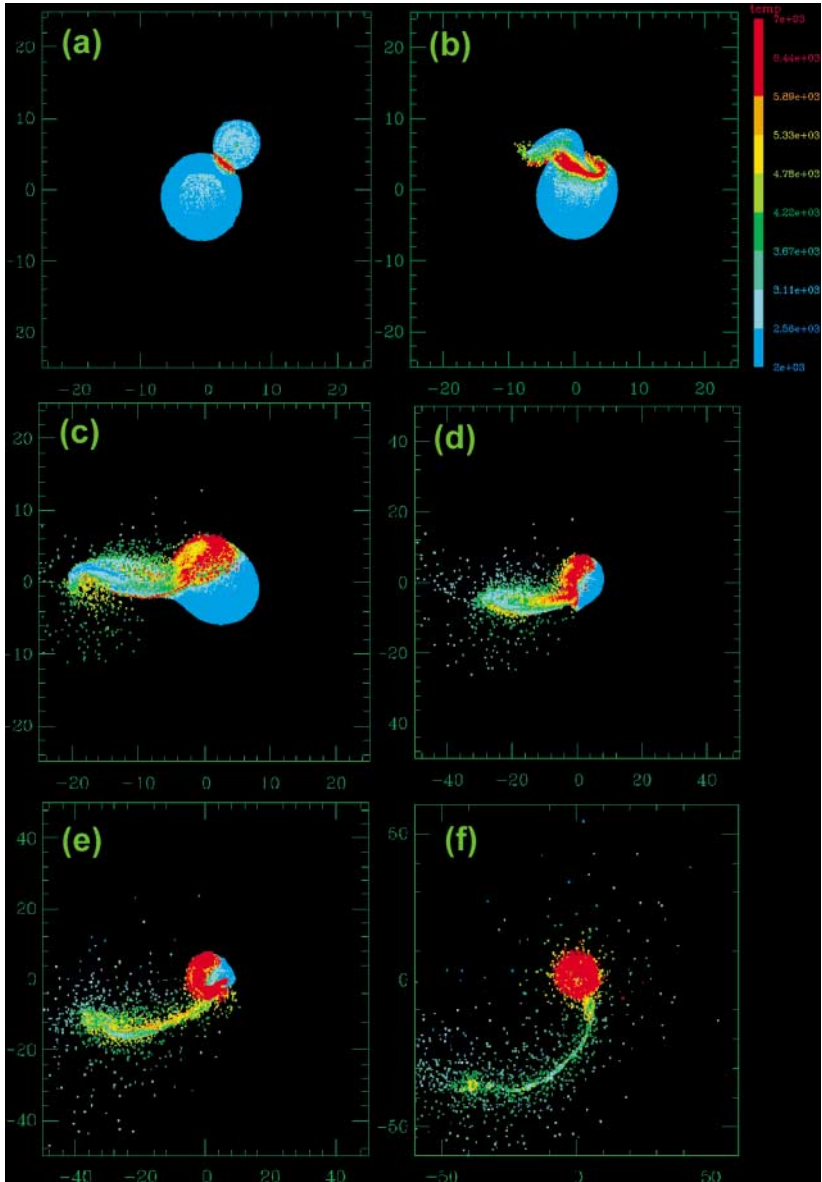


Figure 2 Time series of an impact with $N = 60,000$, $\gamma = 0.13$, $v = v_{esc}$, and $b' = 0.730$ (from Canup 2004). Times in hours for frames *a* through *k* are 0.11, 0.32, 0.86, 1.40, 2.16, 4.85, 5.93, 13.48, 18.87, 21.02, and 26.95, respectively. Color scales with particle temperature in degrees K; frames *a* through *k* are looking down onto the plane of the impact with red indicating particles with $T > 6440$ K. Distances are shown in units of 1000 km. Frame *l* is the system at 27 h, viewed edge-on; here the temperature scale has been shifted so that red corresponds to $T > 9110$ K.

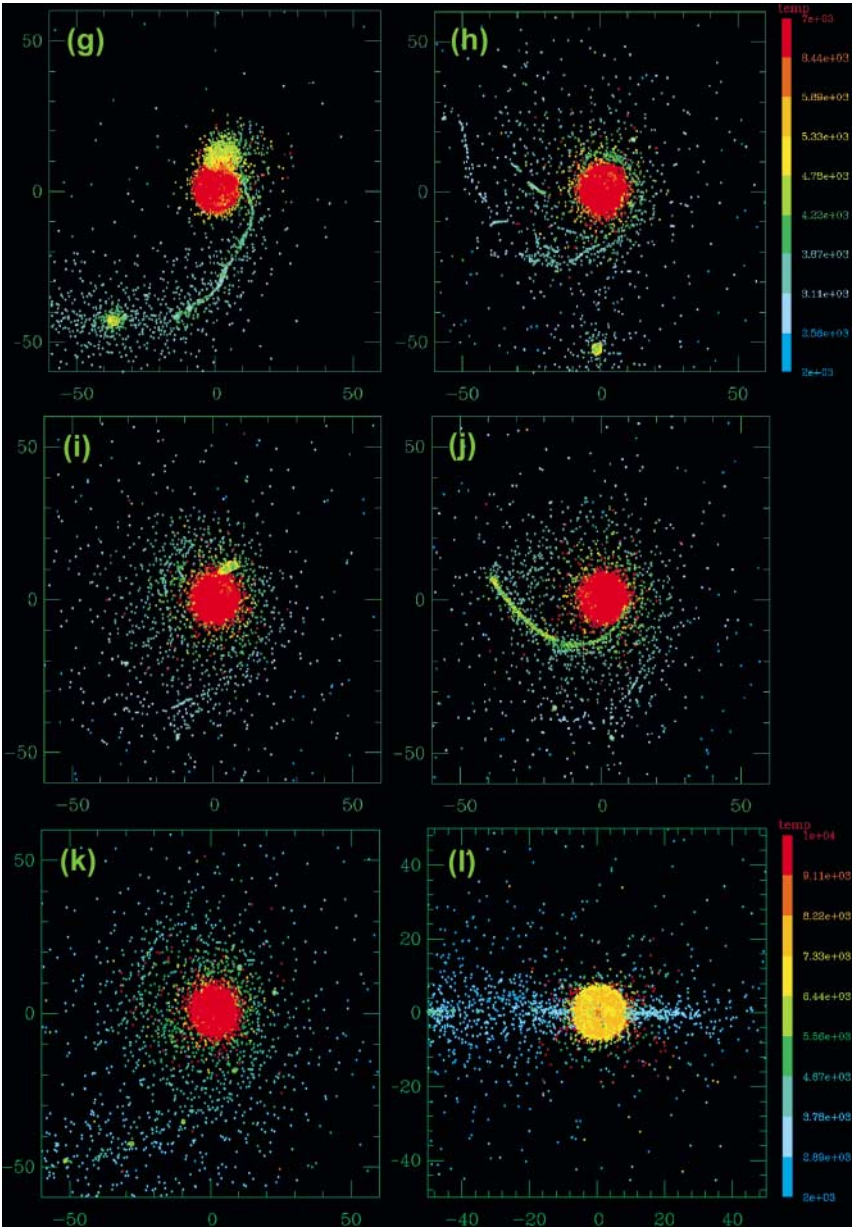


Figure 2 (Continued)

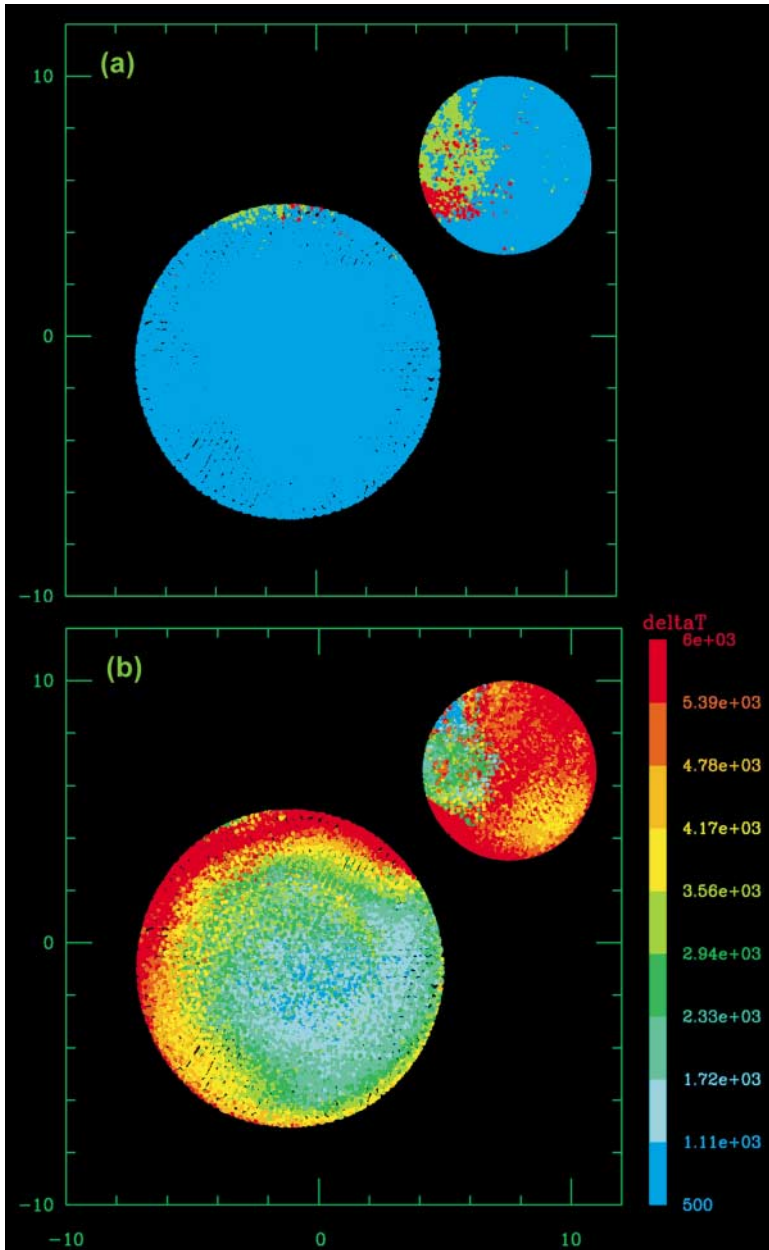


Figure 3 Mapping of final quantities from the impact shown in Figure 2 onto the original figures of the target and impactor. (a) Mapping of final particle states; yellow particles end up in the orbiting disk, red ones escape the system, and blue ones are located in the protoearth. (b) Change in particle temperatures experienced during impact shown in Figure 2; color scales with temperature in degrees K.

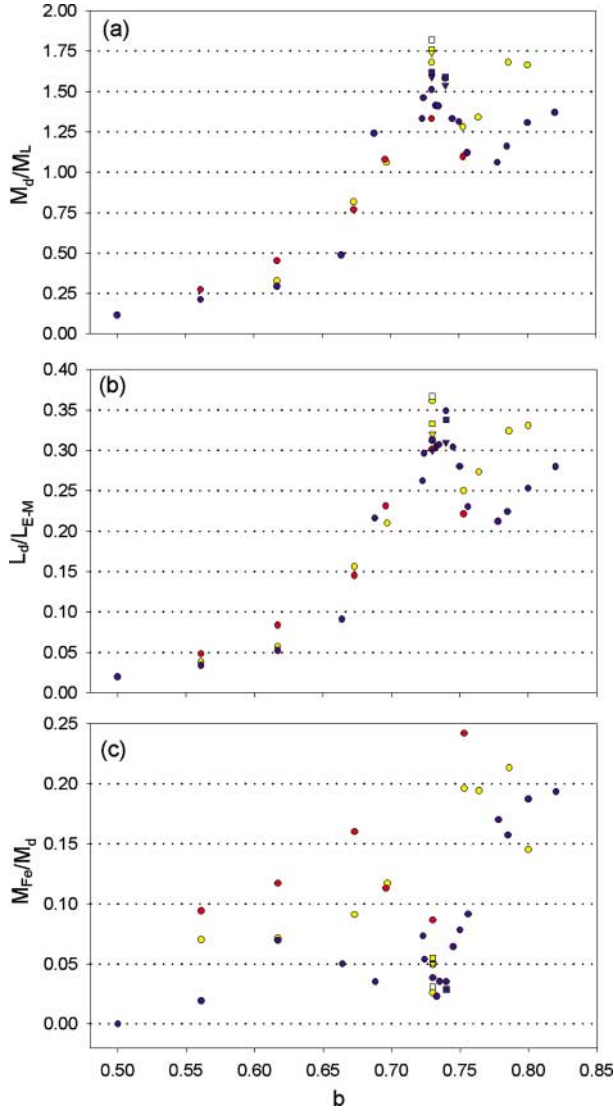


Figure 5 Results from 40 impact simulations from Canup (2004) that had $M_T = 1.02 M_\oplus$, $\gamma = 0.13$, and varied resolutions and impact velocities. Blue, white, yellow, and red symbols correspond to $(v_{imp}/v_{esc}) = 1.00, 1.02, 1.05$, and 1.10 impacts, respectively. Triangles, circles, squares, and inverted triangles correspond to $N = 20,000, 30,000, 60,000$, and $120,000$ particles, respectively. (a) Orbiting mass in lunar masses versus normalized impact parameter, (b) orbiting angular momentum in units of $L_{\oplus-M}$; and (c) mass fraction of orbiting iron.

CONTENTS

FRONTISPIECE, <i>Adriaan Blaauw</i>	xii
MY CRUISE THROUGH THE WORLD OF ASTRONOMY, <i>Adriaan Blaauw</i>	1
ASTROPHYSICS WITH PRESOLAR STARDUST, <i>Donald D. Clayton</i> <i>and Larry R. Nittler</i>	39
THE FIRST STARS, <i>Volker Bromm and Richard B. Larson</i>	79
ISO SPECTROSCOPY OF GAS AND DUST: FROM MOLECULAR CLOUDS TO PROTOPLANETARY DISKS, <i>Ewine F. van Dishoeck</i>	119
NEUTRON STAR COOLING, <i>D.G. Yakovlev and C.J. Pethick</i>	169
INTERSTELLAR TURBULENCE I: OBSERVATIONS AND PROCESSES, <i>Bruce G. Elmegreen and John Scalo</i>	211
INTERSTELLAR TURBULENCE II: IMPLICATIONS AND EFFECTS, <i>John Scalo</i> <i>and Bruce G. Elmegreen</i>	275
GRS 1915+105 AND THE DISC-JET COUPLING IN ACCRETING BLACK HOLE SYSTEMS, <i>Rob Fender and Tomaso Belloni</i>	317
IMPULSIVE MAGNETIC RECONNECTION IN THE EARTH'S MAGNETOTAIL AND THE SOLAR CORONA, <i>A. Bhattacharjee</i>	365
ABUNDANCE VARIATIONS WITHIN GLOBULAR CLUSTERS, <i>Raffaele Gratton,</i> <i>Christopher Sneden, and Eugenio Carretta</i>	385
DYNAMICS OF LUNAR FORMATION, <i>Robin M. Canup</i>	441
EROS AND FAINT RED GALAXIES, <i>Patrick J. McCarthy</i>	477
FINE STRUCTURE IN SUNSPOTS, <i>John H. Thomas and Nigel O. Weiss</i>	517
PLANET FORMATION BY COAGULATION: A FOCUS ON URANUS AND NEPTUNE, <i>Peter Goldreich, Yoram Lithwick, and Re'em Sari</i>	549
SECULAR EVOLUTION AND THE FORMATION OF PSEUDOBULGES IN DISK GALAXIES, <i>John Kormendy and Robert C. Kennicutt, Jr.</i>	603
YOUNG STARS NEAR THE SUN, <i>B. Zuckerman and Inseok Song</i>	685

INDEXES

Subject Index	723
Cumulative Index of Contributing Authors, Volumes 31–42	749
Cumulative Index of Chapter Titles, Volumes 31–42	752

ERRATA

An online log of corrections to *Annual Review of Astronomy and Astrophysics* chapters may be found at <http://astro.annualreviews.org/errata.shtml>