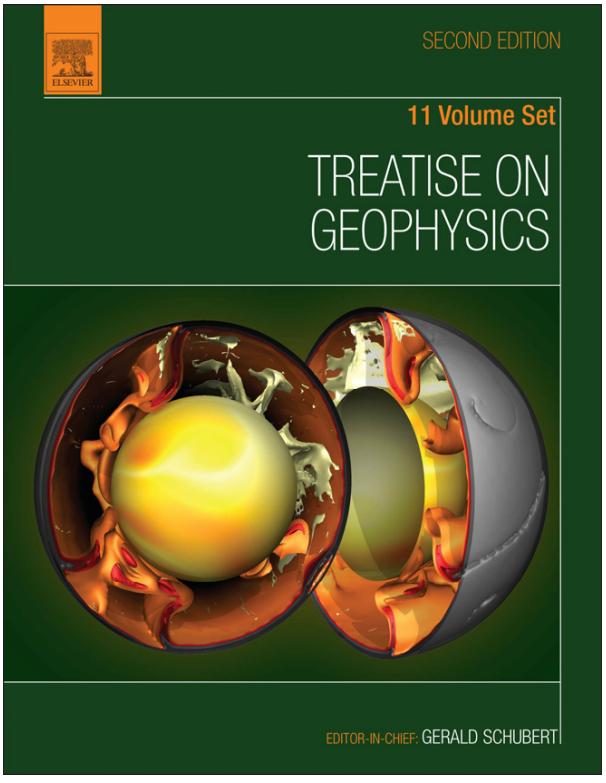
Provided for non-commercial research and educational use.
Not for reproduction, distribution or commercial use.

This article was originally published in Treatise on Geophysics, Second Edition, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

http://www.elsevier.com/locate/permissionusematerial

Peale S.J., and Canup R.M The Origin of the Natural Satellites. In: Gerald Schubert (editor-in-chief) Treatise on Geophysics, 2nd edition, Vol 10. Oxford: Elsevier; 2015. p. 559-604.

Author's personal copy

10.17 The Origin of the Natural Satellites

SJ Peale, University of California, Santa Barbara, CA, USA
RM Canup, Southwest Research Institute, Boulder, CO, USA

© 2015 Elsevier B.V. All rights reserved.

This chapter is a revision of the previous edition chapter by S J Peale, Volume 10, pp. 465–508, © 2007, Elsevier B.V.

10.17.1	**Introduction**	559
10.17.2	**Earth–Moon System**	560
10.17.3	**Mars System**	567
10.17.4	**Jupiter System**	569
10.17.4.1	SEMM Model	570
10.17.4.2	Planetesimal Capture Model	572
10.17.4.3	Starved Accretion Disk Model	572
10.17.5	**Saturn System**	578
10.17.6	**Uranus System**	583
10.17.7	**Neptune System**	584
10.17.8	**Pluto System**	586
10.17.9	**Irregular Satellites**	588
Appendix A	**Accretion Disks**	589
Appendix B	**Tides**	594
Acknowledgments		598
References		598

10.17.1 Introduction

This chapter is an update to the chapter of the same title in the first edition of this volume of the *Treatise on Geophysics*. The information on, for example, the Mars and Neptune systems has not changed much, whereas new research and observations have motivated significant changes in the discussions of especially the Earth–Moon, Pluto, Saturn, and Uranus systems. This chapter will not be all inclusive, but we will attempt to select those thoughts and mechanisms that are currently prominent in the literature and will describe both virtues and caveats in the discussion. The authors' prejudices and judgments will no doubt be evident, but we hope advice from our colleagues has tempered the impact of such. The origin of the satellites constrains the origin of the solar system, and we hope this chapter will be a suitable introduction to researchers who seek to understand the origin of our solar system and that of planetary systems in general.

Each system of satellites is unique, and it is impossible to apply the same detailed sequence of events to account for the origin of the different systems. Of all the planetary satellites, our own Moon has generated the most curiosity about its origin and the subsequent tidal evolution that has placed constraints on the theories. The dynamic mechanisms that have been proposed for the formation of the Moon are evaluated by Boss and Peale (1986), and one or another of these has been proposed for other natural satellites. The mechanisms include what have been called rotational fission, precipitation fission, intact capture, disintegrative capture, binary accretion, and formation by giant impact. In rotational fission, favored by Darwin (1879, 1880), the Moon is derived from material shed from a rapidly spinning Earth as a result of a dynamic fission instability. Precipitation fission (Ringwood, 1970, 1972) requires a massive primordial atmosphere sufficiently hot to vaporize silicates. As the atmosphere cools, the silicates condense and precipitate to the Earth's equatorial plane where they can collect into the Moon, while the volatiles in the atmosphere are swept away by a T Tauri solar wind. For intact capture, the Moon is assumed to form from the accretion of planetesimals in heliocentric orbit relatively close to that of the Earth. During a close approach to the Earth, energy must be removed from the Moon to leave it in a bound orbit. Methods proposed for this energy loss include collision with an already existing satellite (Dyczmons, 1978), atmospheric drag (Horedt, 1976), an encounter with another object in heliocentric orbit when both the Moon and the other object are within the Earth's sphere of influence (Ruskol, 1972a), changes in the masses of the Sun (Szebehely and Evans, 1980) or the Earth (Lyttleton, 1967), and the dissipation of tidal energy during the time of close approach (Alfvén and Arrhenius, 1969, 1972, 1976; Conway, 1982; Gerstenkorn, 1955, 1969; Singer, 1968, 1970, 1972; Singer and Bandermann, 1970; Winters and Malcuit, 1977). Disintegrative capture involves the tidal disruption of a proto-Moon during a close encounter with the Earth, where some of the resulting fragments could be more easily captured into bound orbit to later accumulate into the Moon (Alfvén, 1963; Alfvén and Arrhenius, 1969; Öpik, 1969, 1971). In binary accretion, the Moon forms in the Earth's orbit from an accretion disk simultaneously with the formation of the Earth (Harris, 1977, 1978; Harris and Kaula, 1975; Ruskol, 1961, 1963, 1972a,b,c, 1975). A formation by giant impact takes advantage of the fact that the last stages of accretion of the terrestrial planets involved large bodies (Wetherill, 1985). The result of such an impact is likely to leave sufficient material beyond the Roche radius R_R to accumulate into the Moon (Cameron and Ward, 1976; Canup, 2004a,b; Hartmann and

Treatise on Geophysics, *2nd edition*, (2015), vol. 10, pp. 559-604

Author's personal copy

Davis, 1975; Öpik, 1971). Only the giant impact origin of the Moon has survived rather intense scrutiny (see Boss and Peale, 1986 for a detailed rejection of the other schemes). The details of the Moon-forming impact are undergoing considerable study in an attempt to understand the observation that isotope ratios of several elements are identical on Earth and Moon. Whereas a giant impactor would generally have had an isotope distribution quite different from that of the Earth, the current Moon's surface and the Earth's mantle are made of the same stuff.

As formation of satellites in accretion disks is thought to be the dominant process for the regular satellites of the giant planets (regular satellites are those in nearly circular orbits coplanar with the planet's equatorial plane), we include a description of the processes and parameters appropriate for such accretion disks in **Appendix A**. In **Appendix B**, we outline a simple theory of gravitational tides and the dissipation therein that result in orbital and spin evolution. This evolution in many instances constrains the choice of processes and events involved in the origin of particular satellites. We do not include tables of orbital parameters or physical properties of the satellites, but refer the reader to the websites http://ssd.jpl.nasa.gov/?sat_elem and http://ssd.jpl.nasa.gov/?sat_phys_par for frequently updated total number of satellites with their orbital parameters at the first website and their updated physical properties in the second.

In **Section 10.17.2**, we develop the arguments supporting the giant impact origin of the Earth's Moon, which is necessary to account for the large angular momentum of the system along with a volatile and iron-poor moon. But in addition, we must account for the plethora of identical isotope ratios on the Earth and Moon. The equatorial orbits of Mars' satellites require their origin in a dissipative disk of debris, where means of accounting for a suspected grossly different composition of the satellites from Mars itself are discussed in **Section 10.17.3**. We describe recent attempts at understanding the origin of Jupiter's regular satellites by a generalization of binary accretion during the last stages of Jupiter's formation in **Section 10.17.4**. The constraints imposed by information obtained from the Galileo spacecraft mission guide these attempts, but all involve assumptions of various degrees of uncertainty. But considerable progress has been made recently, and a self-consistent and plausible theory of origin of Jupiter's regular satellites within the theory of the accretion of Jupiter itself is emerging.

In many ways, the Saturn system of regular satellites is much more complicated than Jupiter's, where nearly all of the mass in the system is in the single large satellite Titan, and small icy satellites and icy rings appear inside Titan's orbit where we might have expected large rocky satellites as in the Jupiter system. An additional complication is that the fraction of each satellite that is not ice is not a monotonic function of the orbital semimajor axis. The origin of Saturn's satellites is discussed in **Section 10.17.5**, where we indicate that this origin is among the least well understood in the solar system. The Uranian satellites are covered in **Section 10.17.6**, where the regular nature of these satellites in spite of a planetary obliquity of 97° implies that the disk from which these satellites formed was created in the giant impact that led to Uranus' large obliquity perhaps by disrupting an existing satellite system that was reassembled from the resulting equatorial disk. An alternative would be that the Uranian obliquity resulted from adiabatic processes that allowed the equatorial orbits of the satellites to be preserved, although such a process has not been found. Details of formation of the Uranian satellites in the collision scenario have been investigated but without detailed accretion processes. Neptune's satellite system, discussed in **Section 10.17.7**, is perhaps better understood than most. A plausible series of events centered around the essentially intact capture of the large retrograde satellite Triton account well for all of the observational properties of this system. The Pluto–Charon system (**Section 10.17.8**), like the Earth–Moon system, is characterized by a large specific angular momentum, which likely resulted from an oblique giant impact. The picture is complicated by the discovery of four additional small satellites coplanar with Charon's orbit, which are too distant to have been created simultaneously with Charon and their origin remains a mystery.

Minor satellites include the irregular satellites with large semimajor axes, eccentricities, and inclinations that orbit all of the major planets and the small close satellites often associated with rings of small particles. The former are distinguished by a capture origin (**Section 10.17.9**) and the latter by having been broken up and reassembled, probably several times, by collisions among themselves, with comets or with planetesimals passing through the planets' Hill spheres. The disintegrations of these close satellites are associated with supplying the small particles, which make up the observed rings.

Many asteroids and Kuiper Belt objects (KBOs) also have satellites, where the widely separated binaries probably result from capture (Goldreich et al., 2002) and those in close orbits most probably from collisions (Canup, 2005). But some small, close asteroid binaries may form from the rotational fission of a rubble pile, where the rotational acceleration comes from the thermal Yarkovsky-O'Keefe-Radzievskii-Paddock (YORP) effect (Walsh et al., 2008). Finally, Nesvorny et al. (2010) had proposed formation of the wide binaries of comparable mass components in the Kuiper Belt from gravitational collapse of a swarm of particles. The particles can be brought into the unstable swarms perhaps by assembly in turbulent eddies (Cuzzi et al., 2001, 2008; Johansen et al., 2006) or by a streaming instability (Johansen and Youdin, 2007; Johansen et al., 2007, 2009, 2011, 2012; Youdin and Johansen, 2007). We will not discuss the small body satellites beyond these assertions. Rotational fission, precipitation fission, and disintegrative capture among those failed processes offered for the formation of the Moon also fail to find any likely applications to the satellites of Mars, Pluto, and the major planets.

10.17.2 Earth–Moon System

As additional observational constraints are imposed, the ingenious theories of origin and evolution of the Moon become more complicated. Attempts to accommodate the constraints in modifications of the theories, albeit sometimes successful on the surface, introduce uncertainties and leave gaps in the evolution, which leave us without a completely definitive theory of origin. However, all theories keep the giant impact as the most likely beginning of the process leading to our current Earth–Moon configuration, and details described in this section may generate additional ideas.

The Moon's mass is 1.2% of the Earth's mass, which is the largest fractional mass of all the satellites in the solar system

Author's personal copy

with the exception of that in the Pluto–Charon system (if we ignore asteroid and Kuiper Belt satellites). In addition, the angular momentum of the system of $L_{EM} = 3.5 \times 10^{41}$ g cm^2 s^{-1} would lead to a 4 h rotation period of the Earth if the Moon were incorporated within. This large specific angular momentum is a key constraint on the origin. The expansion of the lunar orbit by tidal interaction with the Earth (Goldreich, 1966; Touma and Wisdom, 1994) is consistent with formation close to the Roche radius $R_R = 2.9R_E$, where R_E is the Earth's radius. Accretion of the Moon inside this distance would be prevented by tidal forces. At a density of 3.34 g cm^{-3}, the Moon is depleted in iron relative to the Earth, and samples of lunar material are depleted in volatiles with vaporization temperatures < 1300 K in solar nebula conditions (e.g., Jones and Palme, 2000). A thick layer of anorthositic crust whose constituents have a positive Eu anomaly is complemented with a Eu deficiency in the mare basalts. The thickness of the anorthositic crust coupled with the complementary Eu anomalies is consistent with the crystallization of the lighter anorthositic material in a few hundred kilometer deep magma ocean with the crystals floating to the top. The floating crystals are enriched in Eu leaving the magma, from which the mare basalts are later derived, deficient in this element (Taylor and Jakes, 1974; Wood, 1972; Wood et al., 1970). Limits on lunar contraction ($\lesssim 1$ km in radius) during cooling have been inferred from the lack of stress features on its surface, which has been interpreted as implying that the interior of the Moon was initially cold beneath the magma ocean (Solomon, 1986). While contraction of an initially hot Moon might be accommodated without the expected surface scarps (see the discussion in the succeeding text), the recent detection of subsurface gravity anomalies by the Gravity Recovery And Interior Laboratory (GRAIL) spacecraft also suggests an initially cool lunar interior (Andrews-Hanna et al., 2013) in contrast to an initially hot moon inferred from the giant impact origin.

With the exception of the volatiles, the lunar material and the Earth's mantle share a number of very close compositional similarities, where isotope ratios for several elements in lunar samples are identical to those on Earth within experimental uncertainty (Lugmair and Shukolyukov, 1998; Touboul et al., 2007; Wiechert et al., 2001; Zhang et al., 2012). The current angular momentum and the Moon's overall depletion in iron are consistent with the formation of the Moon by the impact of a differentiated Mars-sized body with a differentiated Earth (Canup, 2004a), which we shall refer to as the canonical impact. However, it is unlikely that such an impact alone can account for the isotopic similarities between the Moon and the silicate Earth, which has led to a suggestion of mixing of the disk vapor with the Earth's silicate atmosphere after the impact (Pahlevan and Stevenson, 2007) as well as to the recent exploration of alternative impact scenarios (Canup, 2012; Ćuk and Stewart, 2012; Reufer et al., 2012).

The formation of the Moon by giant impact was first published by Hartmann and Davis (1975) with a lunar-sized impactor that could account for the iron and volatile deficits, but Cameron and Ward (1976) realized that the grazing collision of a Mars-sized body could also account for the large angular momentum of the Earth–Moon system. Models of terrestrial planet accretion (e.g., Wetherill, 1985, 1992) showed that the Earth was likely to have undergone such Mars-sized impacts late in its accretion history and it is this scenario that has become the paradigm for the formation of the Moon. (For a delightful account of the development of the idea of giant the impact origin of the Moon by the four authors, see MacKenzie (2003), reviewed by Burns (2003).)

Hafnium–tungsten chronology has been used to estimate the time of the solidification of the lunar magma ocean. Radioactive ^{182}Hf decays to ^{182}W with a half-life $\tau_{1/2} = 9$ million years. Both elements are refractory, so the planets should have the same overall abundances as the solar system as a whole. The most recent analysis was limited to metals in returned lunar samples. This choice avoids augmentation of the ^{182}W from the capture of cosmic ray-generated neutrons by ^{181}Ta, since Ta is excluded from the metals. Any excess ^{182}W in the metals can thereby be attributed solely to production from ^{182}Hf decay. Recent experiments have detected no excess ^{182}W in metallic parts of 18 lunar samples, which implies that Hf was extinct at the time of solidification (Touboul et al., 2007). This result implies that solidification occurred more than 60 million years after the formation of the oldest constituents of meteorites, the calcium–aluminum-rich inclusions (CAIs). This is in contrast with earlier Hf–W chronology, which inferred a solidification age of 30–50 million years (Kleine et al., 2005), which was thought consistent with the timescale within which a giant impact was likely to have occurred. The late solidification of the Moon is supported by ^{207}Pb–^{206}Pb, ^{147}Sm–^{143}Nd, and ^{146}Sm–^{142}Nd chronologies performed on anorthositic material, which are consistent with crystallization as much as 200 million years after the formation of the CAIs (Borg et al., 2011). Since the solidification time of a magma ocean is ~10 million years (Elkins-Tanton et al., 2011), the giant impact may have occurred much later than previously assumed, or a scheme different from crystals of anorthositic material floating to the surface of a magma ocean to account for the anorthositic crust must be invoked. Contradicting the inference of a late formation of the Moon is the work of König et al. (2011), who find evidence for an early formation of the Earth's core associated with an inferred concurrent formation of the Moon. But the Earth's core probably formed long before the impact. Using Lu–Hf chronology, Taylor et al. (2009) argued for a rapid lunar differentiation implying an early formation. Still, Touboul et al. finding no excess ^{182}W in lunar metals is a rather clean result favoring a formation of the Moon after ^{182}Hf is extinct.

The angular momentum relative to the center of mass of the target (mass M_t) delivered by an impactor of mass $M_i = \gamma M_T$, where $M_T = M_t + M_i$ is the total mass, is (Canup, 2004a)

$$L_i = bM_T^{5/3} f(\gamma) \sqrt{\frac{2G}{\left(4\pi\rho_p/3\right)^{1/3}}} \left(\frac{v_i}{v_{esc}}\right) \approx 1.3 L_{EM} b \left(\frac{M_T}{M_E}\right)^{5/3} \left(\frac{\gamma}{0.1}\right) \left(\frac{v_i}{v_{esc}}\right) \quad [1]$$

where $b = \sin\xi$ is the impact parameter divided by $R_t + R_i$, the sum of the target and impactor radii; ξ is the angle between the line connecting the centers of M_t and M_i and the impact trajectory at the time of first contact, such that grazing incidence has $\xi = 90°$ or $b = 1$; G is the gravitational constant; ρ_p is the common density of M_t and M_i; v_i is the velocity of the impactor; $v_{esc} = \sqrt{2GM_T/(R_t + R_i)}$ is the mutual escape velocity, $f(\gamma) = \gamma(1-\gamma)\sqrt{\gamma^{1/3} + (1-\gamma)^{1/3}}$; and M_E is the Earth's mass. The complication of eqn [1] comes from introducing v_{esc} into the equation for angular momentum.

Author's personal copy

For the simplest case in which a single impact provides all the Earth–Moon system angular momentum, L_i must be comparable to or greater than that in the current Earth and Moon (L_{EM}). Some angular momentum will be lost in escaping material both from the initial impact and during the assembly of the Moon from the impact-generated disk. The magnitude of the former depends primarily on the impact velocity, while models of the Moon's accretion suggest that the latter would have been of order of a few $\times 10^{-2}\, L_{EM}$ (Kokubo et al., 2000; Salmon and Canup, 2012). The Earth–Moon angular momentum would have also been decreased by a few percent by tides raised on the Earth by the Sun that slow the Earth's rotation (e.g., Canup et al., 1999). For a given value of L_i, there is a multidimensional parameter space involving b, M_T, γ, and v_i that must be explored to find successful candidate impacts. Accommodation of values of $L_i \approx 2L_{EM}$ in more energetic impacts (Canup, 2012; Ćuk and Stewart, 2012) will be discussed later.

A successful lunar-forming impact produces a disk capable of later yielding an iron-depleted moon outside the Roche limit. This condition is met either when a lunar mass or more is placed into orbits outside R_R by the impact itself or when the disk has sufficient mass and angular momentum so that viscous spreading after the impact can deliver sufficient mass beyond R_R. Simulations of the disk's evolution typically find that a single moon accretes just outside R_R, with inner disk material driven into the planet and $<5\%$ of the disk's mass escaping from the Earth's orbit (Crida and Charnoz, 2012; Ida et al., 1997; Kokubo et al., 2000). A mass and angular momentum conservation argument can be used to estimate the mass of the moon (M_M) that would accumulate from a disk as a function of the initial disk mass (M_D), the initial disk angular momentum (L_D), the initial position of the moon ($a = \lambda R_R$), and the angular momentum of material that escapes or collides with the Earth. Based on results of disk simulations and with $\lambda = 1.3$, the predicted moon mass is

$$\frac{M_M}{M_D} \approx 1.9\left(\frac{L_D}{M_D\sqrt{GM_E R_R}}\right) - 1.1 - 1.9\left(\frac{M_{esc}}{M_D}\right) \qquad [2]$$

where M_{esc} is the escaping mass (Ida et al., 1997; Kokubo et al., 2000). This equation provides a simple way to determine whether a disk produced by a hydrodynamic simulation could later yield a lunar mass moon. It is physically invalid for cases that would give $M_M/M_D > 1$; these correspond to initial disks with specific angular momenta too high to be consistent with the assumption that the final moon forms at $1.3R_R$.

Smooth particle hydrodynamics (SPH) is the method most used in numerical exploration of the impact parameter space. SPH is a Lagrangian method that describes the colliding objects by a multitude of particles that are evolved due to self-gravity and pressure forces. Over time, the spatial resolution of such calculations has increased from $\sim 10^3$ particle simulations in Benz et al. (1986, 1987) to 10^5–10^6 particle simulations in recent works (e.g., Canup, 2004b, 2008, 2012; Canup and Asphaug, 2001; Ćuk and Stewart, 2012; Reufer et al., 2012). At the same time, there have been algorithm and equation of state (Melosh, 2007; see also the discussion in Canup, 2004b) improvements that have influenced the results. Alternatively, impacts can be modeled with a grid-based, Eulerian hydrocode (Canup et al., 2013; Wada et al., 2006). Simulations with SPH and the grid-code CTH yield broadly similar results, with the predicted protolunar disk mass and its angular momentum converging to typically within about 10% for simulations of very similar impact conditions (Canup et al., 2013).

SPH simulations show that for $L_i \approx L_{EM}$, those impacts that produce a relatively iron-poor disk having M_M (from eqn [2]) of at least a lunar mass typically have $b \sim 0.7$ ($\sim 45°$ impact angle), a low impact velocity $v_i/v_{esc} < 1.1$, and an impactor mass ratio $0.1 < \gamma < 0.15$ (a roughly Mars-sized impactor) for $M_T \sim M_E$ (Canup, 2004a; Canup and Asphaug, 2001). An impact involving a low-velocity, Mars-sized impactor with $L_i \approx L_{EM}$ has generally been considered the 'canonical' impact for forming the Moon, because it accounts for the main dynamical properties of the Earth and Moon and the Moon's relative lack of iron while relying on conditions that are probable during the late stages of terrestrial planet accretion, that is, large, randomly oriented impacts, with 45° being the most probable impact angle (e.g., Agnor et al., 1999). However, because such cases yield a disk formed primarily from material originating from the impactor, the observed identical isotope ratios for lunar and Earth mantle materials cannot prevail.

It has been known for decades that the Earth and Moon share identical oxygen isotope compositions that are distinct from those of most meteorites and Mars (e.g., Wiechert et al., 2001). This would be consistent with the Moon's composition if the Moon formed directly from the Earth's mantle. If instead the disk (and the Moon) formed primarily from impactor material, an impactor with an identical oxygen composition as the Earth would then be required. Because oxygen compositions vary with heliocentric position, it was suggested that this similarity would be expected if the impactor formed near the Earth (e.g., Belbruno and Gott, 2005; Wiechert et al., 2001). This seemed to agree well with the canonical impact, which required a low impact velocity and therefore an orbit close to that of the Earth. However, this explanation was brought into serious question by Pahlevan and Stevenson (2007), who argued that it is very improbable that the impactor and Earth would share the same oxygen composition even if they were in close orbits, given the degree of radial mixing predicted by planet accretion models. Using the impacts predicted in one of Chambers' (2001) terrestrial accretion simulations, together with the assumption that the oxygen composition of planetary embryos varied linearly with initial heliocentric distance, they estimated that the typical difference between the composition of an impactor and its final target planet would have been comparable to the current difference between Earth and Mars, which is about 50 times greater than the allowable difference in the Earth–Moon oxygen compositions. A moon accreted from a disk composed primarily of impactor-derived material would then be expected to have a substantially different oxygen composition than the Earth. Furthermore, the isotopic composition of multiple other elements now also appears identical in the silicate Earth and Moon, including chromium (Lugmair and Shukolyukov, 1998), titanium (Zhang et al., 2012), tungsten (Touboul et al., 2007), and silicon (Georg et al., 2007). Georg et al. pointed out that Si partition into the core is a high-pressure process that depletes the heavy isotopes more than the light ones and that requires at least 0.15 Earth masses to become

Author's personal copy

effective. That means that even if the oxygen, chromium, and titanium isotopic compositions were identical in the Earth's mantle and the Mars-sized impactor, the silicon composition of the impactor would likely not be because of partitioning of the heavier Si isotopes into the Earth's core. This supports the Pahlevan–Stevenson assertion that Earth and impactor could not have identical isotopic compositions before the impact.

As a solution, Pahlevan and Stevenson (2007) proposed that the vapor components of the disk and the Earth's silicate atmosphere diffusively mixed after the impact via vigorous convection, allowing the disk to compositionally equilibrate with the Earth before the Moon forms. However, as stressed by Melosh (2009), for this to occur, one must assume that the convection mechanism does not also transport angular momentum at the same efficiency, because if it did, the disk would spread into the Earth on a similar timescale and diffusion from the Earth's atmosphere would have to operate upstream against the flow of this incoming material. Still, this may be a valid assumption, since thermal convection is now thought to be inefficient in transporting angular momentum in disks (Stone et al., 2000). Equilibration is estimated to require $\sim 10^2$–10^3 years, so that the Moon's accumulation must be delayed by this amount of time after the impact (Pahlevan and Stevenson, 2007). The mixing must involve the whole of the disk to equalize the planet-disk compositions, which may be difficult if a substantial fraction of the Moon's mass is initially outside the Roche limit (e.g., Salmon and Canup, 2012). It is not clear that vapor equilibrium of refractory elements in a canonical impact (including notably titanium) is possible, although if the disk lasts for $\geq 10^2$ years, there may be sufficient time for even refractory elements in the magma layer to equilibrate with the disk's vapor atmosphere (Zhang et al., 2012). In addition, the titanium isotope composition of Mars appears much closer to the Moon's composition than is true for oxygen. Within current uncertainties, the Mars–Earth Ti difference may only be about a factor of 2 larger than the Moon–Earth difference (e.g., Canup, 2012, supplemental material), so that a disk composed primarily of material from a Mars-composition impactor might not require much equilibration of Ti to be consistent with the current Moon. Thus, titanium does not provide as strong a constraint on equilibration as oxygen, for which the Earth–Mars difference is much larger.

Such issues motivated additional impact studies with L_i close to L_{EM} with varying assumptions to see if a disk with planet like composition could be produced. Comparisons between simulations performed with SPH versus CTH found no systematic influence of simulation method on predicted disk composition (Canup et al., 2013). To quantify the compositional difference between the silicate portions of the disk and planet, one can define a deviation percentage (after Reufer et al., 2012),

$$\delta f_T = \left[\frac{F_{D,tar}}{F_{P,tar} - 1}\right] \times 100 \qquad [3]$$

where $F_{D,tar}$ and $F_{P,tar}$ are the mass fractions of the silicate portions of the disk and the planet derived from the target's mantle. Identical disk-planet compositions have $\delta f_T = 0$, whereas a disk that contains fractionally more (less) impactor-derived silicate than the final planet has $\delta f_T < 0$ ($\delta f_T > 0$). For a Mars-like composition impactor, consistency with the Earth–Moon oxygen composition requires $|\delta f_T|$ less than a few percent (e.g., Canup, 2012).

Prior works had assumed a target and impactor that were not rotating prior to their collision, an almost certainly incorrect assumption. Canup (2008) examined the effect of pre-impact rotation in impacts that leave a planet-disk system with an angular momentum, L, comparable to L_{EM}. She found that a target proto-Earth with a pre-impact retrograde rotation (i.e., in the opposite rotational sense as the impact itself) allowed for somewhat larger impactors and produced disks containing up to 40% target material. Reufer et al. (2012) considered higher-velocity, more head-on impacts in which a substantial fraction of the impact angular momentum escapes, so that L_i may be much greater than L_{EM} and still leave a planet-disk system with $L \sim L_{EM}$. They found that collisions with $1.2 \leq v_i/v_{esc} \leq 1.3$, $0.15 \leq \gamma \leq 0.2$, and $\xi = 30$–$40°$ produced appropriately massive disks that contained up to $\sim$60% target material.

The canonical impacts typically have $-90\% \leq \delta f_T \leq -70\%$, whereas the successful cases (producing a disk leading to a $\sim 1 M_M$ Moon) in Canup (2008) have $\delta f_T \leq -50\%$ and those in Reufer et al. (2012) have $-40\% \leq \delta f_T \leq -30\%$. The more target-rich disks produced by Canup (2008) and Reufer et al. (2012) have values of $|\delta f_T|$ far greater than the required few % that would account for the Earth–Moon oxygen isotope similarity.

These failures to obtain a disk made up almost entirely of Earth mantle material have motivated proposals for additional modifications of the impact scenario, where the angular momentum of the resulting system is no longer constrained to be near L_{EM}. Ćuk and Stewart (2012) chose high-velocity, equatorial plane impacts ($v_i = 15$–25 km s^{-1}, 1.5–3 times v_{esc}) onto rapidly spinning Earth-mass targets ($M_t = M_E$) (periods 2.3–2.7 h corresponding to angular momenta of 1.9–$3.1 L_{EM}$), where $0.026 \leq M_i/M_t \leq 0.1$. The requirement of a near-fission rotation rate of the Earth prior to impact may be quite restrictive. Rapidly rotating planets are seen in N-body simulations of late-stage accretion (e.g., Agnor et al., 1999). However, such models tend to overestimate planetary rotation rates because they assume pure mergers and ignore the possible ejection of mass and angular momentum during large impacts. The most successful outcomes in the Ćuk–Stewart simulations, where the disk was sufficiently Earth mantle-dominated to satisfy the isotopic constraints (with $|\delta f_T|$ $\sim$several percent) and the predicted mass of the accreted satellite was near M_M, occurred for head-on or slightly retrograde impacts onto the higher angular momentum initial targets with impact velocities ~ 1.5–$2.5 v_{esc}$. In the successful Ćuk–Stewart impacts, the Earth rotation period is reduced, but the total remaining angular momentum is still $\gtrsim 2 L_{EM}$.

Canup (2012) reported simulations successful in producing a disk with the same composition of the Earth's mantle by colliding nonrotating targets and impactors of similar mass ($M_i/M_T \geq 0.4$) at low velocity. Here, the impact parameter was varied over a wide range yielding $0.35 \leq b \leq 0.7$, which determined the angular momentum of the Earth + disk and yielded a disk of sufficient mass to produce a Moon mass satellite. The silicate mantles of target and impactor were well mixed by the impact, so the resulting Earth mantle and disk had essentially the same silicate composition to satisfy the isotope constraints

Author's personal copy

(Nakajima and Stevenson, 2013). The most successful Canup examples yielded $\delta f_T = -0.8\%$ and -0.3%, predicted satellite masses of 1.64 and $1.09 M_M$, but left the Earth spinning with period of about 2 h with total angular momenta of 2.45 and $2.22 L_{EM}$.

The kinetic energy of both the new impact scenarios is comparable (although the Canup model involves a much larger impactor, the Cuk and Stewart model involves a higher impact velocity) and is about a factor of three larger than that of the canonical impacts. Such violent events may have difficulty explaining the recent inference that the Earth's mantle may not have been thoroughly mixed (Touboul et al., 2012), although Ćuk and Stewart (2012) argued that at least in the case of their impacts, this may result from nonuniform heating during the collision. In general, relaxing the angular momentum constraint in two different giant impact scenarios has allowed disk and Earth mantle to have compositions that are sufficiently alike in both the Ćuk and Stewart (2012) analysis and that of Canup (2012) to satisfy the isotope constraints in oxygen, titanium, and chromium (see succeeding text for the discussion on tungsten and silicon), but both leave the system with excess angular momentum that must be eliminated.

Ćuk and Stewart (2012) offered an ingenious scheme for reducing the system angular momentum. As the Moon recedes from the Earth due to torques from tides raised on the Earth, it is captured into the evection resonance with the Sun. The evection resonance occurs when the precession period of the Moon's perigee is commensurate with the Earth's orbital period about the Sun. For an Earth–Moon system with an angular momentum $\sim L_{EM}$, the resonance occurs when the Moon's semimajor axis is $a \approx 4.6$ Earth radii. This resonance had been previously investigated by Touma and Wisdom (1998), who found that it could play a key role in setting up capture into a resonance involving the lunar inclination that could explain the Moon's initial orbital inclination (see succeeding text). For the more rapidly spinning Earth ($\sim$2.5 h day), corresponding to a system angular momentum $\sim$2–$2.5 L_{EM}$, the Earth's increased oblateness causes the location of the resonance to shift outward to $a \approx 7$ Earth radii. At this larger distance, the rate of the Moon's orbital expansion due to tidal interaction with the Earth is slower, which increases the likelihood of capture into evection. In the evection resonance, the line of apsides of the lunar orbit remains perpendicular to Earth–Sun direction. Angular momentum is tidally transferred to the Moon orbit from the Earth and from there to the Earth–Moon orbit about the Sun because of resonance. The orbital eccentricity increases to a value near 0.6 for $A_Q \approx 1$ (see eqn. 4) as the Moon's orbit expands until the expansion of the Moon's orbit stalls due to a balance between Earth tides (which cause a to expand) and satellite tides (which cause a to contract for high-eccentricity orbits). As long as the evection resonance is maintained, angular momentum of the Earth–Moon system is being reduced, and in the Ćuk and Stewart simulations, the Moon's prolonged residency in the resonance removes the excess angular momentum until the resonance is broken and the Moon commences its tidal evolution to its current distance from the Earth. However, there are some caveats in this scheme for removing the excess angular momentum.

A measure of the relative tidal dissipation in the Moon and Earth is given by the ratio

$$A_Q = \frac{k_{2M}}{k_{2E}} \frac{Q_E}{Q_M} \frac{M_E^2}{M_M^2} \frac{R_M^5}{R_E^5} \qquad [4]$$

where the subscript Q indicates the constant Q tidal model. The successful cases of Ćuk and Stewart, where systems with $L \approx L_{EM}$ were obtained, occurred only for a very narrow range of A_Q near 1. If lunar tides are too strong relative to Earth tides, the system leaves the resonance before sufficient angular momentum is removed, whereas if the lunar tides are too weak compared to the Earth tides, the lunar orbital eccentricity reaches overly high values that destabilize the resonance (Ćuk and Stewart, 2012). Both the Earth and Moon are likely to be nearly all molten initially with values of k_2 close to the fluid values, and $A \sim 1$ implies $Q_M \approx 10 Q_E$ (little dissipation in the Moon) from eqn [4]. There is no reason for Q_M to be so much larger than Q_E, but see the succeeding text.

The reader should be warned that the results of the calculations depend on the tidal model. Ward and Canup (2013), using the constant Δt ($Q \propto 1/\omega$) described in **Appendix B**, were also in some cases able to remove all the excess angular momentum while the Moon is trapped in the evection resonance. In their case, the measure of dissipation in the Moon to that in the Earth is given by

$$A_{\Delta t} = \frac{k_{2M}}{k_{2E}} \frac{\Delta t_M}{\Delta t_E} \left(\frac{M_E}{M_M}\right)^2 \left(\frac{R_M}{R_E}\right)^5 \qquad [5]$$

and a value of $A_{\Delta t} \approx 9$ is used in a successful case that removes nearly all of the excess angular momentum. Recall for this tidal model that $Q = 1/\omega \Delta t$ (**Appendix B**), so we could replace $k_2 \Delta t$ in eqn [5] with $k_2/Q\omega$, where Q is now that value appropriate for the frequency ω. The dominant tidal frequency in the Moon is close to the mean orbital motion n, whereas that in the Earth is the much larger $\dot{\psi} - n$, where $\dot{\psi}$ is the angular velocity of the Earth. Then, $k_{2M}\Delta t_M/(k_{2E}\Delta t_E) = (k_{2M}/k_{2E})(Q_E/Q_M)\left(\left(\dot{\psi} - n\right)/n\right)$, where again the Qs are those appropriate for the frequencies they multiply. But for constant Δt, $Q_E(\omega) = Q_E(n)(n/\omega)$, so $A_{\Delta t} \approx 10 Q_E(n)/Q_M(n)$ if the k_2s are comparable. The Ward–Canup result with $A = 9$ would then require the Qs for the Moon and Earth evaluated at the frequency n to be comparable, which contrasts with the constant Q model. Another reservation is that the large increase in the eccentricity during the angular momentum transfer could lead to a drastic increase in the lunar dissipation and the value of A_Q would drift far outside the narrow range for stability of the evection resonance (Wisdom, private communication). There may be a similar restriction on $A_{\Delta t}$. We should therefore regard both the Ćuk–Stewart and the Ward–Canup results as indicative that removal of the angular momentum in the evection resonance could be possible with the reservation that at least the constant Q results have to be relatively finely tuned, in which tuning might be frustrated by the increased dissipation in the Moon during the high-eccentricity phase, and that neither tidal model represents the behavior of real materials (e.g., Castillo-Rogez et al., 2011).

There are thus two viable schemes for creating a moon with identical isotopes to those of the Earth's mantle that are consistent with a giant impact origin of the Moon (Canup, 2012; Ćuk and Stewart, 2012) without requiring postimpact equilibration. Although both require a system angular momentum close to twice L_{EM}, at least a process has been demonstrated that could reduce the angular momentum to a value close to

Author's personal copy

L_{EM}, but not without important caveats. What is less clear is whether these models can account for the Earth–Moon similarities involving elements that partition strongly into metals and therefore can be affected by core formation, including tungsten and silicon. Whether the disk is formed from the target's mantle (as in the Ćuk and Stewart impacts) or from an approximately equal mix of target and impactor mantles (as in the Canup impacts), the core of the impactor is absorbed by the planet, while the disk contains little iron. Mixing and equilibration of the impactor's core with the target's mantle would tend to alter the planet's tungsten and silicon compositions relative to that of the disk (and the resulting Moon). However, it is possible that rapid merging of the impactor's large core with the Earth's core precludes efficient silicate–metal equilibration, because the latter requires emulsification down to centimeter scales (Dahl and Stevenson, 2010; Rubie et al., 2003).

Although the overall compositional and dynamical properties of the current Earth–Moon system appear consistent with these latest versions of giant impact origin, the transition from the distribution of material following the impact to the accreted Moon a little outside the Roche radius R_R that is also consistent with current observations remains less well understood. An impact-produced disk contains silicate magma and vapor, with initially 10–30% vapor by mass for canonical impacts and up to 90% vapor by mass for the high angular momentum impacts. Much or perhaps most of the disk mass initially orbits within the Roche limit. Interior to R_R, clumps among the magma formed by local gravitational instability are sheared apart by planetary tidal forces (Ward and Cameron, 1978). This generates a large viscosity that dissipates heat and causes the Roche interior disk to spread inward onto the planet and outward past R_R on a timescale τ_v. Material from the inner disk that spreads beyond R_R, together with material that was placed into orbits outside R_R by the impact itself, provides the material that can accrete into the Moon.

Early estimates that assumed a pure magma disk (no vapor) suggested $\tau_v \sim O(1)$ year (Ward and Cameron, 1978). This timescale is consistent with that seen in N-body lunar accretion simulations of a vapor-free disk of particles that find that the Moon accretes in less than a year (Crida and Charnoz, 2012; Ida et al., 1997; Kokubo et al., 2000). However, there is an inconsistency in modeling the protolunar disk as vapor-free, because the energy dissipated as a disk of molten or solid particles viscously spreads is sufficient to vaporize the disk. Vaporization renders the disk gravitationally stable unless the energy is radiated away. Silicate vapor and liquid phases coexist when the disk's photospheric temperature is of order ~2000 K. If the phases are vertically well mixed, the radiation rate of the disk at this temperature can limit the disk's spreading to a much longer timescale, $\tau_{rad} \sim O(10^2)$ years (Thompson and Stevenson, 1988). Thompson and Stevenson (1988) suggested that a lower viscosity consistent with this timescale could result as the disk hovers on the brink of local gravitational instability with a sound speed just below the critical value for marginal stability. The latter can occur in a two-phase disk if the gas mass fraction is very low – only a few percent – with the bulk of the material comprised of liquid droplets that are vertically well mixed with the vapor. An alternative, stratified disk structure can result if the liquid instead settles, forming a midplane magma layer surrounded by a vapor disk atmosphere (Ward, 2012). In either case, the lifetime of the Roche interior disk is $\geq 10^2$ years.

The $\leq$1-year accretion timescale predicted by N-body lunar accretion models is at odds with these much longer disk lifetimes. More recent simulations that pair an N-body model for material outside the Roche limit with a fluid model for an inner disk that persists for $\sim 10^2$ years find that the Moon takes $\sim 10^2$ years to complete its accretion (Salmon and Canup, 2012), which may provide sufficient time for the equilibration of oxygen (Pahlevan and Stevenson, 2007). Accretion of the Moon in $\leq$1 year implies a fully molten Moon, whereas accretion over 10^2 years may allow for partial cooling for favorably chosen conditions (e.g., Pritchard and Stevenson, 2000). A molten Moon is not consistent with the most straightforward interpretation of the lack of global-scale thrust faults on the Moon, which implies limited contraction and an initially cool interior (Solomon, 1986). There are uncertainties in the properties of the lunar crust and in the thermal history that seem to allow ranges of parameters and processes that could accommodate an initially molten Moon in spite of the lack of observable indications of a significant radius change (Pritchard and Stevenson, 2000). But recently, the GRAIL spacecraft has identified a system of subsurface gravity anomalies hundreds of kilometers in length that are interpreted as dikes formed during the expansion of the early lithosphere (Andrews-Hanna et al., 2013). Such an early period of expansion is predicted for an initial Moon with a cool interior and an overlying magma ocean as the interior warms during the Moon's first billion years (Solomon, 1977). The GRAIL data thus provide additional evidence for an initially cool lunar interior. Any surface indication of the expansion could have been erased by the late heavy bombardment (LHB) on the Moon, where the existence of the dikes seems to imply that the LHB was not so intense as to create a second magma ocean. But an initially cold Moon is not consistent with the results of the giant impact.

The most likely consequence of the accretion of the Moon from a disk of material generated by a giant impact is a hot, at least partially molten Moon orbiting just outside R_R in the Earth's equatorial plane. Before we leave the Earth–Moon system, we must then account for the Moon's current orbital properties under the assumption that the semimajor axis of the Moon's orbit was increased from its initial to the current value by tidal torques. By integrating backward in time from the current state, Goldreich (1966) found that the Moon's orbit had an inclination to the Earth's equator of about 10° when the lunar semimajor axis was approximately $10R_E$. This result was confirmed by Touma and Wisdom (1994). The 10° inclination is not consistent with the Moon having formed in the Earth's equatorial plane from the debris generated by a giant impact. Moving the Moon from an equatorial orbit to one inclined by 10° is called the 'inclination problem.'

There are four proposals for accomplishing this bit of dynamics. If the Earth had a significant spin before impact like the Ćuk and Stewart initial condition, and a substantial fraction of the Moon remained in a coherent clump after the impact, the latter's orbit could have been inclined to the equatorial plane and subsequent accretion of the remaining lunar mass could leave it inclined (Canup, 2004a). But the equilibration of the isotope ratios would have been frustrated, even if the unlikely event occurred. Alternatively, an additional impact

Author's personal copy

on the Earth after the Moon-forming event could have altered the alignment of the Earth's spin. But this event would have had to have occurred while the Moon was very close (Ward, 2002), and it is not clear how the Moon would not have been contaminated with additional material with different isotope ratios – even if the first impact established equal ratios. The remaining two hypotheses do not obviously frustrate the identical isotope condition, but both still have caveats. They both involve resonant interactions to increase the initial Moon's orbital inclination – one with orbital resonances between the Moon's motion and the Sun (Touma and Wisdom, 1998) and the other with resonances between the Moon's motion and a remnant disk inside R_R (Ward and Canup, 2000).

In an integration that produces the current configuration when the Moon has reached its current distance from the Earth, Touma and Wisdom (1998) started the Moon in the equatorial plane of the Earth at a separation of $3.5R_E$ with an eccentricity of 0.01. The initial obliquity of the Earth is 10° and the initial Earth rotation period is 5 h. There is no circumplanetary disk present. Realistic rates of tidal evolution are used in the symplectic integrations that include the entire chaotic solar system, and dissipation in the Moon is included at a variety of dissipation rates. A tidal model was used employing the constant time lag as discussed in **Appendix B** but with the Mignard (1981a) formulation. The evection resonance is encountered when the Moon is at about $4.6R_E$, where the period of the periapse motion of the lunar orbit relative to an inertial reference is near one year. This resonance is called the evection resonance because the same term in the disturbing function gives rise to the 1.3° amplitude, 31.8 day periodic variation in the Moon's mean longitude called the evection. Capture into the evection resonance is certain if the eccentricity is below about 0.07 as the resonance is approached and if the rate of tidal evolution is sufficiently slow. With the assumed parameters, capture occurs and the eccentricity grows rapidly to large values, where the maximum value reached before the system escapes the resonance is determined by the value of $A_{\Delta t}$ defined by eqn [5]. For $A_{\Delta t}=0$ (no dissipation in the Moon), the maximum eccentricity is about 0.5 before escape and the eccentricity continues to climb after escape from the resonance because of tides raised on Earth and no dissipation in the Moon (Goldreich, 1963). For $A_{\Delta t}=10$ (high dissipation in the Moon), the maximum eccentricity is only about 0.15. For $1 \le A_{\Delta t} \le 10$, the energy dissipated in the Moon in about 8000 years is in the range $\sim 2\times10^{35}$–1.5×10^{36} ergs, which could lead to substantial melting (Touma and Wisdom, 1998).

After escape from the evection resonance, the continued expansion of the orbit further decreases the prograde motion of the orbit periapse and twice the time derivative of the evection resonance variable plus the retrograde motion of the lunar orbit node approaches zero near $6R_E$. The term in the Hamiltonian corresponding to this resonance has $e_M^2 i_M$ in the coefficient, but this resonance affects the inclination more than the eccentricity. Touma and Wisdom named this resonance the eviction – changing the e in evection to i to emphasize the inclination and noting that this resonance 'evicts' the Moon from an equatorial orbit. If A is not too large, the eviction resonance is approached with high eccentricity in the wrong direction for capture. Passage through the resonance leaves the eccentricity large and excites an inclination of 2°–3°. If A were now to increase drastically – perhaps because the continued high eccentricity has partially melted the interior – the dissipation in the Moon becomes so high that the semimajor axis is decreased as the eccentricity is reduced. This takes the system through the eviction resonance from the other direction where capture and subsequent evolution force the inclination to values between 9° and 13°. Escape from the resonance is effected because of the continued decrease in e_M, but the remnant inclination is preserved. Subsequent evolution brings the Moon to the current configuration. The Touma and Wisdom analysis was the first that allowed the Moon to evolve from an equatorial orbit to its current configuration from the effects of tidal dissipation alone.

An important constraint on the Touma–Wisdom model is that the Moon needs to begin its tidal evolution cold, which is contrary to our conclusion earlier that the Moon most likely started hot in at least a partially molten state (e.g., Pritchard and Stevenson, 2000). In addition, a more rapidly rotating initial Earth as advocated in Ćuk and Stewart (2012) moves the position of the evection outward, so that it takes longer for the Moon to encounter the resonance. However, from the expression for the tidal variation in the mean motion (Peale, 1988)

$$\frac{dn}{dt} = -\frac{9}{2}k_2\frac{M_M}{M_E}\frac{R_E^5 n^{16/3}}{(GM_E)^{5/3}}\frac{1}{Q_E} \qquad [6]$$

the timescale for expanding the lunar orbit from $3R_E$ to $7R_E$ is only $\sim 32Q_E$ years, where Q_E is likely to be < 10 for a partially molten Earth. The timescale for solidifying a magma ocean on the Moon is 1000 years for the first 80%, appropriate to maintaining melt on the surface from foundering of the cooled material, and 10 million years for the remainder appropriate for the conductive timescale of an anorthositic stagnant lid (Elkins-Tanton et al., 2011). The further distance of the evection resonance for the rapidly spinning Earth does not allow sufficient time for significant cooling of the Moon. The probable hot Moon remains a caveat in the Touma–Wisdom solution to the inclination problem. But if the inference from the GRAIL data that the Moon was initially cold can be supported by a believable evolution of the system, perhaps the Touma–Wisdom ideas can gain credibility.

The scheme involving disk interactions to change the Moon's inclination (Ward and Canup, 2000) depends on the coexistence of the Moon with a 0.5–1 lunar mass Roche interior disk for 10–100 years. Waves in the disk are generated at locations where the ratio of the local mean motion to that of the Moon involves two small integers. Torques that result from the gravitational interaction of the Moon with the distribution of mass in the wave structures can affect the lunar orbit. Such mean motion resonances (MMRs) from planar interactions are called Lindblad resonances, and they generate spiral density waves. Spiral bending waves are associated with the inclination of the Moon's orbit, where material in the disk is lifted out of the plane (e.g., Shu, 1984). When the Moon is just outside R_R, there are many MMRs in a disk that extends nearly to R_R. Torques from the waves generated in the disk force the Moon to larger semimajor axes, while the back reaction reduces the angular momentum contained in the disk. The resonances thereby move out through the disk and eventually leave its outer edge until the last strong one to leave the disk is the 3:1 inner vertical resonance (IVR) when the Moon's semimajor axis is a little over twice the disk radius.

Author's personal copy

The effect of the torque $T_{IVR} \propto \sigma_g \sin^2 I$ from the so-called bending waves, where I is the orbital inclination for the Moon and σ_g is the surface mass density of the inner disk, in addition to contributing to the growth of the semimajor axis a of the Moon's orbit, is an increase in the inclination at the rate (Ward and Canup, 2000)

$$\frac{dI}{dt} = \frac{T_{IVR}}{M_M a^2 \Omega_K \sin I}\left(\frac{3}{2}\cos I - 1\right) \quad [7]$$

where M_M is the lunar mass and Ω_K is the Kepler mean motion. With a disk mass of $0.75M_M$ and an initial $I = 1°$, inclinations reached 12.3° and 14.5° as the Moon nears $6R_E$ for initial disk spreading times of 37 and 50 years, respectively (Ward and Canup, 2000). The growth of I stops if either the resonance moves outside the disk or the disk is depleted. The disk interaction mechanism requires control of the evolution by the 3:1 vertical resonance. Other resonances, if still present, can lead to inclination damping (e.g., Borderies et al., 1984; Ward and Hahn, 1994). However, the Lindblad resonances invoked by Borderies et al. to damp the inclination are out of the disk when the 3:1 vertical resonance is near the disk edge, and the co-orbiting vertical resonances used by Ward and Hahn are at the position of the Moon, which is outside of the disk. Once the inclination is at 10° near the Earth, tidal evolution can carry the Moon to the current distance with the orbit inclined by about the observed 5° relative to the ecliptic. Whether the Ward–Canup scheme would be as effective in a full model of the Moon's accretion has not yet been assessed. For example, it is not clear that a disk of sufficient mass could even exist after the Moon was formed or could persist long enough to increase the lunar inclination to the necessary value given the very high efficiency of accretion. Ward and Canup considered the IVR torque due to the total inner disk surface density in their analysis. This inherently assumes that the waves in both the disk's gas and magma layers damp sufficiently. If the gas waves do not damp sufficiently, then the torque could be dominated by that of the magma layer, an issue that should be addressed by future models. The Moon's resonant interactions with a Roche interior disk would generally predate its encounter with the evection resonance. An inner disk contributes to the lunar perigee precession, shifting the location of evection outward. The disk mass must be reduced to $< 0.1M_M$ before the evection resonance moves inside $10R_E$ where it can encounter the Moon (Canup, 2004a), thereby making the Ward–Canup and Touma–Wisdom schemes mutually exclusive.

If the Moon is partially molten in the Ward–Canup scenario, the high dissipation limits the eccentricity growth in a subsequent passage through the evection resonance as the disk is dissipated and the lunar orbit continues to expand, even if trapped for a time (Touma and Wisdom, 1998). The passage through the eviction resonance is in the wrong direction for capture so the inclination will not change significantly after the Ward–Canup mechanism has brought it above 10°.

The major recent change in the giant impact paradigm has been the accommodation of the identical isotope ratios in Earth and Moon. Two successful modifications of the giant impact scenario involving an initial value of system angular momentum $\sim 2L_{EM}$ have been proposed, and there may be a way to remove the excess angular momentum. But getting from the disk of essentially identical isotope composition as the Earth's mantle to the Moon at $\sim 10R_E$ with orbit inclined 10° to the equator has many uncertain steps and generally has a deficiency of investigations. The reader can identify several processes needing more thorough investigation, and even those processes that are reasonably well understood have uncertainties that remain to be resolved. For example, how can we reconcile the initially hot Moon inferred from the giant impact origin and rapid accretion with those observations consistent with an initially cold moon? We are sure that imaginative minds will develop new ideas perhaps motivated or constrained by new data on the properties of the Moon that must be produced in a complete theory of origin.

10.17.3 Mars System

Mars' inner satellite Phobos is located at a distance of 2.77 Mars radii (R_M) from the center of Mars, which is well inside the corotation radius of $\sim 6.0R_M$, and the outer satellite, Deimos, is just outside the corotation radius at $6.91R_M$. The tides raised on Mars thus cause Phobos to be spiraling toward Mars and Deimos to be spiraling away. In fact, Phobos is inside the Roche radius for a density of 1.9 g cm^{-3} and would be torn apart by tidal forces if it were a fluid. It needs shear strength of only 10^5 dynes cm^{-2} to resist disruption (Yoder, 1982) – a loose rubble pile would survive (Dobrovolskis, 1982; Soter and Harris, 1977).

Determinations of the dissipation function Q (**Appendix B**) of Mars are $O(100)$ (Bills et al., 2005; Duxbury and Callahan, 1981; Shor, 1975; Sinclair, 1978) from observations of the secular acceleration of Phobos' orbital mean motion. If we choose a constant value of $Q_M = 100$ with Love number $k_M = 0.16$ (Bills et al., 2005), Deimos' orbit could have expanded by < 200 km in 4.6×10^9 years. Phobos' initial semimajor axis would have been near $5.6R_M$ under the same assumptions. The rotation period of Mars has been essentially unaffected by the exchange of angular momentum with the satellites and would have been increased by only about 10 min due to solar tides. Deimos has essentially its initial orbit, and Phobos having started inside the corotation radius is consistent with the measured current value of $Q_M \approx 100$.

Both Phobos and Deimos are synchronously rotating; Deimos would have reached this state in $< 10^8$ years from an unlikely small initial spin period of 4 h, where a rigidity of 5×10^{11} dynes cm^{-2} and $Q = 100$ were assumed. Under the same assumptions, Phobos would have reached this state in $< 10^5$ years at its current separation from Mars and in $< 10^7$ years at its likely initial separation near the corotation radius. (See Peale (1977) for a detailed discussion of the rotation histories of all of the satellites known at that time.)

The two satellites of Mars are in nearly circular equatorial orbits and this argues for their accretion in situ from a debris disk in the equatorial plane – like the regular satellites of the major planets. However, the density of Phobos of 1.88 ± 0.07 g cm^{-3} determined from the 5-day rendezvous by Soviet spacecraft Phobos 2 with $GM_P = 7.22 \pm 0.05 \times 10^{-4}$ km^3 s^{-2} (Kolyuka et al., 1990) coupled with the volume of 5748 ± 190 km^3 from Thomas (1993) has been verified by the analysis of the flybys of the Mars Express spacecraft (1.876 ± 0.020 g cm^{-3}) (Andert et al., 2010). The density of Deimos is 1.48 ± 0.22 g cm^{-3} (Rosenblatt, 2011). Early evaluations of albedos near 5% and the reflection spectra are consistent with carbonaceous chondritic material (Pang et al.,

Author's personal copy

1978; Pollack et al., 1978). But this interpretation has been challenged by Rosenblatt (2011), who finds the spectra to be inconsistent with meteoritic material, whereas the thermal IR is consistent with silicate material. Nevertheless, these early interpretations of the spectra and the fact that the densities are much less than the mean density of Mars of 3.9 g cm^{-2} lend support to the suggestion that the satellites are composed of material that did not originate in the vicinity of Mars, although porosity is another means of reducing the density (25–45% for Phobos, 40–60% for Deimos, Rosenblatt, 2011). This interpretation inspired the consideration of intact capture (Mignard, 1981b; Pollack and Burns, 1977), but Szeto (1983) showed several seemingly insurmountable inconsistencies with the capture hypothesis independent of the impossibility of relaxing the captured satellites into the circular equatorial orbits where they are found. If the debris were indeed carbonaceous chondritic, which is not at all certain, one possible way it could have gotten into orbit about Mars would be the shattering of a planetesimal that was formed in the asteroid belt region of the nebula when it collided with a denser object already in orbit about Mars. Samples of both Phobos and Deimos would tell us if such a contrived origin were necessary. Alternatively, a giant impact on Mars could have launched a disk of material that would settle to the equatorial plane (Craddock, 2011). Craddock pointed out basins on Mars that could be the signature of such an impact. Whether an impact consistent with such basins could have produced a sufficiently extended disk to produce Phobos and Deimos at their initial locations close to the current corotation radius has not been determined. If an accretion disk is indeed the birthplace of Phobos and Deimos, many Phobos-like objects may have impacted Mars, leaving a distribution of craters with asymmetrical or elongate ejecta (Schultz and Lutz-Garihan, 1982), and we are seeing only the surviving members of the swarm. Rosenblatt and Charnoz (2012) had demonstrated the feasibility of accreting Phobos and Deimos near the corotation resonance. Any scheme to capture these satellites intact and bring them into their current orbits cannot survive close inspection of the assumptions involved.

The hypothesis that Phobos and Deimos are remnants of material in a dissipative disk, whatever the disk's origin, is not compromised by the inferred orbital history of the satellites. The initial equatorial orbits remain equatorial in spite of the chaotic, large amplitude variations in the obliquity of Mars (Laskar and Robutel, 1993; Touma and Wisdom, 1993; Ward, 1979) and despite the precession of the spin axis of Mars (Goldreich, 1965). The solid angle described by the orbit normal as the satellite orbit precesses due to Mars' oblateness is an adiabatic invariant (Goldreich, 1965), as the precession rates for the Martian satellites (periods of 2.3 and 57 years for Phobos and Deimos, respectively; Peale, 1977) are fast compared with rates of change of Mars' spin axis direction relative to inertial space (timescales $O(10^5)$ years; Folkner et al., 1997; Touma and Wisdom, 1993).

If we insist that both satellites started with nearly circular orbits, how then can we explain the current eccentricity of Phobos' orbit $e_P = 0.0151$? Both the tidal dissipation in Phobos and that in Mars due to tides raised by Phobos damp the eccentricity (see eqn [B.28]). If the orbital motion is integrated backward in time, this eccentricity grows to large values and collisions with Deimos would have been likely (Szeto, 1983; Yoder, 1982), even if there were no tidal dissipation in Phobos. Significant dissipation in Phobos reduces the timescale for a crossing orbit with Deimos to $< 10^9$ years in the past (Yoder, 1982). The current eccentricity cannot therefore be a remnant eccentricity from tidal decay beginning 4.6×10^9 years ago. Yoder (1982) had identified three commensurabilities (defined when two characteristic periods in the description of the motion are in the ratio of small whole numbers) that Phobos has passed through within the past 10^9 years that provide likely gravitational excitations of Phobos' eccentricity during its inward spiral. The commensurabilities are encountered at semimajor axis $a = 3.8$, 3.2, and $2.9R_M$, where the earliest resonance was encountered only 5×10^8 years ago. The first and third are 2:1 and 3:1 commensurabilities between the rotation of Mars and the orbital mean motion, where the resonant interaction is with Mars' axial asymmetry. At $3.2R_M$, the 2:1 commensurability is between the apparent mean motion of the Sun and the periapse motion of Phobos' orbit, where the latter's secular motion is due to the oblate figure of Mars. This resonance is like the evection resonance for the Moon. There is also a 3:2 spin-orbit resonance excitation of the eccentricity when $a = 4.6R_M$, but this excitation happened so long ago that there would be no contribution to the current eccentricity. The eccentricity would have decayed after each excitation and plausibly arrives at the current eccentricity after the series of kicks and subsequent decays (Yoder, 1982). Orbital inclination can also be excited, and even though the resonance interaction is not as strong as it is for the eccentricities, the excited inclinations do not decay. Still, the current inclinations of the orbits are consistent with the resonance passages (Yoder, 1982).

There is a condition on the dissipation in Phobos for this scenario to work. Yoder (1982) had calculated the dissipation in the satellite accounting for the tidal dissipation both due to eccentric orbit as discussed earlier and due to the forced libration of the very asymmetrical satellite. This libration was thought to have an amplitude of 3.9° (Duxbury and Callahan, 1981; Yoder, 1982) at the time and would cause twice the tidal dissipation in Phobos that would occur if Phobos were nearly axially symmetrical in the same eccentric orbit (Yoder, 1982). Recall that dissipation in Phobos and that in Mars due to tides raised by Phobos both damp the eccentricity (eqn [B.28]). There cannot be too much damping since the series of eccentricity excitations or the current eccentricity would be less than that observed. Since the dissipation in Mars can be presumed known from the measurement of Q_M, and the magnitude of the probable excitations can be reasonably estimated from the resonance passage analysis, the current value of e_P limits the contribution by Phobos. Yoder finds that $\mu_P Q_P > 3$–6×10^{12} dynes cm^{-2} or, if $Q_P \sim 100$, $\mu_P \gtrsim 3$ to 6×10^{10} dynes cm^{-2}, which is about that of ice. However, subsequent determinations of the amplitude of physical libration for Phobos are near 1° (Borderies and Yoder, 1990; Duxbury and Callahan, 1989; Jacobson, 2010; Willner et al., 2010). This reduces the dissipation by about a factor of two from what Yoder assumed, thereby relaxing the constraint on the rigidity by about a factor of 2. The properties of Phobos are not sufficiently well known to be sure that the rigidity could satisfy this new constraint, but $\mu \gtrsim 6 \times 10^{10}$ to 1.2×10^{11} is not unreasonable.

During Phobos' spiral toward Mars, it is likely that it passed through the 2:1 orbital mean motion commensurability with

Author's personal copy

Deimos. Such a passage would excite an eccentricity of about 0.002 in Deimos orbit if the eccentricity of Deimos were much smaller than this before resonance passage (Yoder, 1982). The time of this commensurability is known if the current dissipative properties of Mars have not changed substantially since the resonance encounter. This places a lower bound on the dissipation in Deimos if the current eccentricity (0.0002) is the tidal remnant from an initial value of 0.002 excited by the resonance passage. Yoder (1982) found $\mu_D Q_D(1-\alpha_D)^2/\alpha_D^2 \lesssim 10^{10}$ dynes cm^{-2}, where $\alpha_D = 3(B-A)/C$ with $A<B<C$ being the principal moments of inertia of Deimos. This limit may be unreasonably low, but the dissipation in Deimos may be increased if the forced libration is nearly resonant with the free libration. The enhanced amplitude of libration would lead to higher dissipation and relax the constraint on $\mu_D Q_D$. The free libration period could be better constrained by an estimate of α_D from a more accurate determination of Deimos' shape along with an accurate measure of its physical libration amplitude. With semi-axes of 7.5, 6.2, and 5.4 km for Deimos (Duxbury and Callahan, 1989) and a mean density of 1.5 g cm^{-3} (Rosenblatt, 2011), $\alpha_D = 0.354$ and $\mu_D Q_D \lesssim 3.3 \times 10^{10}$ dynes cm^{-2}. A homogeneous Deimos modeled as a triaxial ellipsoid would have a free libration period about 1.3 times the orbital period, which is not sufficiently close to resonance to enhance the forced libration significantly, but the fact that Deimos is a porous rubble pile may enhance the dissipation above that for a less rigid solid body.

In any case, Yoder's hypothesis that the satellite orbits have always been regular and current properties of the system then attributed to the effects of resonance passages by Phobos is reasonably well supported. This is consistent with our presumed origin from a dissipative disk of small particles, although the details of the origin of the disk have yet to be constrained.

10.17.4 Jupiter System

The Galilean satellites of the Jupiter system and those small satellites inside Io's orbit are nearly coplanar with Jupiter's equator implying their formation in a dissipative disk (**Appendix A**). Additional constraints on the formation of the Galilean satellites of Jupiter have been established by theoretical analysis and the Galileo spacecraft observations. The mean densities of Io, Europa, Ganymede, and Callisto are, respectively, 3.53, 3.01, 1.94, and 1.83 g cm^{-3}. Ganymede and Callisto are about 40% ice and 60% rock by mass (Sohl et al., 2002). Europa has a layer of water and ice 80–170 km thick (Anderson et al., 1998b). CO_2 has been detected on the surfaces of Ganymede and Callisto (Hibbitts et al., 2000, 2002, 2003; McCord et al., 1998) and in the atmosphere of Callisto (Carlson, 1999). The discovery of an intrinsic dipole magnetic field on Ganymede and the low amplitude of higher-order multipoles in the field implies generation of the field deep within the satellite probably by dynamo action in a molten metallic core (Kivelson et al., 1997, 1999, 2002). Induced fields in Europa and Callisto and possibly Ganymede from the variation in the external field due to Jupiter's rotation are consistent with induced fields expected from a conducting liquid water layer at rather shallow depths (Kivelson et al., 2002). It has been speculated that NH_3 is sufficiently abundant on Ganymede and Callisto to lower the solidus temperature of an NH_3–H_2O mixture so that such a liquid layer could be maintained over the age of the solar system (Mousis et al., 2002; Spohn and Schubert, 2003). However, there has been no NH_3 nor N_2 detected on any of the Galilean satellites, but McKinnon (2006) argued that halides in solution may depress the solidus temperature sufficiently to preserve the liquid layers.

The large amount of relatively volatile material in Ganymede and Callisto implies that the disk temperature remained low enough during the accretion process that icy particles could persist. Callisto may not be fully differentiated (Anderson et al., 1998a, 2001; Nagel et al., 2004), and an accretion timescale $\geq 6 \times 10^5$ years is required to keep ice in its interior from melting (Barr and Canup, 2008). The change in composition from rocky to icy from Io to Callisto is consistent with an expected radial temperature gradient in the disk at the time of satellite accretion, but note the existence of the icy satellite Amalthea inside Io's orbit. It is possible that Europa lost initial volatiles due to tidal heating (**Appendix B**). Given its current tidal dissipation rate, a very efficient conversion of tidal heating into vaporization of ice is required over the satellite's entire lifetime in order for Europa to have started as ice-rich as Ganymede and Callisto (e.g., Canup and Ward, 2009). If Europa's ice depletion is primordial, then Io's likely is as well, since by virtue of being interior to Europa, it may have formed in a higher-temperature environment.

The theories for the formation of the Galilean satellites are guided by numerical hydrodynamic calculations – first in two dimensions (2-D) (e.g., Lubow et al., 1999) and later in 3-D (e.g., Bate et al., 2003; d'Angelo et al., 2003; Machida et al., 2008; Tanigawa et al., 2012), where the flows near the Hill sphere are strikingly different in the two types of calculations. It is clear that 3-D calculations are required to constrain the conditions of satellite formation in planetary accretion disks, but the incompleteness of these calculations (e.g., there are no numerical calculations that follow the development of the disk from the collapse phase of Jupiter's atmosphere to the satellite formation stage) has led to a variety of models involving speculation of the disk properties and the dominant processes therein.

The formation of the regular Jovian satellites must necessarily occur late in Jupiter's accretion to allow the hot inflated protoplanet to cool and contract to a radius less than that of the closer satellites. This can plausibly occur by the time the planet is accreting the last few tens of percent of its mass, depending primarily on the planet's opacity (Ward and Canup, 2010). Minimum mass disk models (called minimum mass sub nebula (MMSN) models) for the formation of the satellites require sufficient mass of condensable material to form the satellites augmented with gas to reach solar composition in the initial disk (e.g., Lunine and Stevenson, 1982), where no mass is added to the disk during the satellite accretion process. Mosqueira and Estrada (2003a,b) modified the MMSN model by enhancing the fraction of solid material in the disk (solids-enhanced minimum mass), so there is less gas than in the traditional MMSN models. Various theoretical considerations of disk processes and histories (**Appendix B**) will lead us to favor models (with some caveats) where the current satellites form during the very last stages of Jupiter's

Author's personal copy

accretion when the accretion disk was 'starved' of gas input (Alibert et al., 2005a,b; Canup and Ward, 2001, 2002, 2006, 2009; Mousis and Alibert, 2006; Ogihara and Ida, 2012; Sasaki et al., 2010; Stevenson, 2001a,b; Ward and Canup, 2010). In these latter models, considerably more condensable mass passes through the disk than is contained in the current satellites, and these objects form in a disk with several orders of magnitude less gas than in an MMSN.

The shocks near the L_1 and L_2 Lagrange points, after a gap is formed in the solar nebula, which are found in the 2-D calculations of Bryden et al. (1999), Kley (1999), and Lubow et al. (1999), are modified considerably in the higher-resolution 3-D calculations of Machida et al. (2008). For a 1 Jupiter mass protoplanet, the shocks are displaced about 45° from the Sun–Jupiter line and stand off from the Hill sphere by about 20% of R_H. A reservation about this comparison is that the calculations correspond to a full density solar nebula and are local. The local nature of the Machida et al. calculations means the creation of a deep gap in the disk as Jupiter's mass grows cannot be treated (Machida et al., 2010). In the final stages of accretion, the circumplanetary disk is more likely to behave like a standard accretion disk rather than being subject to the shock-driven accretion found in the 2-D calculations (Bate et al., 2003). In both 2-D and 3-D calculations, a prograde accretion disk is formed, which is the basis for the formation of a coplanar set of satellites in nearly circular orbits.

In spite of Jupiter having likely opened a gap (**Appendix A**) in the solar nebula when its mass exceeded approximately 1/3 of its current value (e.g., Bryden et al., 1999; Ward and Hahn, 2000), 3-D numerical simulations (e.g., Kley et al., 2001) find continuing accretion onto Jupiter at a rate of 10^{-4}–$10^{-2}\,M_E$ year^{-1}, depending on the assumed nebular viscosity, surface mass density, and scale height. Bate et al. (2003), in a 3-D hydrodynamic simulation, found an accretion rate close to $10^{-4}\,M_J$ year^{-1} ($3\times10^{-2}\,M_E$ year^{-1}) as Jupiter approaches its final mass, M_J. The latter authors find that accretion only stops when the planet at 5.2 AU reaches about $10M_J$, with an undepleted solar nebula. The accretion of Jupiter was therefore most likely protracted, with the rate of accretion tailing off as the solar nebula is depleted. A striking conclusion from the 3-D calculations (Bate et al., 2003; Machida et al., 2008; Tanigawa et al., 2012) is that relatively little material from the midplane of the solar nebula is accreted onto the disk. Most of the material rains down on the disk from above and below the midplane. This phenomenon is incorporated into starved disk models discussed in the succeeding text, but first, we describe a recently developed solids-enhanced minimum mass (SEMM) model (Estrada et al., 2009; Mosqueira and Estrada, 2003a,b) and a gas-free accumulation of the satellites from planetesimals captured from the heliocentric swarm (Mosqueira and Estrada, 2006).

In the following discussions of gaseous disk models, we will use the turbulent viscosity parameter α, defined by $\nu_t=\alpha c_s$, where ν_t is a characteristic velocity of a turbulent eddy and c_s is the speed of sound in the gas (**Section A.2**). The value of α is very uncertain, but typical values are chosen near 10^{-3}. Sources of turbulence in the circumplanetary disks include the influx of material onto the disk (e.g., Cassen and Moosman, 1981) and/or turbulence generated by the Magnetohydrodynamic (MHD) instability (Turner et al., 2014). The latter source requires a certain degree of ionization of the disk material that may be difficult to attain in the environments of most planetary accretion disks. See Turner et al. (2014) for a thorough discussion of the conditions where MHD turbulence may be effective in generating a viscous evolution of a disk and Lubow and Martin (2013) who consider the possibility of a layered disk in which the disk's midplane is nonturbulent, while the outer layers are MHD-active.

10.17.4.1 SEMM Model

We first outline the motivation for the initial conditions in the Mosqueira–Estrada (2003a,b) model. The angular momentum of a planetesimal (or a small volume of gas) relative to the Jupiter embryo is $\mathbf{L}=m\mathbf{r}\times\mathbf{v}$, where m, $\mathbf{r}$, and $\mathbf{v}$ are mass, position, and velocity relative to a frame centered on Jupiter but not rotating. If we write $\mathbf{v}=\mathbf{v}_{rel}+\mathbf{n}_J\times\mathbf{r}$, where $\mathbf{v}_{rel}$ is the velocity of the particle relative to a frame centered on Jupiter and rotating with the mean orbital angular velocity $\mathbf{n}_J=\sqrt{GM_\odot/a_J^3}\,\boldsymbol{e}_z$ with G, $M_\odot$, a_J, $\boldsymbol{e}_z$ being the gravitational constant, solar mass, semimajor axis of Jupiter's orbit, and normal to the orbital plane, respectively, $\mathbf{L}=m\mathbf{r}\times(\mathbf{v}_{rel}+\mathbf{n}_J\times\mathbf{r})=m\mathbf{r}\times\mathbf{v}_{rel}+mr^2\mathbf{n}_J$, where the second form is for the assumption that $\mathbf{r}$ is in the plane of the disk and thereby perpendicular to $\mathbf{n}_J$ along the orbit normal. The specific angular momentum is

$$\ell=\mathbf{r}\times\mathbf{v}_{rel}+r^2\mathbf{n}_J \qquad [8]$$

where the first term is the specific angular momentum relative to the rotating system and the second is the correction for inertial space representation.

Before the creation of a gap in the heliocentric nebula, ($M\ll M_J$), it is assumed that all of the flux encountering the Hill sphere of radius $R_H=(M/3M_\odot)^{1/3}a$ is captured by the proto-Jupiter. Lissauer and Kary (1991) found for the sum of all the specific angular momenta

$$\ell=n_J\frac{\int_0^{R_H}-1.5x^3\,dx}{\int_0^{R_H}x\,dx}+n_JR_H^2=\frac{1}{4}n_JR_H^2 \qquad [9]$$

where x is the difference in the semimajor axes of the particle and Jupiter, where we have dropped the vector notation, since it is assumed that all contributions to the specific angular momentum of accreted material are perpendicular to the disk. With conservation of angular momentum, the centrifugal radius is defined by $r_c=\ell^2/GM=R_H/48\approx15R_J$ for the value of ℓ given in eqn [9] and for R_H determined for $M=M_J$, where R_J is Jupiter's current radius. The centrifugal radius r_c is interpreted as the disk size. For $M\ll M_J$ appropriate for eqn [9], the disk before gap opening is quite compact.

After the gap in the heliocentric disk becomes large compared to R_H, material enters the Hill sphere near the Lagrange points in the 2-D calculations of Artymowicz and Lubow (1996) and Lubow et al. (1999). In the 3-D calculations of Machida et al. (2008), material enters the Hill sphere, after passing through the shocks, in the first and third quadrants projected onto the $z=0$ plane with the x-axis pointing away from the Sun. Nearly all of the 3-D calculations show material falling vertically onto the disk (d'Angelo et al., 2003; Machida,

Author's personal copy

2009; Machida et al., 2008, 2010; Tanigawa et al., 2012). After gap opening, Mosqueira and Estrada (Estrada et al., 2009) assumed that gas enters the Hill sphere at very small relative velocity. The first term in eqn [9] thereby vanishes and the specific angular momentum becomes $R_H^2 n_J$ and $r_c = R_H/3 \sim 250 R_J$.

Based on this, Mosqueira and Estrada (2003a,b) proposed a two-component disk: a massive disk for ($r \lesssim 15 R_J$) that persists before gap opening surrounded by an extensive outer disk ($r \sim 150 R_J$) of much lower surface mass density, emplaced after a large gap has formed. The sparseness of the outer disk follows from the much reduced gas flow into the Hill sphere after the gap approaches its maximum extent, and the large radius of the disk follows from the higher specific angular momentum of the material entering the Hill sphere at the Lagrange points as discussed earlier. The gas influx rapidly goes to zero due to gap opening, but solid material continues to be added to the disk through ablation of fragmented planetesimals passing through the disk (e.g., Mosqueira et al., 2010) or collisions of such fragments with existing solid satellitesimals already in the disk (e.g., Estrada and Mosqueira, 2006, see succeeding text).

The gas flow is assumed to wane on the same timescale as the gap opening timescale of $\sim 10^3$ years for a weakly turbulent heliocentric disk (Bryden et al., 1999, 2000; Morbidelli and Crida, 2007). The cessation of the gas flow onto the disk means the presumed chief source of turbulence in the disk is eliminated and the disk viscosity and temperature plummet. Note that the gas flow ends when $M \approx M_J$, which is contrary to the conclusion of Bate et al. (2003), where accretion stops only when $M \approx 10 M_J$, although the latter is for an undepleted solar nebula.

The postaccretion low disk viscosity and low temperature are involved in various schemes for avoiding the short timescales for the important processes in the disk discussed in **Appendix A**. The system is closed, since continuing accretion is omitted, where the minimum mass tag implies there is only sufficient solid material to make the satellites, but the gas content of the disk is down from a solar composition by a factor of 10 (Estrada et al., 2009). A $1/r$ temperature distribution is assumed with $T = 250$ K at Ganymede's distance in order to accommodate the stability of icy particles at greater distances. A feature of the model is a surface density distribution ($\sigma_g \propto 1/r$) with sufficient solids to make Io, Europa, and Ganymede inside $\sim 15 R_J$, leading to an optically thick, massive inner disk. Mosqueira and Estrada set the value of the gas surface density to $\sigma_g \sim 3 \times 10^4$ to 3×10^5 g cm^{-2}, with the lower end of the range indicating a disk enhanced in solids compared to the standard MMSN disk. But the surface density falls drastically beyond about $15 R_J$ to a value determined by spreading Callisto's mass in the outer disk out to $r \sim 150 R_J$. With an opacity due only to the gas (i.e., without dust grains), an optically thin outer disk results from which it is argued that Callisto is accreted slowly ($\sigma_g \sim 1/r$ for $r > 26 R_J$). The $1/r$ temperature dependence in the inner disk goes to a constant temperature in the outer disk of 130 K, like that of the solar nebula. A controversial assumption in the model is that the sharp gradient in the surface density between the inner and outer disks is maintained throughout the accretion timescale of Callisto, which they estimate to be $O(10^5$–10^6 years).

The steep surface density gradient between the inner and outer disks is maintained by assuming that turbulence has been nearly completely damped leading to $\alpha \sim 10^{-5}$–10^{-6} in both the inner and outer disks. Such low values of α might not be definitely ruled out, since there is no continuing accretion to stir the disk. Small but decoupled objects are rapidly lost to Jupiter through gas drag. But the accumulation of dust and rubble into larger objects is more rapid than the decay of these objects by gas drag. The small particles are assumed entrained in the gas and encounter larger particles that are at Kepler velocities. The growth timescale is $M/\dot{M} = 4\rho_p R(3\rho_s \Delta v)$, where Δv is the difference between the Kepler and gas velocities. Sweepup time is smaller than gas drag time for $r < 38 R_J$ independent of particle size or distance range where $T \propto 1/r$, such that satellite formation is favored. Then, the timescale for accretion is determined by the rate at which gas drag can bring say 100 m or larger objects to the protosatellite. Accretion times for Ganymede under these assumptions are $O(10^3$–$10^4)$ years. Since they have little or no ice, Io and Europa can accrete more rapidly than Ganymede.

The short migration timescales indicated in **Appendix A** are frustrated by the assumption that the inner three Galilean satellites open gaps in the disk and change from type I to type II migration. The latter is on the viscous timescale, which exceeds the supposed timescale for elimination of the gas disk by photoevaporation (Hollenbach et al., 2000; Shu et al., 1993), because the turbulent viscosity characterized by $\alpha = O(10^{-6})$ is so small. The steep gradient in the surface mass density between about 20 and $26 R_J$ causes type I drift to reverse sign in this region so that a satellite would spiral out instead of in. Callisto is hypothesized to have ended up where the type I drift vanishes at the value of the density gradient where torques from the inner Lindblad resonances balance those from the outer ones. The steep gradient in σ_g between the inner and outer disks must be maintained not only through the accretion phase but also throughout the dissipation of the disk. Alternatively, Callisto could open a gap, shift from type I to type II drift and thereby end up where it is because of the low viscosity assumed. But why did it wait until $r = 26 R_J$ to open the gap?

The SEMM model remains at this point a conceptual framework for satellite formation; there have been no explicit, time-dependent models demonstrating either the formation of the assumed SEMM initial conditions or that the proposed SEMM disk would ultimately accrete into a system of four Galilean-like satellites. Many of the subjective criticisms of the Mosqueira–Estrada model, such as the very low values of α assumed from the outset to preserve the steep surface density gradient between the inner and outer disks, are muted by uncertainties about which processes are ongoing and on the values of various parameters therein. However, many aspects of the model are ad hoc, such as the $1/r$ temperature distribution normalized to allow ice to condense just beyond Ganymede's orbit, the low viscosity (and maintenance of the steep surface density gradient between the inner and outer disks), and the assumption of the existence of a minimum mass, closed disk in the first place. The latter assumption ignores the history of the disk during the long period of waning gas accretion. How could the massive inner disk resist viscous

Author's personal copy

spreading during this period when accretion still generated a reasonable turbulent viscosity?

The SEMM model, as well as the model of Sasaki et al. (2010) discussed in the succeeding text, is not consistent with recent results from increasingly sophisticated numerical simulations of planetary accretion (e.g., Bate et al., 2003; Machida et al., 2008), which show the protracted nature and persistence of the accretion process as the solar nebula is dissipated. These latter results confirm intuitive conclusions about the accretion histories of the giant planets. The neglect of this history and the added criticisms in the preceding text mean that the SEMM and MMSN models in general are not likely to represent the actual sequence of events that led to the formation of the regular natural satellites around the giant planets. In spite of these criticisms, we note that the Mosqueira–Estrada papers contain discussions of many, if not most, of the processes that occur in disks that effect the satellite development. These processes should be studied in the development of a comprehensive model of origin of the regular satellites.

10.17.4.2 Planetesimal Capture Model

In a rather detailed and careful analysis, Estrada and Mosquiera (2006) reexamined the hypothesis that collisions of planetesimals within Jupiter's Hill sphere in a gas-free environment led to a buildup of captured orbiting material that settles to an equatorial disk through mutual collisions where they could have accumulated into the satellites (e.g., Safronov et al., 1986). Constraints are placed on numerous parameters and assumptions that could allow such a scenario to produce the observed masses and angular momentum distributions, but the authors note that whether or not the parameter values and assumptions are likely to obtain is very uncertain. Perhaps the greatest value of this paper is to show how difficult it is to make a gas-free planetesimal capture scenario a viable alternative for satellite formation. Major caveats are the difficulty in producing the composition gradient in such a scenario (unless later Roche lobe interlopers or tidal dissipation preferentially removes volatiles from the inner satellites), keeping a size distribution of planetesimals that favors collisional capture, and maintaining the supply of planetesimals within Jupiter's feeding zone over the 10^5–10^6-year timescale of the satellite accumulation. It is not clear that a sufficient gas density in the solar nebula to allow a gas drag replenishment of planetesimals in Jupiter's feeding zone would not also keep a gas accretion disk replenished by streaming across the gap. A problem with planetesimals later colliding with the accreted satellites to preferentially remove volatiles from the inner satellites is attaining a sufficient flux of such planetesimals from the depleted zone around Jupiter.

Another problem that is not addressed is the reduction of the optical depth in the circumplanetary disk as the satellitesimals accumulate into larger objects on short timescales, thereby reducing the collisional capture probability of additional mass. In addition, the scenario is not placed in the context of the formation of Jupiter that yields a plausible route to the assumed gas-free initial conditions where Jupiter has no satellites. Presumably, Jupiter has already gone through the last stages of gas accretion through streams of material across a gap in the solar nebula, and we shall see later that it should have acquired a stable satellite system already during that process. With the caveats pointed out by Estrada and Mosqueira and the additional ones we have mentioned here, it seems unlikely that the satellites formed solely from collisionally captured planetesimals in a nearly gas-free environment.

10.17.4.3 Starved Accretion Disk Model

Accumulation of the satellites of Jupiter in an actively supplied, gas-starved accretion disk instead of a MMSN was proposed by Canup and Ward (2001) and separately by Stevenson (2001a) at the Jupiter conference in Boulder, Colorado, in June 2001 (http://lasp.colorado.edu/jupiter/index.html). In this construct, the planet contracts to a size smaller than the satellite region before the nebula dissipates, with subsequent inflow producing a circumplanetary accretion disk (e.g., Ward and Canup, 2010). The satellites form in the disk as the inflow is ongoing, rather than after the inflow has stopped (as in the MMSN and SEMM models). As the planet nears its final mass and the inflow wanes, the disk becomes increasingly gas-starved. Motivations for the starved disk model include a more consistent treatment of the conditions of satellite formation based on planet formation models, the necessity of disk temperatures low enough to allow water ice to condense, and a sufficiently slow accretion of Callisto to allow possibly incomplete differentiation (e.g., Anderson et al., 2001; Kuskov and Kronrod, 2005; McKinnon, 1997; Schubert et al., 2004). Stevenson (2001b) summarized remarks at the conference, whereas Canup and Ward (2002) developed a detailed steady-state model based on their conference paper. Their model was extended by Alibert et al. (2005a,b) who accounted for the time variation of disk properties as accretion wanes. Major features in the Alibert et al. approach are that material from the solar nebula is deposited at the outer edge of the disk with a mass flux through the disk that is in quasi-steady state and that the last part of the satellite formation occurs in a closed disk after accretion has ended. Canup and Ward (2002) had the mass raining down onto the disk consistent with a major finding of the 3-D calculations of Bate et al. (2003), Machida et al. (2008), and Tanigawa et al. (2012) showing the vertical deposition. Canup and Ward (2006) extended their model to include an exponential decrease in the accretion rate with time and a direct model of satellite growth, in which many of the accreted satellites are lost to Jupiter from migration and only the last set formed near the end of accretion remain. Ward and Canup (2010) considered the formation of a disk around a growing gas giant planet, including the energy and angular momentum budget of the inflowing material and the coupled thermodynamic evolution of the planet itself. They found that by the time the planet accretes the final ~10% of its mass, an accretion disk with low gas surface densities and properties broadly similar to those considered in Canup and Ward (2002, 2006) has formed. Sasaki et al. (2010) and Ogihara and Ida (2012) considered two variations on the Canup–Ward model: an inflow that ends abruptly rather than exponentially due to gap opening (and in this these two works share a common element with the SEMM model) and a disk with a magnetically sustained inner cavity. It is the Canup–Ward (2002, 2006, 2009) model that we now describe.

Author's personal copy

Ward and Canup (2010) argued that the vertical deposition of material onto the disk and planet falls in the range $r_c/4 < r < 9r_c/4$, where the upper bound is close to the $30R_J$ used in their 2002 model. For accretion during gas inflow, the steady-state disk with uniform deposition inside a radius of $30R_J$ treated in 2002 is generalized in 2006 to have a time-dependent influx density of

$$F_{in} \propto \left(\frac{1}{r}\right)^{\gamma_{in}} \exp\left(-\frac{t}{\tau_{in}}\right) \quad [10]$$

where F_{in} is the vertical flux density onto the disk's upper and lower surfaces, r is the radial distance from the center of Jupiter, τ_{in} is the time constant for exponential decay of the influx rate, and γ_{in} is an adjustable constant parameterizing the radial variation in the flux density over the disk inside the outer bounding radius. The radius r_c is a free parameter constrained by the specific angular momentum of inflowing material. A range of 16–46 planetary radii for Jupiter, Saturn, and Uranus is consistent with Lissauer's (1995) 3-D calculation (with 16 being the value for Jupiter in particular). The disk outer boundary is set to $r_d \approx 150R_J$. If the viscous timescale $\tau_\nu \approx r^2/\nu$, with $\nu = \alpha c_s H$ (c_s = midplane sound speed and H = gas scale height, **Appendix A**), is much smaller than τ_{in}, the gas surface density σ_g maintains a quasi-steady state like that assumed in Canup and Ward (2002). For $\tau_{in} \sim 10^6$ years, this condition is valid for $\alpha \geq 10^{-4}$.

Solids smaller than ~1 m are entrained in the gas and, for a planet that has not opened a gap, will be delivered with it to the disk from heliocentric orbit. The delivery of small particles to the disk is more complex once the planet opens a gap. At the outer gap edge, the gas disk's rotation is super-Keplerian because the local gas surface density increases with orbital radius, thereby reversing the radial pressure gradient. The gas drag that caused the inward migration of small particles is reversed, causing them to stall and build up at the outer gap edge (Paardekooper and Mellema, 2006a,b). However, this buildup cannot proceed indefinitely, because once the surface density of small particles becomes comparable to that of the local gas, the back reaction of the particles on the gas causes its motion to become more Keplerian and particles begin to flow into the gap (Ward, 2009). A simple model accounting for this effect suggests that particles mm-sized and smaller could enter the gap and be delivered to the disk (Ward, 2009), although this issue merits further investigation for larger particles.

It has traditionally been thought that most solids in the heliocentric disk would be in large planetesimals by the time of Jupiter's formation, although these would probably be diminished in size through fragmentary collisions after Jupiter had gained most of its mass. However, recent works argue that a large fraction of the disk's solid mass was contained in millimeter- to centimeter-sized 'pebbles,' based both on observations of extrasolar disks (e.g., Testi et al., 2003; Wilner et al., 2005) and on dynamic models (e.g., Lambrechts and Johansen, 2012). The latter found that pebble accretion can lead to formation of the gas giant cores in a time shorter than the nebular lifetime. Such conditions would also imply a substantial population of pebble- and grain-sized particles that could be delivered to the circumplanetary disk by the gas inflow.

It may also be the case that an alternative delivery of solids to the disk may be viable. Once in the disk, the solids are assumed to accrete rapidly enough to decouple from the gas before they can spiral into Jupiter from gas drag. In an inflow-supplied disk, the accretion timescale for a satellite of mass m_s is $\tau_{acc} \approx f m_s/(2\pi r \Delta r F_{in})$, where $f \approx 100$ is the gas to solid ratio for the influx and Δr is the width of the annulus over which the satellite accumulates material. With $\Delta r/r \approx 2e$, where $e \approx (H/r)(m_s/4\pi r H \sigma_g)^{1/3}$ is the maximum eccentricity obtained by m_s from a balance between eccentricity damping by density waves and excitation by scattering from similarly sized objects (Ward, 1993), the accretion timescale for a uniform inflow varies with orbital radius as $\tau_{acc} \propto 1/r^{4/5}$ (Canup and Ward, 2006). The inverse dependence on r is because accretion is being regulated by the rate of the inflow, rather than by the orbital frequency as is the case in disks with a fixed total mass. For a uniform inflow per area, mass is added more quickly to an outer satellite's feeding zone due to its much larger surface area, so that objects in the outer disk grow more quickly than those in the inner disk.

Satellite growth continues, while each satellite spirals toward Jupiter from type I drift (**Appendix A**). The rate of type I drift is proportional to the satellite mass and to the gas surface density in the disk (**Appendix A**), and the satellites tend to accelerate as they drift toward Jupiter due to their accretion of some additional inner disk material. The timescales for type I drift (**Appendix A**) are generally shorter than τ_{in} in eqn [10], so that many satellites are lost to Jupiter as the influx winds down. But as the inflow wanes, the gas surface density drops and the type I timescale lengthens; the last surviving satellites are then those for which $\tau_1 \sim \tau_{in}$. Each satellite reaches a critical mass m_{crit} where the timescale τ_{acc} for further growth is comparable to its type I orbital decay timescale after which further growth is limited before the satellite is lost to Jupiter.

The model includes disk heating and cooling, where the heating comes from the gravitational energy lost by the incoming material, viscous dissipation within the disk, and even radiation from a still warm Jupiter with only radiative cooling to keep the disk temperature below the ice point at least in the outer regions. Disk temperature distributions and surface densities are constrained by assumptions about the disk opacities and the viscosity parameter α defined in **Appendix A**. The timescale for type I drift depends on the scale height, which depends on the temperature of the disk, and that temperature distribution depends on the opacity of the solids–gas mixture in the disk, which itself is a function of temperature (Pollack et al., 1994; **Appendix A**). Higher opacities lead to slower required inflow rates for similar temperatures. Opacity is then important in determining both the disk structure and the manner in which the migration rate changes with r. Unfortunately, the opacity and its dependence on radial distance from Jupiter are highly uncertain, since it depends on the amount, composition, and size distribution of the dust in the nebula. Canup and Ward (2002, 2006) used a constant opacity in their model, and it would be of interest to determine the effect of other assumptions about the opacity distribution and the general thermal properties of the disk on their conclusions.

For the case where $\gamma_{in} = 0$ in eqn [10] (uniform deposition inside r_c) Canup and Ward (2006) find in their eqn [2],

Author's personal copy

$$\frac{m_{\rm crit}}{M_{\rm J}} \approx 5.4\left(\frac{\pi}{C_{\rm a}}\right)^{5/9}\left(\frac{H}{r}\right)^{26/9}\left(\frac{r}{r_{\rm c}}\right)^{10/9}\left(\frac{\alpha}{f}\right)^{2/3}(\Omega_{\rm K}\tau_{\rm G}f)^{1/9}$$
$$\approx 5.6\times10^{-5}\chi\left(\frac{3.5}{C_{\rm a}}\right)^{5/9}\left(\frac{H/r}{0.1}\right)^{26/9}\left(\frac{r/r_{\rm c}}{0.5}\right)^{10/9}\left(\frac{\alpha/f}{3\times10^{-5}}\right)^{2/3} \quad [11]$$

where $\chi=[(1\text{ week}/\{2\pi/\Omega_{\rm K}\})(\tau_{\rm G}/10^7\text{ years})]^{1/9}$ and $\tau_{\rm G}=M_{\rm J}/(dM_{\rm J}/dt)$ is the timescale for the growth of $M_{\rm J}$. Because the ratio $m_{\rm crit}/M_{\rm J}$ depends so weakly on the inflow rate ($\tau_{\rm G}^{1/9}$), a similar maximum satellite mass results for a wide range of inflow rates. Next, the dependence of $m_{\rm crit}/M_{\rm J}$ on α/f for a given inflow rate can be understood from the following. A larger α leads to a lower surface mass density and hence slower migration, which allows the satellite mass to grow more before it is lost, and a smaller f means more solids are delivered also resulting in an increased satellite mass.

Canup and Ward (2002, 2006) used a type I timescale appropriate for optically thin, isothermal disks (Tanaka et al., 2002), with $C_{\rm a}=2.7+1.1\zeta$, where ζ is the exponent for the radial surface mass distribution as in $\sigma_{\rm g}\propto 1/r^{\zeta}$. $C_{\rm a}=3.5$ for a disk with $\zeta=3/4$ (Canup and Ward, 2002). It is possible that even a gas-starved disk can be optically thick, depending on the abundance and size of dust in the disk. Recent works have shown that type I migration is altered in optically thick, adiabatic disks due to opposing effects of corotation torques (e.g., Kley and Crida, 2008; Paardekooper and Mellema, 2006a,b). Paardekooper et al. (2010) provided a revised $C_{\rm a}$ accounting for such effects, where

$$C_{\rm a}=-\frac{2}{\gamma_{\rm a}}\left[-2.5-1.7\beta+0.1\zeta+1.1\left(\frac{3}{2}-\zeta\right)+7.9\frac{\beta-(\gamma_{\rm a}-1)\zeta}{\gamma_{\rm a}}\right]$$

where $\gamma_{\rm a}$ is the adiabatic index (1.4 for diatomic hydrogen), β is the exponent in the disk temperature profile $T_{\rm d}\propto 1/r^{\beta}$, and $C_{\rm a}>0$ (<0) is inward (outward) migration. For values appropriate for an optically thick inflow-supplied disk, $C_{\rm a}=0.5$, indicating inward type I migration that is about seven times slower than that estimated by Tanaka et al. (2002) for optically thin, isothermal disks. Because the satellite mass is set by a balance between satellite growth and type I loss, a reduction in $C_{\rm a}$ to 0.5 in eqn [11] would produce a comparable critical satellite mass for an α/f ratio that is a factor of 5 smaller, reducing, for example, the needed level of viscosity to achieve a desired satellite mass.

In these analytic approximations, each satellite grows to mass $m_{\rm crit}(r)$ within an annulus of width Δr, after which it is rapidly lost from type I migration. But in the time to migrate into Jupiter, another $m_{\rm crit}$ accumulates. So, the total mass $M_{\rm T}$ in satellites at any time is given by eqn [4] of Canup and Ward (2006),

$$\frac{M_{\rm T}}{M_{\rm J}}=\int_{R_{\rm J}}^{r_{\rm c}}\frac{m_{\rm crit}/M_{\rm J}}{\Delta r}dr$$
$$\approx 2.1\left(\frac{\pi}{C_{\rm a}}\right)^{4/9}\left(\frac{H}{r}\right)^{10/9}\left(\frac{\alpha}{f}\right)^{1/3}\frac{1}{(\Omega_{\rm K}\tau_{\rm G}f)^{1/9}} \quad [12]$$
$$\approx 2.5\times10^{-4}\frac{1}{\chi}\left(\frac{3.5}{C_{\rm a}}\right)^{4/9}\left(\frac{H/r}{0.1}\right)^{10/9}\left(\frac{\alpha/f}{3\times10^{-5}}\right)^{1/3}$$

where it is assumed that H/r and f are approximately constant across the disk and that $R_{\rm J}\ll r_{\rm c}$. An important characteristic of eqn [12] is that $M_{\rm T}/M_{\rm J}$ is comparable to the observed satellite–planet mass ratios for Jupiter with plausible choices for the parameters, and by replacing $M_{\rm J}$ by $M_{\rm S}$ or $M_{\rm U}$, the same applies to Saturn and Uranus. Notably, $M_{\rm T}/M_{\rm J}$ is insensitive to the inflow rate through χ, lacks a dependence on $r_{\rm c}$, and depends only weakly on α/f.

Perhaps the most notable result of this model is the prediction that all the major planets should have values of $M_{\rm T}/M_{\rm P}\approx10^{-4}$, which is what is observed. The value of $M_{\rm T}/M_{\rm P}$ is determined primarily by the value of α/f, and the weak dependence of the mass fraction on this ratio ($(\alpha/f)^{1/3}$) means that systems comparable to those of Jupiter, Saturn, and Uranus result for $10^{-6}<\alpha/f<5\times10^{-4}$. Neptune is excluded because the capture of the large satellite Triton disrupted the original system. The application of the theory to Uranus requires additional processes that lead to the large obliquity after the satellite system was formed. An impact that tilted Uranus would disrupt the Canup and Ward-formed satellite system leading to mutual collisions among the satellites and reassembly in the new equatorial plane, or if an adiabatic process can be found to tilt Uranus, the satellite system would remain equatorial during the process (see **Section 10.17.6**). Both α and f are most likely to vary both in time and from planet to planet given the difference in the planet compositions and the likely decrease in f from solar composition as the solar nebula is evaporated. Whether or not the given influx of material can cause enough turbulence for values of α/f within this range has not been ascertained. Magnetorotational instability is another potential source of turbulence, but its effectiveness is greatly diminished by the likely presence of dust grains and likely insufficient ionization (e.g., Fujii et al., 2011).

To check these conclusions, satellite formation is simulated in Canup and Ward (2006) by the N-body SyMBA code (Duncan et al., 1998) based on the algebraic mapping integration scheme of Wisdom and Holman (1991), modified to include the interaction of the masses with the gas disk and ongoing mass inflow (Canup and Ward, 2006). The satellites interact gravitationally with Jupiter and with each other. The background disk is treated analytically and interactions of the satellites with it are treated as small perturbations of the orbits. The radial surface density profile is approximated by $\sigma_{\rm g}(r)=\sigma_0(R_{\rm J}/r)^{\gamma_{\rm g}}$, and $H/r=c_{\rm s}/r\Omega_{\rm K}=h_0(r/R_{\rm J})^{\gamma_{\rm c}}$ gives the radial variation of the scale height. The scale height depends on the radial temperature distribution determined by the opacity. Values of $\gamma_{\rm g}=0.75$, $\gamma_{\rm c}=0.13$, and $h_0=0.07$ are estimated from the steady-state disk model of Canup and Ward (2002) with a constant opacity (see **Appendix A**). Satellites experience a drag force because the gas is partially supported by the radial pressure gradient discussed in **Appendix A**, where the satellites are given kicks at each time step proportional to the area to mass ratio in a direction opposite to the relative velocity between the satellite and the gas. The gas drag is only important while the satellites are small, where type I interactions dominate the reduction in eccentricity, orbital inclinations, and semimajor axis for larger satellites. After Papaloizou and Larwood (2000), type I interactions are simulated with added accelerations $\vec{a}_{\rm I}=-\vec{v}/\tau_{\rm I}$, and $\vec{a}_{\rm damp}=-2\left(\vec{v}\cdot\vec{r}\right)\vec{r}/(r^2\tau_{\rm e})-2\left(\vec{v}\cdot\hat{k}\right)\hat{k}/(r^2\tau_{\rm i})$, where $\hat{k}$ is a unit vector perpendicular to the plane of the disk and $\vec{v}$ is the satellite velocity. The timescale $\tau_{\rm I}$ for type I reduction of the semimajor axis is related to the timescale $\tau_{\rm e}\approx1.4\tau_{\rm i}$ for the damping of the eccentricity and inclination by (Tanaka and Ward, 2004)

Author's personal copy

$$\tau_I = \frac{1}{C_a \Omega_K}\left(\frac{M_J}{m_s}\right)\left(\frac{M_J}{r^2\sigma_g}\right)\left(\frac{H}{r}\right)^2 = \frac{C_e}{C_a}\frac{\tau_e}{(H/r)^2} \qquad [13]$$

where C_a and C_e are constants of order unity.

The solid inflow is represented by adding orbiting masses with random positions for $r < r_c$ at a rate proportional to F_{in}/f. Constraints on CPU time for the calculations forced the added objects to be 50–600 km in radius – far larger than the small objects expected to follow the gas during the inflow. But once in orbit, the rapid accretion times imply that the particles would coagulate into sizes that are decoupled from the gas in times less than the viscous spreading time of the gas and to accumulate into larger objects on timescales short compared to gas drag timescale τ_{GD}. If these assumptions are valid, starting with 50–600 km objects to accommodate the computational constraints should have little effect on the outcomes for satellites much larger than these sizes. Collisions within the disk are treated as inelastic mergers.

Since the satellites continue to accrete objects after they reach mass m_{crit}, their migration into Jupiter is hastened, and the value of M_T/M_J given by eqn [12] is an upper bound. This is illustrated in **Figure 1** for a constant and uniform flux density F_{in} ($\gamma_{in} = 0$), $\tau_G = 5 \times 10^6$ years, $r_c = 30R_J$, and $\alpha/f = 10^{-6}$, 5×10^{-5}, 5×10^{-4}. The solid lines indicate the oscillating values of M_T/M_J resulting from the occasional loss of a satellite into Jupiter as a function of time normalized by the accretion timescale, and the dotted lines are the values of M_T/M_J from eqn [12], both for the indicated values of α/f. Generally, the total mass of solids delivered to the disk will exceed M_T, so that several satellite systems with $M_T/M_J = 2 \times 10^{-4}$ would be provided with the earlier ones being lost to Jupiter.

Figure 2 shows that for time-dependent influx rates with $\tau_{in} \sim 1.5 \times 10^6$ years, outer radius for flow deposition $r_o = 30R_P$,

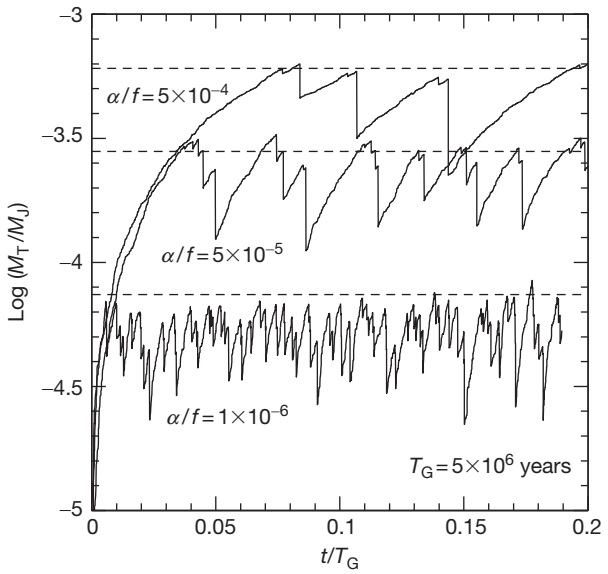


Figure 1 Total mass of satellites as a function of time for a constant rate of inflow for three values of α/f. Modified from Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441, 834–839, with permission.

$\gamma_{in} = 0$, and $\alpha/f = 6.5 \times 10^{-5}$ with $\alpha = 6.5 \times 10^{-3}$, a system of satellites resembling the Galilean satellites is produced with $M_T/M_J \sim 3 \times 10^{-4}$. Recall that M_T/M_J is almost independent of τ_G (eqn [12]) but that disk temperatures are affected by F_{in} (with effective disk temperature $T_{eff} \propto F_{in}^{1/4}$ (Canup and Ward, 2002)). Early in the simulation when the influx is high, the masses of the satellites are given by eqns [11] and [12], but the satellites accreted will be rocky. The properties of the disk must evolve to the point where the midplane temperature of the disk is below 200 K at least beyond the point in the disk where Ganymede accretes a significant fraction of its mass and the accretion of at least Callisto has a timescale $> 10^5$ years to accommodate the latter's possible lack of full differentiation. This timescale requirement generally persists even if Callisto turns out to be differentiated (McKinnon, 1997; Peale, 1999), since influx to the disk and hence the maximum accretion rate are limited by the low disk temperature constraint needed for ice at Ganymede. As the influx wanes, the disk temperature is expected to drop and the component of ice in the satellites is increased. A final generation of Jovian satellites from solar composition inflow with $\tau_{in} = 10^6$ years would form within a disk having temperatures below 200 K exterior to 15 R_J with $\alpha =$ a few $\times 10^{-3}$, a disk opacity $\kappa = O(0.1)\,\mathrm{cm^2\,g^{-1}}$, and a planet temperature $T_J = 500$ K (Canup and Ward, 2002). The constraint of the outer satellites having a large icy component is thereby satisfied. The corresponding timescale for the accretion of the final generation of satellites is $\sim\tau_{in} \sim 10^6$ years, which is consistent with only partial differentiation according to the criteria of Stevenson et al. (1986) and Barr and Canup (2008). Neither the ratio of the total mass of the satellites to the planetary mass, M_T/M_J, nor the distribution of the final satellites is affected much for $-1 < \gamma_{in} < 1.8$, so the results are given in Canup and Ward for $\gamma_{in} = 0$. However, whether most of the mass is deposited close to the planet or far from the planet should affect the distribution of the ice fraction, but this effect of γ_{in} has not been explored. In addition, larger values for γ_{in} may reduce the number of final large satellites (Ogihara and Ida, 2012).

In reality, the satellites are likely to be trapped into orbital resonances sometimes through differential migration of the forming satellites. Such captures were common in the integrations of Canup and Ward (2006) when there were larger outer satellites catching up with smaller inner satellites through type I drift (**Appendix A**). Peale and Lee (2002) demonstrated an assembly of the Galilean satellites into the set of 2:1 MMRs that comprise the present Laplacian relation, where Ganymede is assumed to migrate on the 10^5-year timescale of Canup and Ward and Io and Europa migrate on timescales that are double this value. All the current properties of the resonant system are produced in this scenario, which also requires significant tidal torque from Jupiter to prevent disintegration of the resonant structure from dissipation in Io. However, the assumption of the relative migration rates is ad hoc, and it remains to be shown that detailed plausible disk properties including the opacity as a function of temperature and thereby position in the disk (**Appendix A**), assumed viscosity, etc. can lead to relative migration rates that are consistent with the hypothesis. Still, during the satellite accretion, it is relatively certain that inner small satellites will be trapped into orbital resonances with outer larger satellites that are migrating faster. Such

Author's personal copy

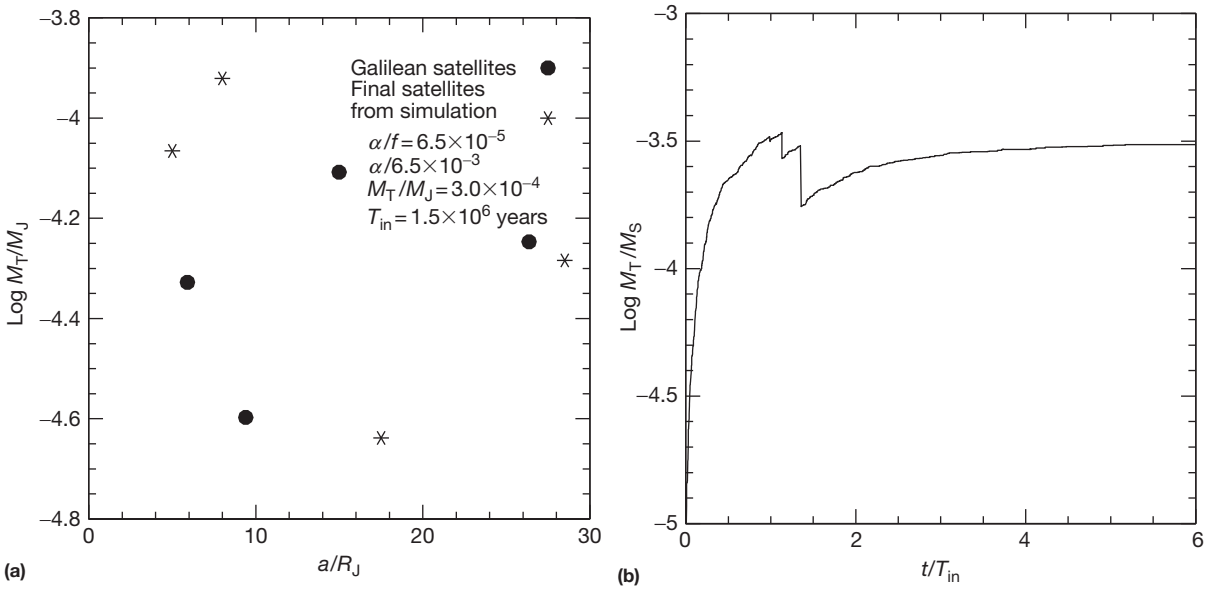


Figure 2 An example of a Jupiter-like system for a waning influx of material onto the disk. (a) Comparison of the masses and locations of the real Galilean satellites with the final set of satellites of a Jupiter-like system from a simulation of a waning influx of material. (b) The time dependence of M_T/M_J for the simulation resulting in the final set of satellites shown in (a). Modified from Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441, 834–839, with permission.

trapping in resonances will hasten the demise of the small satellites interior to larger satellites, where such effects are naturally accounted for in the simulations of the mutual gravitational interactions during differential migration.

The Canup–Ward model successfully produces systems similar to the Galilean satellites (Figure 2). Its virtues include the following: (1) The model is formulated within a plausible model of a waning accretion of the planets, naturally producing not only the observed satellite–planet mass ratios for Jupiter but also those for Saturn and Uranus. (But see Sections 10.17.5 and 10.17.6 for possible caveats.) (2) The waning accretion causes the disk to evolve to sufficiently low disk temperatures to allow water ice condensation and incorporation of that ice into the satellites. (3) The main conclusions are insensitive to the inflow rate and only weakly dependent on the notoriously uncertain value of α. (4) The last generation of Jupiter's satellites accretes sufficiently slowly to allow possible incomplete differentiation of Callisto. (The latter constraint could be relaxed somewhat if it turns out that Callisto is in fact completely differentiated (McKinnon, 1997; Peale, 1999).) (5) The theory is clearly formulated in terms of parameters whose values can be adjusted and the consequences ascertained as new observations and theory provide more constraints. (6) Numerical simulations are consistent with the analytic approximations used in estimating the critical satellite mass attained before loss to the planet and for the total mass fraction M_T/M_J. (7) The simulations produce satellite systems similar to those observed for Jupiter, Saturn, and Uranus for plausible values of the parameters.

Possible caveats to these conclusions include the following: (1) The model requires that the planet contracts to a scale smaller than the satellite system while the nebula is still present, so that an accretion disk forms. While this appears consistent with current contraction models (e.g., Papaloizou and Nelson, 2005; Ward and Canup, 2010), the alternative picture in which a disk is created by spinout from a planet that contracts after the nebula has dissipated would likely produce a very different outcome. (2) A disk with viscosity α values comparable to those inferred for circumsolar disks is assumed. This seems consistent with what is needed to explain the accretion of mass to the planet, given the relatively high specific angular momentum expected for the inflowing gas. But quantitative models justifying such α values in this environment are lacking. Although the effect of the value of α on the conclusions is suppressed by small fractional exponents, the range leading to systems consistent with those of the gaseous planets still implies relatively large values of α. What values of α can be expected from the influx rate as a function of time or from other means of generating turbulence such as stirring by the satellite embryos themselves or MHD-generated turbulence? (3) The model assumes that the outer edge of the inflow region corresponds roughly to the observed scale of the outermost regular satellites. Determination of r_c from hydrodynamic simulations is very challenging, due to the required high resolution, the need to account for the tendency of the circumplanetary disk to viscously spread, and the dependence of results on the assumed nebular scale height. But on balance, current models suggest that the inflow region extended somewhat beyond the current large outer satellites (e.g., Ward and Canup, 2010). Are there processes that would have inhibited satellite accretion in the outer disk? Or could r_c be set by another process, such as the radial extent of a midplane dead zone where MHD turbulence is inactive (Lubow and Martin, 2013)? (4) The assumption that all of the solids are delivered with the gas is dependent on the view that much of the solid mass is in sizes much less than a meter and that small particles do penetrate the gap. An abundance of small particles is predicted by current 'pebble accretion' models, although their

Author's personal copy

delivery to the disk in the presence of the gap should be more thoroughly investigated. How does a realistic estimate of the solid particle size distribution affect the solid inventory of the disk? Are there viable alternative schemes for solid delivery? (5) The critical dependence on disk properties on the opacity affects in particular the inferred disk thermal profile for a given inflow rate and how type I migration rate varies with r. It is possible that the latter could affect the relative growth rates of satellites at various distances. The Canup–Ward models do not include a time-dependent opacity nor a disk that may evolve from optically thick to optically thin. Has the assumption of a constant opacity seriously altered the evolution?

Sasaki et al. (2010) considered a Canup and Ward-type model, only with two important differences. First, they terminate the accretion shortly after Jupiter opens a gap in the heliocentric nebula, which leaves a hot massive disk, which they assume is quickly reduced by a short viscous spreading timescale. This leaves the satellites that are currently in the disk to interact with the remaining gas. The properties of the disk and satellites therein are now 'frozen' at the time of gap formation, and the disk is now closed. In essence, this marries one of the main premises of the SEMM model (an abrupt end to the inflow upon gap opening) to an actively supplied disk model. A unique and crucial assumption in their model is the creation of an inner cavity in the disk through magnetic coupling to the rapidly rotating Jupiter that extends out to the synchronous orbit (2.25 R_J assumed). This inner gap is motivated from inferences about the differences between classical and weak-line T Tauri stars, where the former's properties are assumed to result from formation of an inner cavity created by the magnetic field of the star. An eccentricity barrier at the inner disk edge (Ogihara et al., 2010) is proposed as the means of stopping the satellites at the boundary. Here, a satellite crossing the boundary because of its eccentric orbit forced by a MMR experiences a positive torque because of a tailwind near apoapse but has no negative torque at the periapse because it is not in the disk on this part of its orbit. There are at least two caveats for this hypothesis: (1) Sakaki et al. adopted the timescales obtained by Ogihara et al. (2010) for type I drift in an isothermal disk. These timescales are appropriate to opposing inside and outside torques. These type I drift timescales are applied to all the satellites in the resonances – even to the satellite at the disk's inner edge. But the outer Lindblad resonances are still in the disk for that satellite and there are no opposing torques from the inside, so the satellite will experience a continued type I torque forcing it toward Jupiter that is now enhanced because there are no interior Lindblad resonances contributing an opposite torque. The timescale for the type I drift is thereby reduced for the satellite at the disk edge from that which was assumed. A possible solution could be obtained by considering the opposing action of the corotation torque at the outer edge of the disk cavity (Massett et al., 2006). (2) Type I timescales are derived for a circular satellite orbit, but the forced eccentricity of the satellite at the disk edge introduces additional resonances in the disk which contribute to the torque. A complete treatment of the disk edge dynamics must be completed before one can have confidence that the eccentricity trap will work.

If the barrier does exist, subsequent satellites are now trapped in a series of MMRs and migration of the whole stack of satellites is resisted by the cavity barrier holding the innermost satellite in place. Once the gas infall terminates, the disk is depleted by viscous spreading and the disk temperature decreases exponentially with time. When the mass of the trapped, resonantly bound satellites exceeds the remaining mass in the disk, the barrier is breached and the innermost satellite is forced into Jupiter presumably still in the MMR with the outer satellites. The next satellite in line then is halted at the barrier but may be pushed into Jupiter as the remaining satellites continue to accrete mass from the disk. The ice content of the outer satellites increases as the disk cools and the ice line moves in. When the gas content of the disk is sufficiently depleted, migration stops leaving a system comparable to the current Galilean satellites.

When Sasaki et al. considered an abrupt cessation of inflow to the disk, their outermost large satellite typically forms too quickly to be consistent with an incompletely differentiated Callisto. However, by assuming a more gradual reduction in the infall, they found a $\sim 10^6$-year accretion time for the outermost body. Callisto accretes slowly in the outer part of the disk cool enough to accrete a major fraction of ice and migrates so slowly that it is often not trapped into a resonance.

Sasaki et al. modeled satellite accretion using a semianalytic procedure developed by Ida and Lin (2004, 2008) to model planet formation. Ogihara and Ida (2012) considered the same physical conditions as Sasaki et al. but modeled satellite accretion with an N-body simulation. They found similar results for Jupiter-like systems, including the finding that a Galilean-like configuration in which several satellites are captured into mutual 2:1 resonances is a very common outcome when an inner disk cavity is included with the eccentricity trap.

There are several other aspects of the Sasaki et al. (2010) proposal that are controversial: (1) The assumption that accretion is halted at the time of gap formation is not consistent with the global 3-D calculations of Bate et al. (2003), who found accretion is not halted until $M \sim 10 M_J$. (2) There is no analysis of the disk ionization needed to establish the magnetic coupling to the gas that creates the inner cavity. (3) It is not clear that the large magnetic field necessary to effect the disk standoff is actually to be expected. On the other hand, the existence of a cavity at some point may well be required to explain the current subcritical rotation rates of Jupiter and Saturn (Takata and Stevenson, 1996). (4) And their model assumes a satellite accretion timescale that increases with orbit radius, whereas in an inflow-regulated disk, this timescale decreases with increasing r. This appears likely to have affected the architecture of their final satellite systems, particularly in the case of Saturn-like systems without an inner cavity (see the discussion in the next section).

The Canup–Ward (2002, 2006) model is a reasonably complete and plausible model of satellite formation within a model of the latter stages of the accretion of Jupiter, albeit not without caveats. It is consistent with the composition and positional constraints and produces the gross features of the satellite systems of the giant planets for plausible choices of the controlling parameters, which differ from planet to planet. Some choices of the parameters yield systems comparable with the observed Galilean satellites. The vertical accretion onto the disk is consistent with the high-resolution 3-D simulations. The accretion of the satellites in the circumplanetary

Author's personal copy

disk is specifically treated as a function of disk properties that change as the accretion winds down, where plausible timescales for the various processes are assigned that are consistent with the observed properties of the satellites. The encouraging conclusions from the model depend on the assumed value of α, the presence of sufficient small solids that can be efficiently delivered to the disk by the gas, and the disk's thermal model (and therefore the assumed opacity). Refinement of the model should develop realistic constraints on these parameters.

10.17.5 Saturn System

The Saturn system is far different from the Jupiter system, and aside from the conclusion that the regular satellites formed in a cold, dissipative disk, the details of the formation of the satellites therein are less understood. The Saturn system is dominated by a single satellite, Titan, which contains 96% of the combined mass of all Saturn's satellites. Whereas Jupiter's satellite densities decrease progressively with increasing distance from the planet, at Saturn, they are strikingly nonmonotonic: the densities of the satellites Mimas, Enceladus, Tethys, Dione, Rhea, Titan, Hyperion, and Iapetus are 1.15, 1.61, 0.97, 1.48, 1.23, 1.88, 0.57, and 1.08 g cm^{-3}, respectively (e.g., Jaumann et al., 2009). These densities imply a substantial ice content for all the satellites, but in common with Jupiter, the fraction of the planet's mass contained in the satellites is near 2×10^{-4}. Interior to these satellites are Saturn's main rings, which together contain about the mass of a Mimas-sized satellite, and multiple smaller ringmoons. The ringmoons are dynamically young (e.g., Esposito, 1986; Goldreich and Tremaine, 1982) and may be recent products of the collisional expansion of ring material beyond the Roche limit (Charnoz et al., 2010). The small mass of the satellites other than Titan, the lack of any rock-dominated satellites inside Titan's orbit, the nonmonotonic distribution of densities, and the origin of the rings comprise the major mysteries of the formation of Saturn's satellite system.

While the low density of Hyperion (mean radius $\langle R \rangle \sim 150$ km) is significantly affected by porosity, this is expected to have a minor effect on the densities of larger Tethys ($R = 530$ km) and Iapetus ($R = 718$ km). In particular Tethys' density is the lowest of all outer solar system objects having $R > 500$ km and, for plausible densities for rock and nonporous ice, implies a rock content between 0% and 6% (e.g., Thomas, 2010). The main rings are also more than 90–95% pure water ice (e.g., Cuzzi et al., 2010). These are extremely ice-rich compositions compared to a bulk solar composition of roughly half rock/half ice. In contrast, the mass fractions of rock in Enceladus, Dione, and Titan are similar to those of Ganymede and Callisto. The current high density of Enceladus may reflect loss of water from the observed plumes at the south pole (Porco et al., 2006). The loss rate of neutral H_2O is estimated to be ~75–1600 kg s^{-1} (e.g., Fleshman et al., 2010). The history of activity on Enceladus is totally unknown, but if current activity has persisted for a significant fraction of Enceladus' lifetime, the satellite could have lost a significant fraction of its ice and started with a more Mimas-like density.

Titan is about 50% water as deduced from its mean density. Estimates of its moment of inertia from Cassini data suggest that its interior may be incompletely differentiated like Callisto (Iess et al., 2010), although this result assumes hydrostatic equilibrium and is therefore uncertain due to the possible contribution of nonhydrostatic structures (e.g., Gao and Stevenson, 2012). The mole fraction of methane in Titan's atmosphere is about 5% in the lower atmosphere and the rest is essentially all N_2 except for traces of argon, a tentative detection of neon, and small amounts of heavier hydrocarbons (Niemann et al., 2005, 2010). Aside from a mole fraction of $2.1 \pm 0.8 \times 10^{-7}$ for ^{36}Ar and $2.8 \pm 2.1 \times 10^{-7}$ for ^{22}Ne, no other primordial inert gases have been detected (Niemann et al., 2010).

If the N_2 in Titan's atmosphere were trapped in the icy planetesimals that went into Titan as direct condensation from the gaseous phase, or if it were trapped in amorphous ice or as a clathrate hydrate, the noble gases would have also been trapped. The low upper bound on the noble gases implies then that the N_2 came into Titan in the form of NH_3 (Niemann et al., 2005, and references therein). The nondetection of krypton and xenon led to several suggestions for the origin of the atmosphere consistent with the lack of these gases. One possibility is that the icy planetesimals formed at temperatures >75 K so that the noble gases and CH_4 were not captured, but NH_3 and CO_2 were. However, Owen and Niemann (2009) showed that the detection of Ar but not Kr or Xe could instead be due to the limited sensitivity of the Huygens probe instrument given the relative abundance of these elements in other solar system bodies, making such arguments inconclusive.

Heavier hydrocarbons are thought to be products of photochemistry initiated by photodissociation of methane (e.g., Niemann et al., 2005). The relatively short timescale for the survival of CH_4 from photodissociation of 10–100 million years (Wilson and Atreya, 2004; Yung et al., 1984), together with the lack of any isotope fractionation in the carbon of CH_4 like that seen in nitrogen and oxygen, means that CH_4 must be continuously supplied from Titan's interior (Niemann et al., 2005). This scenario may require the continuous supply of CH_4 to Titan's atmosphere to be created through chemistry within Titan (Zolotov et al., 2005). We are certainly left with the constraint that the temperature eventually becomes cool enough in the Saturn disk to allow water ice to persist but perhaps not so cold as to permit stable methane clathrates.

Most models of the Saturnian protosatellite disk invoke a surface density, midplane temperature, and pressure that increase with decreasing r as in the Jupiter examples in the preceding text. This would normally favor the formation of rocky satellites close to the planet with icy satellites further out, yet the inner satellites are ice-rich at Saturn. For an accreting disk, the rate of heat generated by viscous dissipation is proportional to the inflow rate (Canup and Ward, 2002), and so for very slow rates of inflow, ice can be stable in the inner disk. In the limit that viscous dissipation is unimportant and the planet's luminosity is the dominant heat source in the disk, the ice stability distance is within three Saturnian radii once the planet's temperature falls below about 500 K (e.g., Canup, 2010). A cold, solar composition inner disk would produce ice-rock satellites, rather than extremely ice-rich objects like Mimas and Tethys. Furthermore, a smooth disk surface density profile consistent with Titan's mass would generally yield much more massive satellites in the inner disk than those observed. Thus, the inner Saturnian satellites are not easily explained by a standard disk model.

Author's personal copy

The Mosqueira–Estrada SEMM model for the Saturnian disk (2003a) starts with a smoothed minimum mass nebula based on the total satellite system mass, rather than on the specific distribution of mass in the current inner satellites. The temperature and surface mass densities are assumed to vary as $1/r$ as in the Jupiter model. As in the Jovian SEMM disk, a steep surface density gradient is assumed just outside the centrifugal radius defined earlier, which for Saturn is just outside the orbit of Titan. We argued earlier that it is improbable that such a steep gradient could have been maintained during the disk's formation, when the viscosity required to transport angular momentum outward so that the planet can accrete mass (e.g., Estrada et al., 2009) would have smoothed out such variations (e.g., Ward and Canup, 2010). But interestingly, capture of Hyperion into its current 4:3 MMR with Titan seems to require such a steep density gradient ($\sigma_g \sim r^{-3}$) if Hyperion formed as Titan formed and was brought into the resonance by gas drag, as seems likely (Lee and Peale, 2000). As Hyperion nears its final mass, which was probably considerably larger than its current mass because pieces chipped off Hyperion typically do not remain in the resonance and are not reaccreted, gas drag overwhelmingly selects the 3:2 MMR preventing migration to the 4:3 if the surface mass density decreases less rapidly than r^{-3} (Lee and Peale, 2000).

In the SEMM model, Titan forms in the inner, dense disk in 10^4–10^5 years and would likely be fully differentiated (Mosqueira and Estrada, 2003a). Melting of ice occurs in Titan's interior if it forms in $\lesssim 10^6$ years even in the limit that there is no radiogenic heating and the satellite forms from small impactors that deposit their energy close to the surface (Barr et al., 2010), so that substantially more rapid formation implies differentiation. In the SEMM model, Titan is saved from loss due to inward type I migration by opening a gap in the very low viscosity disk, with the disk then removed due to nonviscous processes (e.g., photoevaporation) so that Titan is not lost by type II migration. Mosqueira and Estrada form Iapetus in their outer, low surface mass density disk, where it migrates inward by type I drift to the point where this torque vanishes because of the steep density gradient between the inner and outer disks. The ice-rich composition of Iapetus is then ascribed to the preferential delivery of water to the outer disk by the ablation of differentiated heliocentric planetesimals, because objects passing through the low-density outer disk become hot enough to lose their ice but not their rock (Mosqueira et al., 2010).

To account for the ice-rich compositions of the inner moons, Mosqueira and Estrada (2003b) suggested that the inner Saturnian disk was initially too warm for ice and that early rocky embryos that formed there were lost as they spiraled into Saturn. Subsequently, as the disk cooled, the ice condensed to form ice-rich satellites about 10^5 years after the end of Saturn's accretion when the disk gas surface density had decreased to $\sim 10^4$ g cm^{-2}. However, the survival of the inner moons would then require a very rapid removal of the inner gas disk just after they formed; otherwise, they too would be lost to gas drag in the order of a few $\times 10^4$ years (for Mimas) to $\sim 10^5$ years (for Rhea). Mosqueira and Estrada (2003b) qualitatively discussed seven different reasons for why there might be no large satellites interior to Titan. Their preferred explanation is that the tidal torque from a fully formed Titan removed the inner disk before other large inner satellites could form, but how a tidal torque causes the inner material to flow rapidly inward is not explained. For a disk with no inflow, satellite growth would generally proceed more rapidly in the inner disk due to shorter orbital periods and a higher disk surface density, implying that if the inner disk contained a mass much larger than that in the current inner moons (and this is the case for the SEMM disk), satellites much larger than the current inner Saturnian moons would likely form before Titan formed.

An underlying assumption of the SEMM or any minimum mass model is that the satellites form after the planet has finished accreting gas, so that there is no inflow supply to the disk. However, the period of inflow to the disk is long (comparable to the lifetime of the nebula), while the timescale for solids orbiting in the disk to accrete into satellites is short. Thus, it seems more probable that Saturn's satellites formed throughout the lifetime of its disk as it was being supplied by an inflow, rather than waiting until after the inflow had ended, similar to the conditions advocated for the Jupiter system. The Canup–Ward model (2002, 2006, 2009), as well the simulations of Sasaki et al. (2010) and Ogihara and Ida (2012), adopt this premise.

Canup and Ward (2006) simulated the accretion of Saturn's satellites from an actively supplied disk with an N-body model as described in **Section 10.17.4**. Jovian-like systems with four or five similarly sized large satellites were the most common outcome in the Canup and Ward simulations. But depending on the stochastic timing of when the inflow ends relative to the oscillations of satellite growth and loss seen in **Figure 1**, other system structures were also produced. **Figure 3** shows a simulation with $r_c = 30R_S$, $\gamma_{in} = 0$, $\alpha/f = 6 \times 10^{-5}$, and $\alpha = 6 \times 10^{-3}$ that produced a Saturn-like system with one large satellite and numerous small satellites (Canup and Ward, 2006). Initially, four large satellites formed, but the inner three are lost to collision with the planet as the inflow wanes. After the last large inner satellite is lost, the final inflow delivers enough mass to produce Dione- and Rhea-sized satellites in the vacated inner region of the disk. During this period, the inflow rate is so slow that the ice stability line moves inward to a few Saturnian radii. When the infall ends and the gas disk dissipates, a seven-satellite system remains with $M_T/M_S = 1.8 \times 10^{-4}$ and the most massive satellite at $\sim 14.6R_S$ that contains 70% of the mass of the total satellite system mass. Its close orbital spacing to the satellite at $11.3R_S$ suggests a future collision may occur. There are two satellites bracketing the position of Iapetus but with larger orbital eccentricities. The inner satellites in this simulation are sufficiently close together that collisions are likely with a reduction in the total number of satellites.

The Canup and Ward simulations were the first to show that a smooth disk could produce approximate analogs to Saturn's satellite system. However, systems with only one large satellite were not common, and they did not produce any systems in which one satellite contained $>70\%$ of the final satellite system mass. Although their calculations produced the small, ice-rich satellites inside Titan's orbit, they could not account for the large increase in the ice fraction of these satellites over the half rock/half ice content of the influx nor could they account for the nonmonotonic distribution of the rock–ice content of the inner satellites. This difficulty has motivated an alternative scenario described in the succeeding text that

Author's personal copy

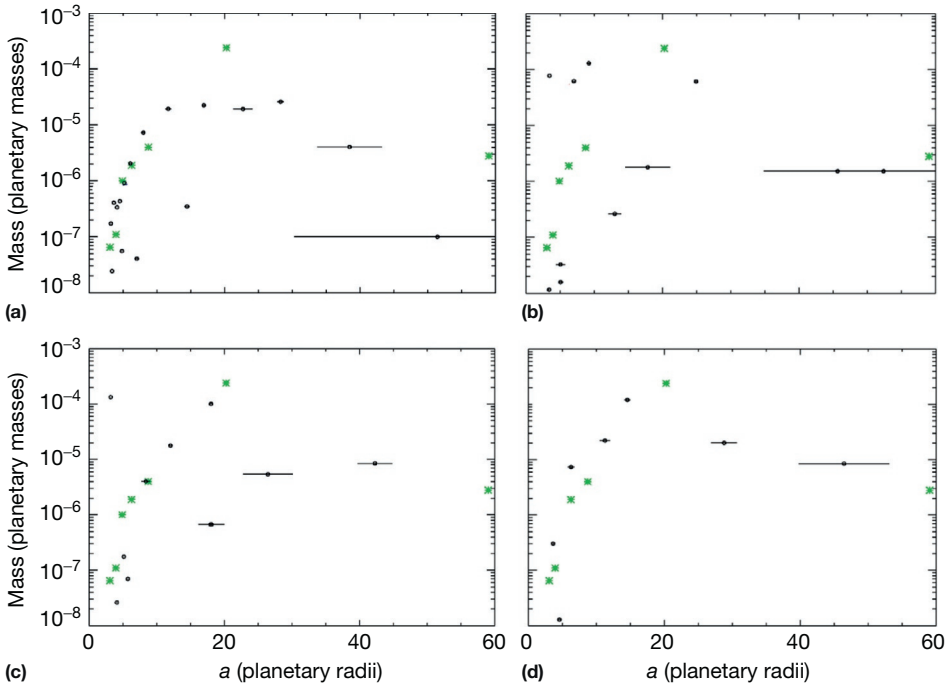


Figure 3 Results of a satellite accretion simulation from Canup and Ward (2006) that produced a Saturn-like system of satellites. Here, the inflow to the disk is assumed to decrease exponentially with time with a time constant τ_{in}. Black circles show the simulated satellites (with horizontal lines proportional to orbital eccentricities); Saturn's satellites are shown as green stars. (a) Multiple satellites form as solid material flows into the disk ($t = 0.2\tau_{in}$). Once satellites grow to $M \sim$ few $\times 10^{-5}$ planet masses, they begin to migrate inward. (b) At $t = 1.4\tau_{in}$, the system resembles Jupiter's Galilean system, with four similarly sized large satellites. The inner three large satellites are lost to collision with Saturn, with the last large satellite lost at time $t = 3\tau_{in}$ (c), when the inflow rate has slowed substantially. The final system (d) at $t = 10\tau_{in}$ has a single large Titan-like satellite orbiting at 15 planetary radii.

produces nearly pure ice inner satellites as well as a nearly pure ice ring system.

Sasaki et al. (2010) simulated the accretion of satellites using a semianalytic, Monte Carlo model. This model closely reproduced the oscillations in the total satellite system mass seen in Figure 1. However, when considering the same conditions as Canup and Ward (an inflow that decayed exponentially with time and a planet without an inner cavity in its disk), they found the most common result was a Saturn-like system with a single large satellite. Ogihara and Ida (2012) performed N-body simulations of the conditions considered in Sasaki et al. and found Jupiter-like systems were the most common result for disks without inner cavities, similar to the Canup and Ward (2006) findings. All three works use the same expressions for the rate of type I migration and the quasi steady-state gas surface density. However, their treatment of the satellite growth rate differs. Sasaki et al. used a semianalytic procedure in which they use a satellite accretion timescale that is proportional to the local orbital period, leading to $\tau_{acc} \propto r^{5/4}$ (their eqn [13]) and a longer growth timescale for outer satellites compared to inner ones. But in an inflow-supplied disk, the rate of satellite growth is controlled by the rate of delivery of material to the disk rather than by the local orbital period (Canup and Ward, 2002). For a uniform inflow per area, this leads to $\tau_{acc} \propto 1/r^{4/5}$ (Canup and Ward, 2006) or $\tau_{acc} \propto 1/r^2$ (Ogihara and Ida, 2012), depending on how the width of a satellite's feeding zone is defined. In both the Canup and Ward (2006) and Ogihara and Ida (2012) simulations, satellites in the outer disk thus have shorter growth timescales than those in the inner disk.

By imposing a slower accretion in the outer disk, the Sasaki et al. model would tend to artificially delay the growth of outer Titan-sized objects in the time it takes an inner Titan-sized object to migrate inward. This appears to have biased their results in favor of single satellite systems. Ogihara and Ida (2012) suggested that a radially nonuniform inflow, in particular one that is centrally concentrated, might be able to produce a similar result, although this has yet to be demonstrated.

Can the very ice-rich compositions of the inner moons be explained in the context of such models? Canup (2010) proposed that as the last of the large satellites was lost, tidal stripping from its outer layers created Saturn's rings and ultimately the inner satellites out to Tethys. This scenario requires that this last satellite be differentiated as it crossed the Roche limit with the outer layer being nearly pure ice or an ice-covered ocean. This possibility is supported by Canup (2010) who notes that accretion should bring a Titan like body to near the melting point of ice even on the relatively long accretion time scale of 10^6 years expected under the waning influx of material to the disk. A sustained eccentricity $\sim 2 \times 10^{-3}$ for $k_2/Q \sim O(10^{-2})$, which is a balance between excitation by perturbations of other satellites and damping by tidal dissipation in the satellite, would provide somewhat more than the latent heat of melting an ice shell in the $\sim 10^4$ years the satellite spends near the Roche limit before colliding with Saturn. This heat input plus that from accretion is not

Author's personal copy

unlikely to produce the needed differentiation into an ice water mantle over a rocky core.

When the satellite's orbit decays within the Roche limit set by its mean density (located at $\sim 1.75R_S$ for a Titan-density satellite, where R_S is Saturn's current mean radius), tidal distortion begins to remove material from its ice shell. The removal of ice causes the satellite's mean density to increase until the remnant satellite is marginally stable. This continues until either the remnant collides with the planet or its higher-density rocky core disrupts. The Roche limit for rock is at $a_{R,rock} \sim 1.5R_S$. Planet contraction models predict that the newly formed Saturn's radius would have been between $R_P = 1.5R_S$ and $R_P = 1.7R_S$ (e.g., Fortney et al., 2007; Marley et al., 2007). For $a_{R,rock} \leq R_P$, the remnant satellite collides with the planet before its core disrupts, and tidal stripping produces a pure ice ring.

Canup (2010) used SPH to simulate a differentiated Titan-sized satellite as it spirals inward from its initial Roche limit. The total mass of ice produced via tidal stripping depends on the location of the planet's surface: In the limit that $a_{R,rock} = R_P$, $\sim 10^{25}$ g of ice is stripped into orbits exterior to the satellite ($\sim$10% of the original satellite's mass), while if $a_{R,rock} < R_P$, less ice is produced before the remnant satellite hits the planet. As each ringlet of ice is stripped, strong shepherding torques from the remnant satellite rapidly repel it, driving exterior material into orbits beyond the Roche limit for ice in 10^2 years. Here, the stripped material accumulates in a pure ice moon whose final mass is $m_o \leq 10^{25}$ g. Canup estimated that the location of synchronous orbit around a newly formed Saturn would be located at about $3R_S$ due to the slower rotation of the more distended planet, so that this ice moon eventually spirals inward due to planetary tides, but at a slower rate than the original satellite due to its smaller mass. For planetary tidal parameters $(k_2/Q)_P$ (with likely values $10^{-6} \leq k_2/Q_P \leq 10^{-5}$; e.g., Charnoz et al., 2009), the ice moon takes $\sim 10^6$ years $[5 \times 10^{-6}/(k_2/Q)_P]$ $[10^{25}\, g/m_o]$ to decay within the ice Roche limit, where it disrupts into a massive ice ring. This timescale is long enough that the ring forms after the remnant satellite has collided with Saturn and when the circumplanetary gas disk has essentially dissipated. Mosqueira and Estrada (2011) claimed that such a ring would be destroyed by gas drag, because its surface density decreases with time (and with it, the ring's lifetime against gas drag) as the ring spreads. But a ring formed at the very end of gas accretion can survive, because there is also a decrease in the gas surface density with time both before the ring forms and as the ring evolves due to the waning inflow, which they did not consider. However, their point is a good one for any rings formed by satellites that decayed within the Roche limit earlier in a planet's gas accretion.

The Canup model implies an initial ring that is a factor of 20–500 times more massive than Saturn's current rings. In contrast, prior ring origin models produce an initial ring comparable in mass to the current rings (Charnoz et al., 2009; Dones, 1991). Salmon et al. (2010) found that such dense rings can evolve to something resembling Saturn's ring by viscous spreading after $O(10^8)$ years, but it is not clear how the ignored MMRs with the closest satellites such as Mimas would affect this evolution. As a ring viscously spreads, it spawns satellites from its outer edge (e.g., Charnoz et al., 2010), and Canup estimated that for such a massive ring, the first spawned satellite would have a mass comparable to that of Tethys, whereas Charnoz et al. (2011) found the first mass spawned to be comparable to Rhea's mass. Spawned moons evolve outward due to first resonant interactions with the ring and then by tidal evolution, with the mass of spawned moons decreasing with time as the mass of the ring decreases due to viscous spreading (Charnoz et al., 2010, 2011). Mimas, Enceladus, and Tethys would be by-products of the formation of Saturn's rings. However, the dynamics of such satellite orbit expansions have not been investigated (see succeeding text).

Extending this idea, Charnoz et al. (2011) proposed that all of the satellites interior to and including Rhea were spawned from an initial ring containing a few $\times 10^{24}$ to few $\times 10^{25}$ g. Dione and Rhea contain $\sim 10^{24}$ g in rock, much greater than the total rock in the inner moons ($\sim 10^{23}$ g) that could be delivered by external bombardment during the LHB (Canup, 2013). The Charnoz et al. (2011) model thus requires an initial ring with a substantial rock component. To leave the rings as essentially pure ice, they creatively suggested that the rock in the rings was initially in the form of large chunks that were able to open gaps in the ring and undergo type II migration. A direct simulation of a Rhea-sized chunk embedded in a ring is used to show that the chunk is expelled from the ring's outer edge, where it would then form the core of a spawned satellite. It is, however, not clear why the ring's rock would be contained in fragments many orders of magnitude larger than the accompanying ice particles, since the strength of rock and ice is similar and because ice and rock particles would likely form rock–ice aggregates of varying mean densities, nor whether a simulation that included a large number of rock chunks would be able to remove them all from the ring with nearly perfect efficiency. A second condition of the model is that Dione and Rhea evolve out to their current orbits through tides, which requires an extremely small tidal Q parameter for Saturn of $Q < 2000$. It is generally believed that the Q is of order 10^4–10^5 for gas giants. Lainey et al. (2012) considered astrometric observations of Saturn's satellites over the last century and found $Q \sim 1700$. This result is controversial, both because it implies a much different Q for Saturn than similar techniques predicted for Jupiter (Lainey et al., 2009) and because their data also showed Mimas migrating inward rapidly, whereas it should be migrating outward due to both tides and ring torques.

There are several important caveats to both the Canup (2010) and Charnoz et al. (2011) models. First, neither can produce the nonmonotonic distribution of satellite compositions without invoking ad hoc degrees of freedom. In Charnoz et al., a pure ice initial Tethys is proposed to result from stochastic variation in how rock is delivered from the initial ring. The Canup model implies an essentially pure ice initial Tethys but then requires additional processes to explain Enceladus' rock content while avoiding the subsequent accretion of too much rock onto Tethys. Next, can the initial satellites accreted outside the Roche radius retreat far enough before the next one is spawned to allow the minimum spacing of the satellites, or is an additional sequence of mergers required? The simulations of Charnoz et al. (2011) show that typically more inner moons are formed than are currently in the system, although they utilize a simplified accretion model. Crida and

Author's personal copy

Charnoz (2012) used an analytic merger model and predicted a mass–distance distribution of satellites in good agreement with that seen at Saturn. However, their accretion model does not include gravitational interactions between the moons, including MMRs.

The most critical constraint comes from the limited tidal evolution of the current satellites. Slow differential expansion of the orbits will lead to a plethora of MMRs that will be encountered, and capture within will lead to eccentricity growth that will destabilize the system. That the tidal evolution of the system has been limited is implied by calculations of evolution of the Saturn system that brings Enceladus to the 2:1 MMR with Dione (Peale, unpublished). Reducing the satellite semimajor axes to start the integrations consistent with their mutual tidal evolution brings Tethys and Dione and, almost at the same epoch, Mimas and Enceladus inside 3:2 MMRs. The integrations had to start outside these resonances to avoid the certain captures therein, which implies that the evolution of the satellite system is limited over the 4.6×10^9-year history to start outside these resonances. There is no obvious means to escape from these resonances and the eccentricity growth therein leads to crossing orbits before the resonances become unstable. The implied lower bound on the average Q of Saturn exceeds 90 000, which is higher than that inferred from the proximity of Mimas to Saturn. The system configuration is such that Enceladus' semimajor axis must start outside 0.9832 its current value if the two 3:2 MMRs are to be avoided.

One way MMRs among the satellites spawned by the massive icy ring in the Canup and Charnoz et al. models can be traversed without capture is to note that the initial orbital expansion of these satellites as they interact with a massive ring will be very rapid compared to tidal evolution. If the orbit of the inner satellite expands the width of the MMR in a time less than the libration period of the resonance variable, capture can be avoided. It is possible that the inner moons (Mimas, Enceladus, and Tethys) achieved nearly their current positions through the ring interactions, and subsequent tidal evolution can be severely limited. But this possibility has not yet been demonstrated. In the Charnoz et al. (2011) model, Dione and Rhea become too distant to experience ring torques and they must migrate over vast distances at the much slower tidal rates. If 90 000 is the real lower limit on the historical tidal Q, such distant tidal migration is unlikely. If Q is much lower than this and the extensive migration could ensue, the capture of Dione and Rhea and the inner satellites into destabilizing MMRs is likely. Details of the orbital evolution of the satellites under the variety of possible constraints discussed have not been investigated.

The challenge of explaining the variation in densities among the mid-sized Saturnian satellites has motivated alternative origin scenarios that form these satellites as intact ejected clumps from large impacts. If the impacting objects are differentiated, such clumps can have compositions that vary significantly from those of the parent objects. Sekine and Genda (2012) considered the collision between a smaller satellite and a Titan-sized body as the latter migrated into the inner disk due to type I migration. Asphaug and Reufer (2013) proposed that an initial multiple moon system at Saturn was destabilized through mutual collisions that ultimately left only Titan and that the current mid-sized moons were clumps ejected from such collisions. Both works use SPH simulations to demonstrate that such collisions can produce large intact fragments having similar sizes and a wide range of rock fractions as observed among Saturn's mid-sized moons. However, both works face a common challenge: how to avoid having the clumps ejected from an impact reaccrete onto the target object. Ejecta are initially on crossing orbits with the target and would generally be rapidly reaccreted. If ejecta are in the form of small particles, then their mutual collision times can be shorter than the reaccretion timescale, and in this case, they can collisionally damp to a ring, which could potentially accrete into separate stable satellites. But this process would also seem likely to remix the compositions of the clumps and seems unlikely to yield multiple satellites with well-separated orbits as observed.

We conclude our discussion of Saturn with Iapetus, which presents a special problem in that its orbit is nearly circular, but inclined by about 7.49° relative to the local Laplacian plane. The latter inclination would imply a capture origin, but because eccentricity damping due to the tidal dissipation in Iapetus is very weak, a capture origin is inconsistent with its current small orbital eccentricity ($e_I = 0.0283$). The Laplacian plane is that plane on which a satellite's orbit precesses at nearly a constant rate with nearly a constant orbital inclination. The Laplacian plane nearly coincides with the equatorial plane for close satellites due to torque from the planetary oblateness, but it coincides nearly with the planet's orbital plane for distant satellites due to the solar torque. An extensive debris disk settling to this plane would therefore be warped as the local Laplacian plane changed from equatorial plane to orbital plane with distance from the planet (e.g., Ward, 1981). The regular satellites accumulated from debris disks at distances where the oblateness of the planet dominated the location of the Laplacian plane. But at Iapetus' distance from Saturn, the local Laplacian plane is inclined by 15° to Saturn's equator, where the ascending nodes on the orbital plane of Saturn's equator (obliquity = 26.7°) and the Laplacian plane coincide. Can Iapetus have accumulated in the debris disk as implied by its small orbital eccentricity yet be inclined to the local Laplacian plane by 7.49° today?

Ward (1981) offered a tidy affirmative answer by pointing out that the debris disk itself contributes to the precession of objects within it. The Laplacian plane thereby would have a lower inclination to Saturn's equatorial plane at the time the disk was present, and Iapetus could have accreted there as assumed in the previously mentioned discussion. Iapetus' orbital plane precesses with a current period of about 3000 years (Ward, 1981), and this period would have been shorter when the disk was present. If the disk could be dispersed on a timescale short compared to the orbit precession period at the time, the Laplacian plane would rotate toward its current position and leave Iapetus with its observed low eccentricity and 7.49° orbital inclination. The short dispersal time necessary, $O(100)$ years, is not consistent with accretion times of the satellites of order 10^3–10^4 years in the theories discussed in the preceding text or for allowing sufficient time for rocky satellites to spiral into Saturn in the inner disk, leading to the formation of icy satellites there. Viscous dissipation of the disk on this short timescale requires values of $\alpha = O(1)$ (Ward, 1981), which is implausible. Because of the difficulties imposed by such rapid disk dispersal, the anomalous orbit of

Author's personal copy

Iapetus is still a puzzle. A capture origin seems unlikely, because sufficient gas drag to circularize the orbit would also have damped the inclination to the local Laplacian plane.

Although the structure of the circumplanetary disk is still poorly constrained and there are uncertainties about the nature of the solid content in the gas inflow, the Canup–Ward model is perhaps a first step toward an understanding of the origin of Saturn's satellite system, since a system with no large satellite interior to Titan appears attainable. The loss of large inner satellites to type I migration provides a means to produce a pure ice initial ring and ice-rich inner satellites. The loss of silicate embryos while the nebula was too warm to allow water ice condensation followed by cooling to allow the remnant ice to accrete onto the small satellites is an attractive alternative (Mosqueira and Estrada, 2003a,b). But there have been no detailed disk evolution models to show that the sequence of events needed for this scenario can occur. Explaining the distribution of densities in the inner satellites remains a challenge, and current ideas ultimately depend on stochastic events, which are difficult to model with great confidence due to inherent uncertainties in the relevant processes. We are much closer to understanding the details of the origin of Jupiter's satellites than we are to understanding the origin of the Saturn system.

10.17.6 Uranus System

The Uranus system of satellites is Jupiter-like in the sense that it contains a few large satellites with $M_T/M_U \sim 1 \times 10^{-4}$ (M_T is the total mass of the satellites). All the satellites out to Oberon with the exception of Miranda have nearly equatorial orbits with densities–radii of the major satellites Miranda, Ariel, Umbriel, Titania, and Oberon of 1.2/236, 1.7/579, 1.4/585, 1.7/789, and 1.6/761 ($g\,cm^{-3}\,km^{-1}$), respectively. Inside Miranda's orbit are 13 small satellites that are the most densely packed in the solar system and are thereby unstable and experience relatively frequent collisions (French and Showalter, 2012). Thirteen remnant rings in the inner system are testimony to the frequent destruction and reassembly of the satellites in the inner system. The Canup–Ward (2006) model can produce a Uranus-like system of the major satellites, where the lower rate of inflow and cooler disk temperatures about the smaller mass planet produce satellites with a mixture of rock and ice that is consistent with the densities. However, these authors note the difficulty of moving the disk material from heliocentric orbit to the equatorial plane of Uranus when the obliquity of Uranus is 98°. On the other hand, if the Uranus obliquity were small when the satellites formed and only later tipped over by an adiabatic process, the Canup–Ward model would be immediately applicable. But with more certainty than the Saturn case, Uranus will not open a gap in the solar nebula, and the influx may have a different distribution over the disk than that appropriate for Jupiter. So, the Canup–Ward model may have to be modified.

Accounting for the maintenance of a coplanar, equatorial system of satellites about Uranus as the planet gains its large obliquity is the major obstacle to understanding their origin. An adiabatic scheme for tilting Uranus, where equatorial orbits would be maintained by adiabatic invariance (e.g., Goldreich, 1965), requires a resonance between the precession rates of the spin axis about the orbit normal and the precession of the orbit about its local Laplacian plane. This requires the spin precession rate to be increased drastically from its current value. Boué and Laskar (2010) accomplished this by assuming that Uranus originally had a massive satellite with an orbital radius of about 0.01 AU. However, Morbidelli et al. (2012) pointed out that the Laplacian plane for such a satellite is close to Uranus' orbital plane and this satellite orbit would not follow the equator as the Uranus obliquity increased. Because of its large mass, this satellite would also hinder at least Oberon and Titania from doing so. Morbidelli et al. concluded that a sudden tilting of Uranus with a giant impact is the only remaining viable means of accounting for the large obliquity, a scheme which had been proposed before (e.g., Slattery et al., 1992).

The most viable timing of the impact is after the planet has completed its gas accretion and thereby already possessed a satellite system. The impact would likely disrupt this system into crossing orbits where collisions would disintegrate the satellites into a debris disk. The latter scenario was verified numerically by Morbidelli et al. (2012), so these authors consider the evolution of a protosatellite disk after the massive collision that tilts Uranus modeled as gravitationally interacting concentric rings. After the tilt, the disk is still in the former equatorial plane, but Uranus' new equatorial plane corresponds to the new 98° obliquity. The rotational oblateness of Uranus, parameterized by the gravitational harmonic coefficient J_2, causes differential precession of the ascending nodes of the disk rings (or of particles in the disk) so the disk becomes a thick torus spanning the equatorial plane. Collisions within the torus damp the ring to the new equatorial plane, where the new disk can accrete into equatorial satellites. However, the J_2 effect competes with the self-gravity of the disk, where the latter dominates for those portions of the disk that are further from Uranus and cause the disk to precess as a unit while maintaining its original inclination to the equatorial plane. Such dominance of self-gravity of the particles was first realized for the epsilon ring of Uranus, where the periapse longitudes precessed together maintaining the ring coherence (Goldreich and Tremaine, 1979).

If the current value of J_2 is assumed, the nodes are randomized by differential precession only slightly beyond Miranda's orbit near 0.001 AU. Beyond Miranda's distance, the inclined ring would lead to satellites highly inclined to the equatorial plane. But existing simulations indicate that the impact should create an equatorial disk near 0.01–$0.03M_U$ extending out to $3R_U$, where R_U is Uranus' radius (Slattery et al., 1992). This equatorial ring enhances J_2 sufficiently to extend the randomization of the node longitudes of the rings comprising the disk out beyond Oberon's orbital distance. The mutual collisions damp the disk down to the equatorial plane where the equatorial satellites can be formed. But if the tilt to the current obliquity is from 0° to 98° with the single impact, the satellite orbits will be equatorial but retrograde! On the other hand, if Uranus had a preimpact obliquity >30°, the probability of prograde orbits exceeds 40% (Morbidelli et al., 2012). This complicates the scenario somewhat as it requires Uranus to have had at least two giant impacts, but such impacts near the end of planetary accretion are thought to have been common (Morbidelli et al., 2012). The equatorial disk formed from the impact probably could

Author's personal copy

not contribute significantly to the major satellites as they are too far away, but contribution to the small inner satellites is possible.

In contrast to the opinion expressed in Section 10.17.4 against starting with a pristine disk for the formation of the regular satellites of Jupiter, here is an example of the creation of a disk around a giant planet from which the satellites containing about 10^{-4} of the Uranus mass could be formed. If the solar nebula had already been removed at the time of impact, there would be no further accretion onto the disk after the disk was formed. If the debris disk from the collisional breakup of the original satellites contained little gas, loss of solids from gas drag would be eliminated, and the total satellite mass would then be $O(10^{-4}\,M_U)$ appropriate for the Canup–Ward model of the original accretion.

An alternative to this scheme has been proposed by Crida and Charnoz (2012). Here, the satellites are created from a massive disk mostly inside the Roche radius that was created at the time of the giant impact that tilted Uranus. The satellites are created one by one just outside the Roche radius as the disk spreads viscously, and they migrate away from the disk from disk torques initially and eventually through tidal expansion of the orbits. This idea has the appealing features that only a single impact is necessary to create a system that will be equatorial and prograde. However, the details of its execution pose some severe problems. First, there most likely would have been an accretion disk about Uranus as it was being formed, and satellites would have formed in this disk. Since Uranus must have been mostly completely formed at the time of the impact, it must have had a preexisting satellite system in its equatorial plane as assumed by Morbidelli et al. (2012). This preexisting system is assumed not to exist in the Crida and Charnoz scheme. Next, the disk formed by the impact is inside about $3R_U$ so that Oberon must be moved from this distance to almost $23R_U$ if it is spawned from the ring. Even if the tidal dissipation function Q for Uranus were sufficiently small to accomplish this feat over the age of the solar system, Titania, more massive than Oberon and at a smaller orbital distance, would have moved out more rapidly. Finally, the disk produced by a Uranus-tilting impact would initially be vapor, which would evolve differently than a disk of solid particles as considered by Crida and Charnoz. It is difficult to see how the Crida–Charnoz scheme could be the origin of the Uranian satellites.

The Slattery (1992) simulation of the giant impact contained only 5000 particles in the proto-Uranus and 3000 particles in a $2M_E$ impactor. Rather substantial changes in the outcome resulted from SPH calculations of the lunar-forming impact on the Earth as the resolution was increased substantially above that of earlier calculations (Canup, 2004b), so it is important to do simulations of the Uranus impact with the highest resolutions currently practical to see if conditions assumed (e.g., the mass and extent of the equatorial disk generated by the impact) in the successful scheme for forming the prograde, equatorial satellites by Morbidelli et al. actually prevail.

10.17.7 Neptune System

A plausible theory has emerged for the current configuration and content of the Neptune system of satellites. Here, the task is simpler than for the other gas planet systems, because Triton's destruction of the original system during its capture has erased all concern for accretion disks, gaps, solar nebula interactions, etc. The dominant characteristic of the Neptune satellite system is the existence of the large satellite Triton (NI) ($R = 1353$ km), in a close ($r \approx 14.3R_N$, $R_N = 24\,766$ km) circular, retrograde orbit (obliquity 156.8°). The remaining regular satellites are prograde. Neptune also has relatively few known satellites compared with the other major planets, and all but two of those, Triton and Nereid (NII), were unknown until the Voyager spacecraft observations (Smith et al., 1989) and subsequent ground-based discoveries (Holman et al., 2004). Nereid's orbital eccentricity of 0.75 brings it no closer than about 1.4×10^6 km from Neptune's center because of its extremely large semimajor axis of 5.51×10^6 km ($222.6R_N$) – well outside Triton's distance.

The massive satellite Triton is blamed for most of the features of this unusual system. The retrograde orbit means that Triton was almost certainly captured intact from heliocentric orbit. It has been proposed that Triton was captured in a three-body exchange where Pluto was expelled from orbit around Neptune (Farinella et al., 1980), that it was captured by gas drag (Mckinnon, 1984; McKinnon and Leith, 1995), or that it collided with a satellite already in orbit about Neptune with a mass a few percent of Triton's mass (Goldreich et al., 1989). The first hypothesis suffers from the likely multiple subsequent close encounters of the ejected Pluto with Neptune and essentially no means of leaving Pluto in the currently stable 3:2 MMR with Neptune, whereas the second has low probability. A collision would have been sufficient to capture Triton into a very eccentric orbit extending a significant fraction toward the Hill sphere boundary at $r_H \approx (M_N/3M_\odot)^{1/3} a_N \approx 4.5 \times 10^3 R_N$ (a_N = heliocentric distance) while not destroying it.

An alternative method of capture of Triton involving the disruption of a binary KBO with characteristics similar to the Pluto–Charon binary pair has been proposed (Agnor and Hamilton, 2004, 2006). Binary systems among the small bodies in the solar system appear to be ubiquitous with ~16% of near-Earth asteroids (Margot et al., 2002), ~2% of main belt asteroids (Merline et al., 2002), and ~11% of KBOs (Stephens and Noll, 2006) having companions. The fraction of small bodies in binary pairs is probably larger than these such that binary–planet encounters are likely. When a binary pair comes sufficiently close to a planet such that the Hill sphere of the pair in the planetary gravitational field is smaller than their separation, they become tidally unbound. Depending on the phase of the relative motion in the binary and the inclination of the orbital plane with that of the encounter trajectory, one of the planetesimals can be left in bound orbit about the planet while the other escapes. A simulation with binary pair masses $m_1 = m_T$ and $m_2 = 0.1m_T$ separated by $20R_1$ in circular orbit, where m_T is the mass of Triton and R_1 is the radius of m_1, closest approach distance $8R_N$ well inside the tidal disruption radius $r_{td} = 26R_N$ (R_N = Neptune radius), resulted in 95% of the binaries being disrupted for $0 < v_\infty < 2.0$ km s^{-1}, ~50% of the encounters resulting in capture of either m_1 or m_2 for $v_\infty < 0.35$ km s^{-1}, and ~50% of the encounters resulting in capture of m_2 for $0.35 \lesssim v_\infty \lesssim 1.55$ km s^{-1}. For the latter velocity interval, capture of m_1 is still possible but rare.

Although either collisional capture (Goldreich et al., 1989) or binary exchange capture (Agnor and Hamilton, 2004, 2006)

Author's personal copy

can yield reasonable orbits that can damp to the observed retrograde circular orbit observed today, Agnor and Hamilton show that with at least 11% of KBOs being binary systems, capture by collision has negligibly small probability compared with the binary exchange capture. If the original satellite system of Neptune were similar to that of Uranus, collision of Triton with any satellites this large would have destroyed it, making the currently inclined orbit of a recollected Triton impossible (Ćuk and Gladman, 2005). So, it is almost certain that Triton was captured gently without such a risk, although a collisional capture remains an improbable possibility. Nogueira et al. (2011) verified the arguments for the capture of Triton by the disruption of a binary KBO, and Vokrouhlicky et al. (2008) argued that the capture must have occurred early in solar system history in the context of the Nice model of this history (Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005).

The properties of a prograde satellite system that likely existed around Neptune prior to Triton's capture are of course unknown. An original satellite system like that of Uranus extending from about $5R_N$ to beyond $20R_N$ would have been perturbed into crossing orbits, and the satellite system would be thereby destroyed. During the times when Triton's periapse was inside $20R_N$, energy would be removed from Triton's orbit impulsively at each passage through the disk of material leading to circularization times of 10^4–10^5 years while Triton cleared most of the material from the disk outside of about $5R_N$ (Ćuk and Gladman, 2005).

If interactions with such a disk were not sufficient to circularize Triton's orbit, subsequent tidal dissipation would suffice. With $k_T(=k_2)=0.1$ and $Q_T=100$, Goldreich et al. (1989) found that an initial Triton orbit with semimajor axis $a_T \approx 10^3 R_N$ would damp to nearly its current circular orbit with $a_T \sim 14.3R_N$ from tidal dissipation in Triton in about $4–5\times10^8$ years – comfortably shorter than the age of the solar system. However, the large amount of dissipation probably melted a significant fraction of the satellite, which drastically hastens tidal evolution (Mckinnon et al., 1995; Shock and McKinnon, 1993). The Love number k_2 would then be $O(1)$ (closer to that of a fluid sphere) instead of $O(0.1)$ and $Q_T \sim O(10)$ instead of $O(100)$, leading to a factor of $\sim$100 shorter eccentricity decay timescales.

We can estimate the time required to damp Triton's eccentricity (e_T) with the following procedure. During the damping of Triton's eccentricity, it probably existed in an asymptotic rotation state defined by the zero of the averaged tidal torque given, for example, by eqn [B.17] in **Appendix B**. We determine de/dt from eqn [B.27], where we assume $\dot{\psi}_s$ is the asymptotic value given as a function of e by the zero of the third component of eqn [B.17]. If we also assume that the periapse position is maintained at Triton's current distance, then the semimajor axis can be represented as $a_0/(1-e)$ where a_0 is the current value. With these assumptions, the time to decrease Triton's eccentricity from 0.9 to 0.01 is $3.3\times10^5\ Q_T/k_T$ years. The tides raised on Neptune by Triton will have a negligible effect (Banfield and Murray, 1992; Goldreich and Soter, 1966). With the Goldreich et al. choices of k_T and Q_T, the time to damp Triton's eccentricity is 3.3×10^8 years, which is the several hundred million years damping time found by Goldreich et al. With the more likely value of k_T/Q_T, this time is only 3.3×10^6 years.

But tidal dissipation may have been less efficient in the circularization of Triton's original eccentric orbit than inferred by Goldreich et al. because solar perturbations cause periodic increases in periapse distances (Benner and McKinnon, 1995). The Goldreich et al. analysis ignored these Kozai oscillations in eccentricity (see **Section 10.17.9**), which would have occurred if the original maximum orbital inclination was comparable to the current one. These 1000-year variations in e_T cause the mean periapse distance to be much larger than that obtained by Goldreich et al. with the consequence that the tidal damping time, with the same values of Q_T and k_T, and expression for dissipation used by Goldreich et al. would yield a timescale of 3.5 Gy for the eccentricity reduction (Ćuk and Gladman, 2005). Reducing this number to 35 My with the more likely values of Q_T and k_T in the preceding text still makes the tides adequate.

Consistent with the capture scenario and subsequent melting, Triton was found to have a young and active surface. The observed, currently active plumes are almost certainly solar-driven (Smith et al., 1989), but the geologically young volcanic plains and cantaloupe terrain are clearly endogenic, where such surface modification may be ongoing (Shenk and Jackson, 1993; Stern and McKinnon, 2000). Triton's density is about 2 g cm^{-3}, which indicates some ice content. Triton's original constituents were obtained directly from the solar nebula at its location of formation, but there may be significant contamination with debris from the original satellite system. The volatile content was probably reduced during the rather extreme tidal heating during the eccentricity damping, although some N_2 was retained from the early dense atmosphere and most of the water is still there (Lunine and Nolan, 1992). Triton and Pluto are similar in size and density, but Pluto may have lost a lot of volatiles from the impact that created Charon (see **Section 10.17.8**), and one cannot infer that this loss is the same as that from Triton. The current density does not constrain the place in the solar nebula where Triton may have been formed. But an initial composition similar to that of Halley's comet is consistent with observations of the surface ices, where the severe reduction of CO relative to N_2 and CO_2 can be understood by hydrothermal processing resulting from the tidal heating (Shock and McKinnon, 1993).

While Triton's orbit was extended, any regular satellites inside about $5R_N$ would also have been disrupted and destroyed from the periodic dips of the periapse inside $5–8R_N$, depending on the initial semimajor axis. There are likely to have been several satellites inside the periapse distance of Triton in comparison with the regular systems of the other major planets. Orbit eccentricities would have been limited to maximum values near 0.3, since the rate of decrease of the eccentricity by tidal dissipation within the satellite exceeds Triton's ability to increase it for larger values (Banfield and Murray, 1992). Two satellites with semimajor axes of $3R_N$ and $5R_N$ would have overlapping orbits that would persist in overlapping for times at least of the order of the eccentricity damping time of 10^5–10^7 years. A collision between any of Neptune's current inner satellites with its neighbor would lead to their mutual destruction. This follows from the relation (Stevenson et al., 1986)

$$\frac{1}{2}m_i v_i^2 \sim m_s S + \frac{3}{5}\frac{Gm_s^2}{\gamma R_s} \qquad [14]$$

Author's personal copy

where m_i and m_s are the masses of the impactor and satellite, respectively; v_i is the relative velocity at impact; R_s is the satellite radius; $S \sim 10^6$ erg g^{-1} is the material binding energy; and $\gamma \sim 0.1$ is a factor introduced to account for the inefficiency in converting the impact kinetic energy into kinetic energy of the fragments. The impactor kinetic energy must exceed the energy stored in material strength plus the self-gravitational energy by a sufficient amount to break up the body. From eqn [14] and eccentricities of 0.3, Naiad (NIII) could destroy all of the satellites except Proteus (NVIII), and any of the satellites NIV to NVII could destroy Proteus. This implies that the current satellite system could not have existed prior to Triton's orbit circularization (Banfield and Murray, 1992). The debris from the first generation of satellites would settle into the equatorial plane in circular orbits and recollect into a second generation of inner satellites with nearly circular orbits and zero inclinations, where all memory of Triton's perturbations would be thereby lost. Had any of these inner satellites survived destruction, their orbits would have been left inclined to Neptune's equatorial plane. The lack of such inclined orbits among the inner satellites must mean that there are no survivors among those early inner satellites.

The inner satellites observed today are not those formed after the circularization of Triton's orbit. The flux of comets through the system would have destroyed all of the current inner satellites except for Proteus (1989N1) over the lifetime of the solar system (Colwell and Esposito, 1992; Smith et al., 1989). The debris from such a destructive collision would settle to the equatorial plane where the satellites can reaccrete. The rings of small particles are likely to be remnants of these collisional breakups. Any distant irregular satellites of Neptune (Holman et al., 2004) would also have their orbits perturbed drastically by Triton if they were captured before or during Triton's capture and orbital decay and their numbers depleted (Ćuk and Gladman, 2005).

10.17.8 Pluto System

Pluto and its large satellite Charon orbit about their center of mass with a period of 6.3872273(3) days at a separation of 19,573(2) km with zero eccentricity (Buie et al., 2012a). (The numbers in parentheses denote the 1σ uncertainty in the last digits.) The rotation of both bodies is synchronous with the orbital motion, which makes this system the only one in the solar system known to have reached the end point of tidal evolution. Like the Earth–Moon system, the Pluto system has a high specific angular momentum that favors the formation of Charon by a large oblique impact (Mckinnon, 1984, 1989; Stern et al., 1997). Canup (2005) had performed SPH simulations of impacts on the Pluto precursor with impactor mass a significant fraction of the target mass with a variety of compositions and differentiation states. The most probable outcome of a simulation that is successful in placing a sufficient amount of mass in orbit outside the Roche distance is for Charon to be launched as an almost intact satellite in a highly eccentric orbit. In these examples, the impactor is undifferentiated and Charon is composed almost entirely of material from the impactor, which results from its being sheared off in the oblique impact. Some of the discussion of the giant impact creating the Moon in Section 10.17.2 is applicable to Pluto–Charon, but the impact associated with the latter involves much less energy. There is little heating of Charon, little water ice is vaporized, but there is probably sufficient energy deposited in Pluto at impact coupled with radiogenic heat to cause its differentiation (Canup, 2005). An impact thus appears a quite natural mechanism to produce Charon, assuming that large collisions were probable in the early Kuiper Belt. However, the impact does not, as it turns out, provide a straightforward explanation for Pluto's tiny satellites discovered in recent years. The objects present a current puzzle, which we describe in some detail later in the text. The gravitational collapse scenario pointed out earlier in the text for widely spaced, massive Kuiper Belt binaries probably is not applicable to the Pluto system because of the difficulty of assembling a sufficient amount of mass into a gravitationally unstable swarm (Nesvorny et al., 2010).

Pluto has four known additional satellites in nearly circular orbits that are nearly coplanar with the orbit of Charon. The first two discovered (Weaver et al., 2006) are named Nix and Hydra. Best-fit Kepler orbits by Buie et al. (2012b) yield periods, semimajor axes, and eccentricities of 24.85494(6) d, 38.20163(7) d; 48 831(25) km, 64 736(11) km; and 0.00232(19), 0.00658(15), respectively. Pluto's fourth satellite, S/2011 (134 340) 1, also known as P4, was discovered on 28 June 2011 (Showalter et al., 2011). The satellite was found in archival images, where the long time span allowed the determination of the following Kepler orbit parameters (Buie et al., 2012b): period = 32.1866(22) d, semimajor axis = 57 890(120) km, and eccentricity = 0.0045(11). The fifth satellite, S/2012 (134 340) 1, also known as P5, was discovered on 14 separate sets of images from the Wide Field Camera on the Hubble Space Telescope on June 26, 27, and 29 and July 7 and 9 (Showalter et al., 2012). P4 is about 1/10 as bright as Nix, and P5 is about 1/2 as bright as P4. The orbital period of P5 is 20.2 ± 0.1 days with orbital semimajor axis near 42 000 km. All of the orbital planes are within $\lesssim 0.15°$ of Charon's orbital plane. With the orbital period of Charon at 6.3872273 days, the ratio of the orbital periods of each of the satellites to that of Charon is 3.16 (P5), 3.89 (Nix), 5.03 (P4), and 5.98 (Hydra) reasonably close to 3:1, 4:1, 5:1, and 6:1, respectively. The parameters of the Pluto system are summarized in Table 1.

Although the orbital periods of Charon, Nix, and Hydra are nearly in the ratios 1:4:6, no resonance variables are likely to be librating at the present time (Lee and Peale, 2006). The small satellites orbit about the center of mass of Pluto and Charon, and that position from the orbit-fitting procedure gives a Charon–Pluto mass ratio of 0.1168(55). Pluto's radius from the mutual event light curves is 1164(22.9) km (Young and Binzel, 1994) and Charon's radius from a stellar occultation is 603.6(1.4) km (Sicardy et al., 2006). With the masses determined by Buie et al., these radii yield densities of 2.03(0.06) and 1.65(0.08) g cm^{-3} for Pluto and Charon, respectively. The physical parameters of Pluto and Charon are consistent with the results of the giant impact origin of the system (Canup, 2005). Lee and Peale (2006) showed that the orbital motions of Nix and Hydra are markedly non-Keplerian because of the large Charon–Pluto mass ratio. As the masses of the new satellites are increased within the range allowed by the uncertainty in the albedos, the close proximity of the 3:2 MMR between Nix and Hydra causes increasing fluctuations in the

Author's personal copy

Table 1 Pluto system

	Mass (g)	*Radius (km)*	*Rotation period (days)*	*Orbital period (days)*	a *(km)*	*e*
Pluto	1.302×10^{25}	1151	6.38723	6.38723	–	–
Charon	$0.1165 m_P$	603.6	6.38723	6.38723	$19573 = 17.0 R_P$	0
P5	?	5–14[a]	?	20.2	$42000 = 36.5 R_P$	~0
Nix	$O(3 \times 10^{-5}) m_P$[b]	23–68[a]	?	24.855	$48831 = 42.4 R_P$	~0
P4	?	7–20[a]	?	32.2	$57890 = 50.3 R_P$	~0
Hydra	$O(6 \times 10^{-5}) m_P$[b]	30–84[a]	?	38.202	$64736 = 56.2 R_P$	~0

[a]Assumed albedo range (0.04–0.35).

[b]The stability of P4 places upper limits on the masses of Nix and Hydra that are about an order of magnitude less than these estimates from astrometry (Youdin et al. 2012).

eccentricities. Although the maximum masses allowed within these constraints allow possible libration of the 3:2 resonance variable involving the periapse longitude of Hydra, these maximum masses are probably already ruled out because they would also excite Charon's eccentricity beyond the current observational uncertainties (see Buie et al., 2012a). A dynamic fit to a sufficiently dense set of observational data will yield constraints on the masses of Nix and Hydra (Lee and Peale, 2006). The fact that Nix, Hydra, P4, and P5 are in orbits coplanar with Charon's orbit is consistent with their being accreted in a dissipative disk of much smaller particles.

One way to create such a disk is from the debris left over from the giant impact that trapped Charon into its orbit about Pluto. However, Nix and Hydra reside at ~42 and 56 Pluto radii (R_P) from the Pluto–Charon center of mass, and debris from the impact will not extend out to these distances (Canup, 2004a, 2005, 2011). A disk extending beyond the location of the 6:1 mean motion commensurability with the initially close orbit of Charon is likely formed from partially differentiated impactor and target although usually not from impacts of homogeneous bodies (Canup, 2011). If the small satellites accreted from this disk of material, an appealing hypothesis for transporting Nix and Hydra, and presumably P4 and P5 discovered later, was investigated by Ward and Canup (2006).

Charon's orbit must expand from its initial size in the process of reaching the current state of completed tidal evolution of dual synchronous rotation (Cheng et al., 2014). One can infer from the closeness of Nix and Hydra to the 4:1 and 6:1 mean motion commensurabilities with Charon that the satellites may have occupied a MMR at each commensurability. Being so trapped would mean they could be pushed out ahead of Charon as Charon's orbit tidally expanded. But the resonance must be chosen with care for this to happen.

There are a number of resonances that can be occupied at the 4:1 and 6:1 mean motion commensurabilities characterized by the angular arguments of terms in the expansion of the perturbing potential

$$\phi_{ml} = (m+1)\lambda - \lambda_C - (m-l)\varpi - l\varpi_C \qquad [15]$$

where λ and λ_C are the mean longitudes of the satellite and Charon and ϖ and ϖ_C are the longitudes of periapse. The angle ϕ_{ml} is written for mth-order resonances of the form $m+1$:1 with $m=3$ for Nix and $m=5$ for Hydra. The terms in the expansion of the perturbing potential are represented by $\Phi_{ml}(a, a_C, e, e_C) \cos \phi_{ml}$, where terms involving orbital inclinations are ignored and where $\Phi_{ml} \sim e^{m-l} e_C^l$. The only resonances at the $m+1$: 1 mean motion commensurability that can preserve the satellites during Charon's orbital expansion are for $m=l$ involving only the eccentricity of Charon's orbit in the coefficient Φ_{ml}. This follows from the fact that all those resonances involving e and hence ϖ lead to steady growth in e as the orbits expand with eventual instability. The resonance involving only the eccentricity and longitude of periapse of Charon is called a corotation resonance, since the satellite orbits at the same angular velocity as the resonant potential in the expansion of the perturbing potential due to Charon, the so-called pattern speed. In resonances of this type, there are forced eccentricities for Nix and Hydra whose magnitudes vanish with e_C, but these do not lead to crossing orbits or other unstable configurations. The stability of a corotation resonance vanishes as $e_C \to 0$.

Inspection of eqn [B.22] in **Appendix B** and the local text show that tides raised on Pluto can generally either increase or decrease the orbital eccentricity e depending on the values of both rotation rate $\dot{\psi}_P$ and e but that tides on Charon always decrease the eccentricity, since Charon's angular velocity $\dot{\psi}_C$ will always be near the mean orbital motion n. For relatively larger values of $\dot{\psi}_P$, there is a competition between tides on Pluto trying to increase e and those on Charon trying to decrease e. Cheng (2011) and Cheng et al. (2014) had shown that with a judicious choice of the ratio of the rate of dissipation in Charon to that in Pluto, it is possible to keep Charon's e nearly constant near its initial value throughout the tidal evolution from its close proximity to Pluto to its current separation in dual synchronous rotation. Thereby, the condition for maintaining Nix and Hydra in the 4:1 and 6:1 corotation resonances with Charon during the expansion of the latter's orbit can be satisfied, but that is not the end of the story.

There are two difficulties with the Ward and Canup hypothesis. First, in an investigation to test the migration hypothesis for Nix and Hydra, where the two small satellites are initially placed outside the 4:1 and 6:1 commensurabilities and Charon's orbit is tidally expanded, there were no captures into either corotation resonance by itself in thousands of trials (Cheng, 2011; Cheng, Lee and Peale (2014a)). Next, by carefully placing one of the satellites into a corotation resonance, it was possible to transport it to its current position. But the parameter space that permitted the migration of one satellite in a corotation resonance does not overlap with that which permitted the other to be so transported (Lithwick and Wu, 2008). It is not possible to transport both Nix and Hydra within corotation resonances. A possible alternative was found by Cheng (2011) where in many cases, a test particle was trapped into more than one resonance at the 5:1, 6:1, 7:1, and 8:1 commensurabilities during thousands of trials with the

Author's personal copy

gravitational coefficient $J_2 = 0$ for Pluto. Interestingly, when caught in the multiple resonances at a given commensurability, the evolution proceeded to the final state without e growing to instability. All the resonance variables stopped librating at the end of the evolution, but the test particle was often left with a significant eccentricity. There was one example where all of the resonance variables librated about 0° with Charon's and the test particle's orbit aligned that proceeded to the current configuration for Hydra near the 6:1 commensurability with both eccentricities damping to zero. There was no example of stable transport in multiresonances at the 4:1 commensurability.

Unfortunately, this mechanism for transporting Nix and Hydra in multiple MMRs has a fatal flaw. A condition for occupying several resonances at a mean motion commensurability is that the orbits be aligned or antialigned with $\langle \dot{\varpi}_C \rangle = \langle \dot{\varpi} \rangle$. This was possible in the examples because $J_2 = 0$ was assumed. But the rapidly rotating Pluto, softened by the impact, will have a large hydrostatic J_2 (initially >0.1) leading to $\dot{\varpi}_C \gg \dot{\varpi}$. The conditions for multiresonance transport of the small satellites will not prevail.

Lithwick and Wu (private communication) successfully transported Nix and Hydra in three-body resonances (with $J_2 = 0$) that were so delicate that their assembly is very unlikely by natural processes and the presence of P4 and P5 and the finite value of J_2 are major obstacles. Cheng, Peale and Lee (2014b) integrated more than a million sets of initial conditions distributed over eccentricities, semimajor axes, and true anomalies with Nix at a range of initial positions relative to Charon but always starting outside the 4:1 mean motion commensurability. Hydra was always just outside the 3:2 mean motion commensurability with Nix or within one of the resonances at that commensurability. Finite but minimal masses of Nix and Hydra were assumed with Charon's albedo. There were no survivors for either $J_2 = 0$ or $J_2 =$ the hydrostatic value. The frustrating failure of the resonance transport hypothesis to place the small satellites, formed simultaneously with Charon at the time of the impact, at their current positions relative to Charon has inspired alternative suggestions.

Establishing a disk of material after the Pluto–Charon evolution is complete has been suggested by Lithwick and Wu (2008). Such a disk, if sufficiently dissipative, will settle to the plane of the Pluto–Charon orbit. To produce satellites orbiting in the same direction as Charon, the disk would need to have a negative specific angular momentum in the frame rotating with Pluto's orbital motion, since Pluto–Charon have a retrograde orbit. Existing studies of the stability of the Pluto–Charon phase space establish an unstable region spanning Charon's orbit. Stern et al. (1994) found $0.47a_C \lesssim a \lesssim 2.0a_C$ and $0.65a_C \lesssim a \lesssim 2.15a_C$ to be unstable for prograde and retrograde orbits, respectively. Pires dos Santos et al. (2011) found that Nix and Hydra decrease the stability of the region beyond $2a_C$ and that the only stable region between Nix and Hydra is near the 5:1 mean motion commensurability with Charon, where we find P4. The unstable region includes the 2:1 and 3:1 mean motion commensurabilities, and we find P5 just outside the 3:1. It would appear that the small satellites may have selected only the nearest stable regions in the Pluto–Charon phase space. N-body calculations to test this hypothesis have not yet been performed.

A conceivable way of establishing the disk after the evolution of the Pluto–Charon system is a collision of two large objects within Pluto's sphere of influence. At best, this is an event of low probability, but Pires dos Santos et al. (2012) maximized that probability by assuming a heliocentric disk density 1000 times that of the current Kuiper Belt, which would prevail before the orbits of the outer planets were expanded. The probability of collision within the Pluto–Charon sphere of influence is further enhanced by one of the objects in the collision being already in a temporarily captured orbit about Pluto–Charon. With the maximum lifetime of such captured orbits being only $O(100)$ years, Pires de Santos et al. found the time between collisions of objects with sufficient mass to account for Nix and Hydra to be so much longer than this that they abandoned this scenario as a possible origin of a dissipative disk that could have produced the small satellites.

Although it is generally agreed that Charon is the result of a glancing impact on the Pluto precursor with subsequent tidal evolution to the current dual synchronous state, the origin of the small satellites remains uncertain. It seems clear that these satellites formed from a dissipative disk of material, but collisional creation of a disk with sufficient mass that extends throughout the region of the small satellites remains improbable.

10.17.9 Irregular Satellites

The irregular satellites that orbit all of the major planets occupy large orbits with significant inclinations and eccentricities. Recent observations have increased the number of known irregular satellites to 96 (Gladman et al., 1998, 2000, 2001; Holman et al., 2004, http://ssd.jpl.nasa.gov/sat_elem.html, http://ksiezyce.republika.pl/table_en.html). They often extend to distances from their primaries that are close to the stability limits of about half of the planet's Hill sphere radius (Hamilton and Burns, 1991; Hamilton and Krivov, 1997). Similar orbital elements for members of clusters of the irregular satellites imply breakup events after the primary body was captured (e.g., Gladman et al., 2001). The only viable means of establishing these systems of large orbits with high eccentricities and inclinations is through capture from the large number of planetesimals orbiting the Sun early in solar system history (e.g., Burns, 1986). Capture can result from a loss of energy by the planetesimal while it is within the planet's Hill sphere. Such energy loss can be either from gas drag in an extended planetary atmosphere (Pollack et al., 1979), from a collision between two unbound planetesimals leaving one with sufficient energy loss to be bound, or from a collision with a smaller object already in stable orbit about the primary (Colombo and Franklin, 1971). The gas drag mechanism for at least the progenitor of the Himalia group of prograde irregular satellites orbiting Jupiter has been explored in detail by Ćuk and Burns (2004), where it is shown that such capture must occur near the end of a waning gas disk lifetime to avoid losing the satellites to the planet. Gas drag no doubt resulted in the capture of many planetesimals into stable orbits about the major planets, but only the very last ones have survived. The occupancy of orbital resonances by two of the retrograde satellites of Jupiter implies that at least a thin disk was necessarily in place at the time of the capture to allow a weak gas drag to bring the satellites to the resonances (Saha and Tremaine, 1993). An alternative mechanism of capture involves a rapid

Author's personal copy

growth of the primary, while a planetesimal is temporarily trapped in the Hill sphere, to expand the region of stable orbits (Heppenheimer and Porco, 1977; Saha and Tremaine, 1993). The binary exchange capture for the irregular satellites that is highly probable for the capture of Triton (Agnor and Hamilton, 2004, 2006) is probably not as effective for the capture of the distant irregular satellites, since an irregular satellite is less likely to have ever been close enough to the planet for the binary Hill sphere to have been smaller than the separation of the binary members (Vokrouhlicky et al., 2008).

Still another means of capture of the irregular satellites is a two-body encounter within the Hill sphere, where one of the two participants has sufficient energy loss to remain captured. An interesting variation of nondissipative capture has been proposed by Nesvorny et al. (2007) within the Nice model of early solar system evolution (Gomes et al., 2005; Morbidelli et al., 2005; Tsiganis et al., 2005). In the Nice model of solar system evolution, all the major planets were formed between 5 and 20 AU with a sea of planetesimals with total mass between 20 and 35 Earth masses between 20 and 35 AU from the Sun. The initial separation of Jupiter and Saturn is inside the 2:1 MMR. Jupiter migrates inward and Saturn outward from the scattering of the remaining planetesimals in their vicinity such that near 700 million years after formation, Jupiter and Saturn cross the 2:1 MMR in the wrong direction for capture, but Saturn's eccentricity is excited that destabilizes Neptune and Uranus into crossing orbits between themselves and with Saturn. There are repeated planetary encounters, Neptune and/or Uranus is scattered into the sea of planetesimals leading to a vastly increased flux of planetesimals into the inner solar system, and the asteroid belt is disturbed. The flux of objects into the inner solar system results in the late heavy bombardment (LHB) therein. After many close encounters, the orbits of Neptune and Uranus decouple and the eccentricities and inclinations are damped by dynamic friction as the orbits migrate outward by scattering the abundant planetesimals. The current architecture of the outer solar system is thereby produced.

But the many close encounters among Neptune, Uranus, and Saturn within the population of planetesimals led to the capture of planetesimals in stable orbits around these planets. Planetary encounters with Jupiter are rare, so Jupiter's irregular satellites must be captured by some other mechanism, like the gas drag and collisions discussed earlier in the text. In a typical computer run that produces the current architecture, Neptune, Uranus, and Saturn capture many more irregular satellites than are observed with distributions in orbital elements that match the observed distributions remarkably well. So, if the Nice model or some modification of the same that results in many planetary encounters in the presence of planetesimals is close to reality, the Nesvorny et al. scheme must be important in the capture of irregular satellites. Of course, this mechanism of capture only occurs more than 700 million years after the initial formation of the planets, which may mean that irregular satellites captured before this time by, say, the collision or gas drag mechanisms are replaced by the new population – except Jupiter's irregular satellites may be those captured at earlier times. It is interesting that the systems of irregular satellites are so similar if scaled to the Hill sphere of each planet (Jewitt and Sheppard, 2005), when a different mechanism may be important for Jupiter than for the other three giant planets.

The capture of satellites should lead to a random distribution of orbital inclinations relative to the planet's orbital plane (~ecliptic), but there are few irregulars known with inclinations between 47° and 141°. Carruba et al. (2002) gave a lucid explanation of the scarcity of satellites with orbital inclinations within this range. Solar perturbations dominate the variations in the orbits of these satellites, and in an analytic secular theory of the restricted three-body problem involving perturbations of the satellite orbit by the lowest-order solar tidal term, the motion is averaged over the mean motions of both the planet and the satellites. In this approximation, the component of the satellite's angular momentum that is perpendicular to the planetary orbital plane, $\sqrt{Gm_p a_s(1-e^2)}\cos i$, is conserved, so that the variations in e and i are linked. When the satellite orbital plane is closest to the planetary orbital plane, the eccentricity is maximal. The variations in e and i are thus of opposite algebraic sign for prograde orbits ($i<90°$), but for retrograde orbits ($i>90°$), increasing i brings the orbit closer to the planetary orbital plane so variations in e and i have the same algebraic sign. For $i<i_{crit}=39.2°$ or $i>i_{crit}=140.8°$, the amplitude of the variations in e and i are limited, but for inclinations between the values of i_{crit}, large variations in e can drive the periapse into the region of the regular satellites or the apoapse close to R_H when the argument of periapse ω is circulating. Either consequence can lead to the elimination of the satellite. The only way a high inclination satellite is likely to survive is if ω is librating about 90° or 270° (Kozai resonance) where variations in the eccentricity are more modest. By estimating the fraction of the available phase space for a captured satellite to be in Kozai resonance, Carruba et al. predicted that perhaps 10% of captured satellites should be in stable Kozai resonances at high inclination. The irregular satellites S/2003 J10, Kiviuq (S/2000S5), and Ijiraq (S/2000S6) are apparently currently librating in Kozai resonances. But at least Kiviuq is close to the chaotic zone separating libration from circulation, so its libration may be only temporary (Carruba et al., 2003, 2004).

Numerical integrations of the complete equations of motion of irregular satellites, including the perturbations of the other major planets, show that highly inclined orbits outside the Kozai resonance are more unstable for orbital planes further from the planetary orbital plane between the values of i_{crit} (Carruba et al., 2002). As the extremes of the variable inclinations approach the limits of the zone of exclusion, the satellites can survive for longer periods of time, but the unstable region seems to approach the observed limits asymptotically as the integrations progress. The nonrandom distributions of orbital inclinations and values of the other observed orbital parameters of the irregular satellites are therefore completely consistent with the satellites having been captured into random orbits from the solar nebula population of planetesimals by any of the means discussed in the preceding text.

Appendix A Accretion Disks

The regular satellites of the major planets are in nearly circular orbits in the planes of the planets' equators. This means these satellites had to have formed in dissipative disks of gas and particles that settle to the midplane of the disk before the accretion of the satellites occurs. Given that these satellites

Author's personal copy

form in such a disk during the formation of the planet, it is important to understand the nature of the accretion in such disks. Our ultimate goal is the determination of the timescales for accretion of such satellites as a function of the disk and particle parameters, for these timescales constrain the current observable properties of the regular satellites. Once formed, the satellites interact with the disks primarily through gas drag when they are small and through density waves generated in the adjacent regions of the disk at mean motion commensurabilities with the satellite when they are large. If the midplane pressure of the disk decreases with distance from the primary, gas drag causes a satellite to spiral toward the planet. Generally the same type of disk causes a similar spiral toward the planet through interaction of the satellite with density waves generated in the disk.

A.1 Accretion of Solids

First, we consider the idealized case of an isolated spherical body m_1 of radius R_1 in a swarm of particles with a distribution $n(m_2, v)\mathrm{d}m_2\,\mathrm{d}v$ representing the number density of particles of mass within $\mathrm{d}m_2$ of m_2 and speeds within $\mathrm{d}v$ of v. If the relative velocities are isotropic, $n(m_2, v)\sin\theta\, d\theta\,\mathrm{d}\phi\,\mathrm{d}m_2\,\mathrm{d}v/4\pi$ represents the number density of the particles with velocities aimed within $\mathrm{d}\theta\,\mathrm{d}\phi$ of spherical polar coordinates θ, ϕ in a coordinate system where the mean particle velocity is zero. One of these particles of radius R_2 and speed v far from m_1 will collide with m_1 during an encounter if the impact parameter is $<R_g$, where

$$R_g^2 = (R_1+R_2)^2\left(1+\frac{v_e^2}{v^2}\right) \qquad [A.1]$$

with $v_e=\sqrt{2G(m_1+m_2)/(R_1+R_2)}$, the escape velocity of m_2 from m_1 when they are in contact. Equation [A.1] is found by conserving energy and angular momentum when m_2 just grazes m_1 during the encounter. The factor $(1+v_e^2/v^2)=F_g$ is called the gravitational enhancement factor for the collisional cross section of m_1 for particles of radius R_2 and speed v relative to m_1. The mass flux of particles with mass within $\mathrm{d}m_2$ of m_2 and velocity within $\mathrm{d}v$ of v directed within $\mathrm{d}\theta\,\mathrm{d}\phi$ of θ, ϕ onto m_1 is $\pi R_g^2 n(m_2,v)m_2 v\,\mathrm{d}m_2\,\mathrm{d}v\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi/4\pi$ such that

$$\dot{m}_1=\int_{m_2}\int_v \pi n(m_2,v)m_2 v(R_1+R_2)^2\left(1+\frac{2G(m_1+m_2)}{(R_1+R_2)v^2}\right)\mathrm{d}m_2\,\mathrm{d}v \qquad [A.2]$$

is the rate of growth of m_1, if we assume each m_2 that collides is accreted. In eqn [A.2], we have assumed the velocities are isotropic relative to m_1 and integrated over the solid angle and use $R_2=(3m_2/4\pi\rho_p)^{1/3}$, with ρ_p being the mass density of the individual particles. Because $n(m_2, v)$ is not very well constrained, eqn [A.2] is often replaced by the much simpler form

$$\dot{m}_1\approx\frac{\sigma_s}{2H_s}v\pi(R_1+R_2)^2F_g \qquad [A.3]$$

where R_2, v_e, and v are now suitably averaged values. We have replaced the spatial mass density of the particles by $\sigma_s/2H_s$, where σ_s is the surface mass density of solids and $2H_s$ is the disk thickness. H_s depends on the velocity dispersion v and can be estimated by

$$0.5v_\perp^2=\int_0^{H_s}(GMz/r^3)\,\mathrm{d}z=0.5\Omega_K^2H_s^2 \qquad [A.4]$$

where $v_\perp=Kv$ $(K=O(1))$ is the component of the averaged velocity that is perpendicular to the plane of the disk, G is the gravitational constant, M is the mass of the primary, r is the distance from the primary, z is the vertical distance from the midplane, and $\Omega_K=\sqrt{GM/r^3}$ is the Kepler orbital angular velocity of a test particle at distance r from the primary. GMz/r^3 is just the local vertical component of the gravitational acceleration due to the primary. Hence, $H_s=v_\perp/\Omega_K=Kv/\Omega_K$. This value of H_s is generally smaller than the gas pressure scale height c_s/Ω_K (derived later in the text), with c_s being the speed of sound, because of settling of particles toward the midplane.

From eqn [A.3], we can write

$$\dot{R}_1=\frac{\sigma_s\Omega_K}{8K\rho_p}\left(1+\frac{R_2}{R_1}\right)^2F_g \qquad [A.5]$$

where ρ_p is the mass density of the particles as distinguished from ρ_s, the spatial density of solids in the disk. If we assume $R_2\ll R_1$ and ignore factors of order unity, we can write the timescale for the growth of a body as (Canup and Ward, 2002)

$$\tau_A=\frac{R_1}{\dot{R}_1}\approx\frac{\rho_pR_1}{\Omega_K\sigma_sF_g}\sim 50\,\text{years}\left(\frac{R_1}{2500\,\text{km}}\right)\left(\frac{\rho_p}{2\,\text{g cm}^{-3}}\right)\left(\frac{10}{F_g}\right)\left(\frac{3\times10^3\,\text{g cm}^{-2}}{\sigma_s}\right)\left(\frac{r}{15R_J}\right)^{3/2} \qquad [A.6]$$

where the numerical form is for the accretion of Ganymede in a minimum mass disk of particles. The surface mass density σ_s is estimated by spreading Ganymede's mass over an annulus extending from the midpoint between Ganymede's and Europa's orbits to the midpoint between Ganymede's and Callisto's orbit to yield $\sigma_s=3300\,\text{g cm}^{-2}$. With $\rho_p\approx2\,\text{g cm}^{-3}$ and an orbital radius of $15R_J$, we arrive at $\tau_A\sim50$ years when $F_g=10$. This is a lower bound on the timescale for the growth of Ganymede, since σ_s is reduced as the growth proceeds, but it is indicative of the very short timescales for accretion in minimum mass disks.

Safronov (1969), by transforming from relative to random velocities and assuming the vertical and horizontal velocity dispersions are equal, found $K=\pi/4\sqrt{2}$. Safronov also assumed that the velocity dispersion of the accreting particles remained comparable to the escape velocity of the largest embryos as R_1 increased. In this case, $F_g=O(3)$, and R_1 increases linearly with time if the surface mass density σ_s is maintained in the vicinity of m_1. This linear increase in the radius of an accreting body is called orderly growth. $F_g\approx3$ is not that much smaller than our choice of $F_g=10$ in eqn [A.6]. Runaway growth can make these accretion timescales even shorter.

The assumption by Safronov that the velocity dispersion of the smaller particles was always near the escape velocity of the largest bodies proved to be incorrect in numerical simulations by Wetherill and Stewart (1989). As long as most of the mass of the disk was in the smaller bodies, the velocity dispersion was nearly the escape velocity of the smaller particles. Increases in the velocities of small particles by close encounters with the

Author's personal copy

larger bodies were damped by interactions with members of the swarm. F_g is thereby increased by orders of magnitude from that assumed in orderly growth and led to what is called runaway growth. For v fixed near the escape velocity of the smaller bodies, the rate of growth in eqn [A.5] becomes proportional to v_e^2, which, for constant density, increases as R_1^2. The change in scaling results in the largest planetesimal growing much faster than any other, and it 'runs away.' For the two largest planetesimals of the swarm of radii $R_1 > R_2$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{R_1}{R_2}\right) = \frac{R_1}{R_2}\left(\frac{\dot{R}_1}{R_1} - \frac{\dot{R}_2}{R_2}\right) \approx \frac{R_1}{R_2}(R_1 - R_2) \qquad [A.7]$$

shows that if initially $R_1 > R_2$, the ratio R_1/R_2 grows as long as the mass available for accretion has not decreased substantially.

The timescale for accretion in eqn [A.6] is reduced by the larger value of F_g. Both orderly growth and runaway growth timescales depend on the maintenance of the surface mass density of the disk near the accreting satellite and that most of the disk mass is in the small particles. These timescales increase as the nearby material is exhausted and must be eventually governed by how rapidly particles can be brought into the so-called feeding zone of the growing satellite. Still, numerical simulations of lunar accretion in a particle disk verify the rapid accretion timescales even while accounting for the growth of the average particle size during the process (Ida et al., 1997; Kokubo et al., 2000). Mosqueira and Estrada (2003a,b) determined accretion timescales governed by how rapidly gas drag can bring larger embryos to the feeding zone of the satellite.

Although some stages of the accretion process in disks remain unclear, once small solid objects have formed or are deposited in a flat dissipative medium, their accumulation into a coplanar system of satellites seems reasonably well understood in spite of continued haggling over details. The timing of the formation depends on the temperature history of the disk when ices are the predominant constituents of the forming satellites. Once formed, the satellites will spiral toward the primary because of gas drag and interaction with density waves in the gaseous part (type I drift) (Ward, 1997). It is thus likely that at least the large regular satellites formed during the last stages of accretion of their host planet as the gas in the disk and the solar nebula was being dissipated. The small equatorial satellites residing close to their primaries have most likely been repeatedly disintegrated and reformed in a gas-free environment by planetesimals remaining in the heliocentric swarm. Instabilities in the distribution of large objects lead to mergers and eventual separation of the remaining objects by more than several Hill sphere radii, where the Hill sphere radius (sphere of influence) is defined as the distance from the satellite to its inner Lagrange point ($R_H \approx r(M_s/3M_P)^{1/3}$).

Refinements to the accretion process in disks take account of the effect of the primary and depletion of accreting particles in the feeding zone of a growing object (Greenzweig and Lissauer, 1990, 1992; Lissauer, 1987; Wetherill and Cox, 1984, 1985), where an upper bound on the gravitational enhancement factor of about 1000 is determined. Gas drag replenishes the feeding zone and thereby prolongs the accretion of a body in a disk, but only a fraction of such particles (10–40%) brought into the feeding zone by gas drag are actually accreted (Kary and Lissauer, 1995; Kary et al., 1993).

The previously mentioned discussion is focused on gas-free accretion of solids. However, particles near meter size do not tend to stick to one another during collisions, and they tend to spiral into the planet (or star) because of the head wind created by the less than Kepler speed of orbiting by the gas. Youdin and Goodman (2005) described a streaming instability that can concentrate boulders into quite large gravitationally unstable collections. The streaming instability arises because a concentration of boulders can force the gas within the concentration to orbit closer to the Kepler velocity. The interaction with the surrounding gas streaming by, also containing isolated boulders that are spiraling in, causes the concentration of boulders to increase and become gravitationally bound and collapse to large objects (planetesimals in circumstellar disks) as the relative kinetic energy of the boulders within the swarm is dissipated. In the solar nebula, these objects can exceed the size of the dwarf planet Ceres (Johansen et al., 2007, 2011, 2012). The streaming instability appears to be crucial in the accretion of solids in accretion disks to greater than kilometer size.

A.2 Viscous Spreading

An inherent property of disks orbiting a primary is that they tend to spread as angular momentum is exchanged between adjacent regions because of the Kepler shear. Gravitational, magnetic, and viscous torques and wave transport with shock dissipation are all possible candidates for angular momentum transport (Balbus, 2003). Viscous interchange results from the Kepler shear in the motion of the material orbiting the primary. If we represent the shear stress as due to a viscosity η, we can write the shear stress at a distance r from the primary $\omega_{r\phi} = \eta|\mathrm{d}v_\phi/\mathrm{d}r|_{\mathrm{rot}} = \eta r|\mathrm{d}\Omega_K/\mathrm{d}r| = 3\eta\Omega_K/2$, where $|\mathrm{d}v_\phi/\mathrm{d}r|_{\mathrm{rot}}$ is the radial gradient of the azimuthal velocity in the frame rotating at the Kepler angular velocity at distance r. For a gas,

$$\eta = \frac{m\vec{v}}{3A} \approx \frac{mc_s}{3A} \qquad [A.8]$$

where m is the mean molecular mass, $\vec{v} = \sqrt{8kT/\pi m} \approx c_s$ is the mean molecular speed with k being Boltzmann's constant, T is the temperature, A is the collision cross section of the molecules, and c_s is the sound speed (e.g., Sears and Salinger, 1975). Most of the regular satellites around the major planets have a substantial fraction of ice. So, if ice is to be in solid form as satellitesimals, disk temperatures must be $\lesssim 200$ K (depending on the pressure), so for molecular hydrogen, $c_s \approx 10^5$ cm s^{-1}. The kinematic viscosity $\nu = \eta/\rho_g = 2H_g\eta/\sigma_g \approx c_s H_g m/(A\sigma_g)$, where ρ_g is the gas density, σ_g is the surface mass density of the gas, and H_g is the half thickness of the gas disk.

For the gaseous disk, we choose H_g to be the scale height of the disk that is determined by isostasy in the direction perpendicular to the disk, $\mathrm{d}P/\mathrm{d}z = -\rho_g g_z = -mPg_z/kT = -mPz\Omega_K^2/kT$, where P is the pressure and $g_z = GMz/r^3 = z\Omega_K^2$ is the component of gravity in the z direction. The solution of this equation gives $P = P_0 \exp -z^2/H_g^2$ where $H_g = \sqrt{2kT/m}/\Omega_K \approx c_s/\Omega_K$ is the isothermal scale height. As an example, we can estimate the timescale for viscous transfer of angular momentum in a disk surrounding Jupiter that has the minimum mass for accreting

Author's personal copy

Ganymede at Ganymede's distance from Jupiter augmented with volatiles to reach solar composition. The amount of solids in the disk necessary to make Ganymede corresponds to a surface density of 3300 g cm^{-2} spread from the midpoint between Ganymede and Europa, and the midpoint between Ganymede and Callisto is thereby increased to $\sim 10^5$ g cm^{-2} such that the kinematic viscosity $\nu \approx 3 \times 10^5$ cm^2 s^{-1}. The timescale for viscous transfer of angular momentum at Ganymede's distance from Jupiter ($\sim 10^{11}$ cm) is $r^2/\nu = 1.2 \times 10^9$ years. This time is so long compared with other timescales associated with a circumplanetary disk that molecular viscosity cannot be important in the disk evolution.

If the gas is turbulent, a turbulent viscosity is invoked to increase the efficiency of angular momentum transport. In this approximation, the molecular mass in eqn [A.8] is replaced by the mass of a typical turbulent eddy $\sim \rho_g l^3$ and the cross section A by $\sim l^2$ where l is a typical eddy size. Finally, $\bar{v}$ is replaced by the mean random velocity of the turbulent eddy v_t, so $\eta_t = \rho_g l v_t$. Next, $v_t = \alpha c_s$, where α is a free parameter, $l \sim H_g = c_s/\Omega_K$, so that the kinematic viscosity appropriate for a turbulent medium is (Shakura and Sunyaev, 1973)

$$\nu_t = \frac{\alpha c_s^2}{\Omega_K} \qquad [A.9]$$

The free parameter α is used to indicate the strength of the non-molecular viscous coupling, and the models are often called 'α disks.' The choices of α are usually based on unsupported assertions about the strength of the turbulence. Lower bounds on the magnitude of angular momentum transport follow from the masses of the giant planets. An equivalent viscosity must be sufficient to prevent the forming planet from opening a gap in the protoplanetary disk before it has acquired a significant fraction of its observed mass (Lin and Papaloizou, 1985). Typically, assumed values of $\alpha = O(10^{-3})$ in applications. For shear stress $w_{r\phi}$, the torque across a cylindrical surface of radius r centered on the primary is $T_\nu = w_{r\phi} 2\pi r^2 2H_g = 6\pi\eta\Omega_K r^2 H_g = 3\pi\alpha\sigma_g c_s^2 r^2$ for an α disk. For the example of a minimum mass disk at Ganymede's distance from Jupiter with $c_s = 10^5$ cm s^{-1} and $\alpha = 10^{-3}$, $\nu_t \approx 10^{12}$ cm^2 s^{-1} and the timescale for viscous spreading $\sim$400 years.

If we consider the mass dm in the cylindrical region of radial extent dr with specific angular momentum h, we have $D(h\mathrm{d}m)/Dt = -(\mathrm{d}T_\nu/\mathrm{d}r)\mathrm{d}r$ to balance the net torque on the cylinder with the change in the angular momentum, where $D/Dt \equiv \partial/\partial t + \mathbf{v}\cdot\nabla$ is the total derivative. In a steady-state disk,

$$\nu_r \frac{\mathrm{d}m}{\mathrm{d}r}\frac{\mathrm{d}h}{\mathrm{d}r} = F_r \frac{\mathrm{d}h}{\mathrm{d}r} = -\frac{\mathrm{d}T_\nu}{\mathrm{d}r} \qquad [A.10]$$

where F_r is the radial mass flux. For the general case where $T_\nu = 0$ at $r = 0$ and as $r \to \infty$ and with $\mathrm{d}h/\mathrm{d}r > 0$ as for Kepler motion, we see that $\mathrm{d}T_\nu/\mathrm{d}r$ must be positive at small r and negative at large r so that mass moves inward for small r and outward for large r. Angular momentum always moves outward.

Particle disks also spread from angular momentum transfer via particle collisions. For a disk of particles of mass m_p and velocity dispersion $\bar{v}$ about the Kepler velocity at distance r from the primary, the equivalent kinematic viscosity is given by $\nu_s = \eta_s/\rho_s = m_p\bar{v}/(3A\rho_s) = 2m_p\bar{v}H_s/(3A\sigma_s) = 4H_s\bar{v}\rho_p R_p/(9\sigma_s)$ (Daisaka et al., 2001), where σ_s is the surface mass density of particles in the disk, R_p is the particle radius, and ρ_p is the particle density (distinguished from the spatial mass density of particles ρ_s). For a disk of 100 m particles corresponding to the minimum mass particle disk at Ganymede's distance from Jupiter and the assumption $\bar{v} = v_{esc}$ of two touching particles, $H_s = Kv_{esc}/\Omega_K$ and $\rho_p = 1$ g cm^{-3}, ν_s is only a little over 80 cm^2 s^{-1}. So, low dispersion particle disks spread very slowly with timescale $r^2/\nu = 4.5 \times 10^{12}$ years. But inside the Roche radius in the lunar case, gravitational instabilities cause patches of higher-density material that are repeatedly sheared apart, thereby yielding enhanced collisional dissipation with an effective viscosity of $\nu \sim (\pi G\sigma_s)^2/\Omega_K^3$ and a resulting timescale $r^2/\nu_t \sim 1$ year (Ward and Cameron, 1978). This rapid spreading has been confirmed by N-body simulations (Takeda and Ida, 2001). But self-heating during this short timescale spreading is so intense, at least for the protolunar disk, that more silicates are vaporized, thereby decreasing the effective viscosity and allowing longer timescales for spreading (Thompson and Stevenson, 1988). So, the long timescales for spreading of particulate disks from the simple molecular viscosity analogy are not likely to prevail (see Salmon et al., 2010, for a thorough treatment of viscous spreading in particle disks).

A.3 Gas Drag

A forming satellite assumes a Kepler orbit about the primary, but the gas in the disk has its central acceleration reduced by a radial pressure gradient and thereby orbits slower than the satellites. The satellite thus experiences a continuous head wind that causes it to spiral toward the primary (Weidenschilling, 1977; Whipple, 1964). The gas motion is governed by the equation

$$\rho_g \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\rho_g \nabla\Phi - \nabla P \qquad [A.11]$$

where Φ is the gravitational potential per unit mass and P is the pressure. In a steady state, cylindrical coordinates $\ddot{r}, \ddot{\phi}, \ddot{z}, \dot{r}$ and $\dot{z}$ are all zero, and the r component of eqn [A.10] is

$$-\rho_g r\dot{\phi}^2 = -\rho_g \frac{\partial\Phi}{\partial r} - \frac{\partial P}{\partial r} \qquad [A.12]$$

Now, $P = \rho_g kT/m = \sigma_g c_s^2/2H_g = \sigma_g c_s \Omega_K/2$ (m = mean molecular mass). Often, the surface mass density and midplane temperatures are represented by power laws (e.g., Weidenschilling, 1977) like $\sigma_g = \sigma_{g0}(r_0/r)^p$ and $T = T_0(r_0/r)^q$, where σ_{g0} and T_0 are values at distance r_0 from the primary. Then, $\partial P/\partial r = -K\sigma_g\Omega_K c_s/r$, where $K = O(1)$ depends on the choices for p and q. With $\Phi = -GM_p/r$, the solution of eqn [A.11] yields the gas angular velocity

$$\dot{\phi}^2 \equiv \Omega^2 = \frac{GM_p}{r^3} - \frac{K\sigma_g\Omega_K c_s}{\rho_g r^2} = \Omega_K^2 - \frac{2Kc_s^2}{r^2} \qquad [A.13]$$

from which

$$\Omega_K - \Omega \approx \frac{Kc_s^2}{r^2\Omega_K} \qquad [A.14]$$

where $\Omega_K + \Omega \approx 2\Omega_K$ has been used.

For circularly orbiting material, the gas drag force is given by

$$F_d = -C_d \pi R^2 \rho_g v_{rel}^2 \qquad [A.15]$$

where C_d is the drag coefficient, which depends on the Reynolds number $\mathcal{R} = 2v_{re1}R\rho_g/\eta$, where η is the molecular viscosity,

Author's personal copy

and $v_{rel} = r(\Omega_K - \Omega) = Kc_s^2/r\Omega_K$ is the relative velocity between the forming satellite and the gas, that is, the head wind. The Reynolds number dependence of C_d (e.g., Prandtl, 1952) leads to $C_d = 20/\mathcal{R}$ (Stokes drag) for $\mathcal{R} \lesssim 0.5$ and $C_d \approx 0.43$ for $\mathcal{R} \gtrsim 500$ (see Peale (1993) for an application of gas drag with full Reynolds number dependence for C_d). The torque on the satellite is rF_d, which when equated to the time derivative of the angular momentum given by $m_p\sqrt{GM_p r}$ yields an expression for $\dot{r}$. The timescale for gas drag changes in the orbit is then

$$\tau_{GD} \equiv \frac{r}{\dot{r}} = \frac{4}{3K^2}\frac{1}{C_d}\frac{\rho_p R}{\Omega_K \sigma_g}\left(\frac{r\Omega_K}{c_s}\right)^3 \sim 10^3 \text{ years}\left(\frac{10}{C_d}\right)\left(\frac{R}{2500\text{km}}\right)\left(\frac{\rho_p}{2\text{gcm}^{-3}}\right) \times \left(\frac{0.1}{c_s/r\Omega_K}\right)^3\left(\frac{3\times 10^5 \text{gcm}^{-2}}{\sigma_g}\right)\left(\frac{r}{15R_J}\right)^{3/2} \quad [A.16]$$

where we have used $K = 3/4$ for $p = q = 1$, and the numerical form is for Ganymede (Canup and Ward, 2002), where the minimum mass particle disk has been augmented by ~30 times as much gas to yield solar composition.

A.4 Migration from Density Waves

Satellites also interact with disks by creating spiral density waves at the commensurabilities of the mean motions in the disk with that of the satellite (e.g., Ward, 1997). The resulting distribution of mass in the density waves exerts gravitational torques on the satellite that tend to push the satellite away from the site of the density wave. Hence, the density waves generated inside the satellite orbit tend to increase its angular momentum, whereas those outside tend to decrease it. Simultaneously, the ring material receives or gives up angular momentum, which tends to push the ring material away from the satellite. The distribution of the resonances outside the satellite versus those inside leads to the outside density waves dominating, such that the net effect is for the satellite to spiral toward its primary, called type I drift, with a timescale given by (Canup and Ward, 2002; Ward, 1997)

$$\tau_I \equiv \frac{r}{\dot{r}} \approx \frac{1}{C_a\Omega_K}\left(\frac{M_p}{M_s}\right)\left(\frac{M_p}{r^2\sigma_g}\right)\left(\frac{c_s}{r\Omega_K}\right)^2 \sim 10^2 \text{ years}\left(\frac{3}{C_a}\right)\left(\frac{2500\text{km}}{R}\right)^3\left(\frac{2\text{gcm}^{-3}}{\rho_p}\right)\left(\frac{c_s/r\Omega_K}{0.1}\right)^2 \left(\frac{3\times 10^5 \text{gcm}^{-2}}{\sigma_g}\right)\left(\frac{15R_J}{r}\right)^{1/2} \quad [A.17]$$

where C_a is a torque asymmetry parameter that is a function of the disk's radial surface density and temperature profiles (e.g., Tanaka et al., 2002; Ward, 1997); M_p and M_s are the planet and satellite masses, respectively; and r is the distance from the M_p center of mass. The last term in eqn [A.17] is $(H_g/r)^2$. The numerical representation is that centered on Ganymede in a minimum mass Jovian disk. Recent works (e.g., Kley and Crida, 2008; Paardekooper and Mellema, 2006a,b) have developed revised type I torque formula for the case of non-isothermal, adiabatic disks, which depend on local disk properties. These are discussed in the Jupiter section.

The repulsion of the gas disk by the satellite can eventually open a gap in the disk if (e.g., Ward and Hahn, 2000)

$$\frac{M_s}{M_p} > c_v\sqrt{\alpha}\left(\frac{c_s}{r\Omega_K}\right)^{5/2} \sim 10^{-4}c_v\left(\frac{\alpha}{10^{-3}}\right)^{1/2}\left(\frac{c_s/r\Omega_K}{0.1}\right)^{5/2} \quad [A.18]$$

where $c_v = O(1\text{–}10)$. This form of the criterion comes from Ward and Canup (2002), where they pointed out that the limiting mass is just above Ganymede's mass for $c_v = 1$ and $\alpha = 10^{-3}$. If a gap is opened, the type I drift is replaced by type II, whose timescale is now the disk viscous timescale (Ward, 1997)

$$\tau_{II} = \frac{r^2}{\nu_t} = \frac{(r/H_g)^2}{\alpha\Omega_K} \sim 300 \text{ years}\left(\frac{10^{-3}}{\alpha}\right)\left(\frac{0.1}{H/r}\right)^2\left(\frac{r}{15R_J}\right)^{3/2}\left(\frac{M_J}{M_p}\right)^{1/2} \quad [A.19]$$

where again the numerical representation is centered on Ganymede and M_p is normalized by the Jupiter mass M_J. Whether or not a gap is opened by a satellite depends on the value of α and the scale height of the disk at the satellite location. For an alternative gap-opening criterion also depending on these two parameters, see Crida et al. (2012).

A.5 Thermodynamics

The temperature distribution in the disk is of fundamental importance in determining nearly all of the parameters associated with accretion within the disk and whether or not the accreted particles survive migration toward the primary. The growing satellites will be near the midplane of the disk, so the midplane temperature T_c must be known. This in turn is determined by an energy balance equation in steady-state disks, where sources of energy include the gravitational energy deposited by the flux of incoming material, the viscous dissipation within the disk, radiation from the still warm primary, and radiation from the surrounding solar nebula. The sink is radiation from the upper and lower surfaces of the disk at some effective temperature, T_e. If the accretion of the satellites is assumed to take place in a disk with no accretion from the solar nebula, the gravitational source of energy now comes from the redistribution of mass during viscous evolution, and the internal energy of the disk is reduced as it cools.

For optically thick disks, $T_e^4 = 2T_c^4/(1 + 3\tau/2) \approx 4T_c^4/3\tau$, whereas for optically thin disks, $T_e^4 \approx 4\tau T_c^4$ (e.g., Shapiro and Teukolsky, 1983), where $\tau = \int_0^{H_g} \kappa\rho_g dz \approx \overline{\kappa}\sigma_g/2$ is the optical depth from midplane to the surface, with κ being the Rosseland mean opacity ($\text{cm}^2\,\text{g}^{-1}$) (e.g., Schwarzschild, 1958). An example of a surface mass distribution in a steady-state disk from Canup and Ward (2002) (Lynden-Bell and Pringle, 1974) is

$$\sigma_g(r) \approx \frac{4\dot{M}}{15\pi\nu_t}\left[\frac{5}{4} - \sqrt{\frac{r_c}{r_d}} - \frac{1}{4}\left(\frac{r}{r_c}\right)^2\right] \text{ for } r < r_c \approx \frac{4\dot{M}}{15\pi\nu_t}\left[\sqrt{\frac{r_c}{r}} - \sqrt{\frac{r_c}{r_d}}\right] \text{ for } r \geq r_c \quad [A.20]$$

where the kinematic viscosity $\nu_t = \alpha c_s^2/\Omega_K$ is the approximation of Shakura and Sunyaev (1973) for a turbulent medium discussed in the preceding text, r_c is the radius inside of which the mass flux $\dot{M}$ is deposited uniformly, and r_d is the edge of the

Author's personal copy

disk. Another example is a disk with constant mass flux toward the primary throughout the region where the satellites are forming (e.g., Lynden-Bell and Pringle, 1974):

$$\sigma_g(r) = \frac{\dot{M}}{3\pi\nu_t}\left[1-\sqrt{\frac{R_P}{r}}\right] \qquad [A.21]$$

if the viscous couple vanishes at the primary's surface of radius R_P.

Figure 4 shows the Rosalind mean opacities appropriate for low temperatures in an accretion disk. The opacity due to Bell and Lin (1994) accounts for ice grains at low temperatures and 'metallic' grains at high temperatures. Ice particles dominate the opacity up to temperatures slightly above 160 K with approximately a T^2 power law dependence. Above this temperature, the ice evaporates with a sharp drop in opacity and 'metallic' grains dominate the opacity for larger T. The Pollack et al. (1994) opacities include a distribution of organic particles that maintain a high opacity in the region beyond the ice point. These opacities can be modeled by writing $\kappa = \kappa_0 T^{\xi}$, but different values of the exponent ξ apply in different parts of the disk. There have been no studies of the formation of icy satellites with this accounting of the variations in the disk opacities. An example of the energy balance equation for a steady-state disk at distance r from the primary is

$$\sigma_{SB}T_P^4\left[\frac{2}{3\pi}\left(\frac{R_P}{r}\right)^3+\frac{1}{2}\left(\frac{R_P}{r}\right)^2\frac{H}{r}\left(\frac{d\ln H}{d\ln r}-1\right)\right] + F_{grav}(r)+F_{vis}(r) = 2\sigma_{SB}\left(T_e^4-\frac{\beta_n}{4\pi}T_{neb}^4\right) \qquad [A.22]$$

where the first bracketed term on the left is the radiation deposited per unit area from the warm planet (Ruden and Pollack, 1991), $F_{grav}(r)$ is the energy deposited due the accretion, and $F_{vis}(r)$ is the energy deposited per unit surface area from viscous interactions. The fraction of solid angle subtended by the solar nebula is $\beta_n/4\pi$. The most extreme model dependences are in the terms F_{grav} and F_{vis}. There are no constraints on F_{grav} from simulations, and F_{vis} depends on the unknown value of α. The importance of the opacity κ for the midplane properties cannot be overemphasized.

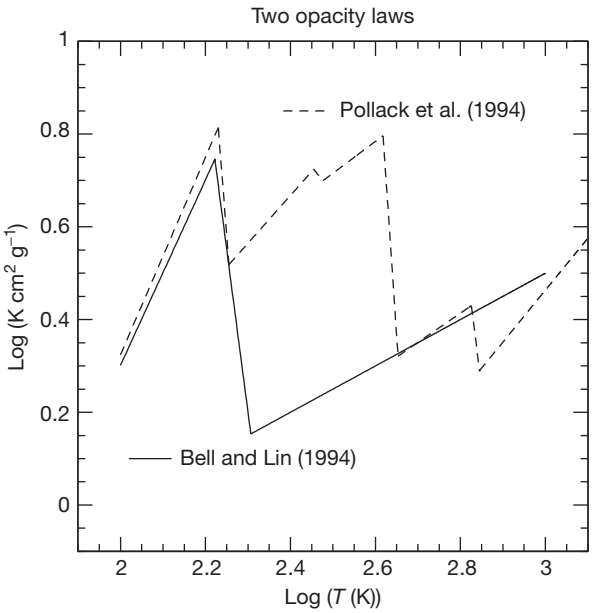


Figure 4 Power law approximations to the Rosseland mean opacities for a cold accretion disk. Adapted from Bell KR and Lin DNC (1994) Using Fu Orionis outbursts to constrain self-regulated protostellar disk models. *Astrophysical Journal* 474: 987–1004; Pollack JB, Hollenbach D, Beckwith S, Simonelli DP, Roush T, and Fong W (1994) Composition and radiative properties of grains in molecular clouds and accretion disks. *Astrophysical Journal* 421: 615–639.

Appendix B Tides

Tidal dissipation can drastically change the orbital configuration and the thermal state of the regular natural satellites from the time of their formation. So, it is imperative to understand this evolution, which puts constraints on the conditions of satellite formation. We therefore outline a theory of tidal effects on satellite orbits (and rotations) in this appendix. The potential of mass element dm_0 at $\mathbf{r}$ (within m_0) due to a mass M located at $\mathbf{R}$ relative to a coordinate system fixed in the body m_0 is given by

$$dV = -\frac{GM\,dm_0}{|\mathbf{R}-\mathbf{r}|} = -\frac{GM\,dm_0}{R}\left[1+\frac{r}{R}P_1(\cos S)+\left(\frac{r}{R}\right)^2P_2(\cos S)+\left(\frac{r}{R}\right)^3P_3(\cos S)\ldots\right] \qquad [B.1]$$

where G is the gravitational constant, P_i are the Legendre polynomials, and $\cos S = \mathbf{r}\cdot\mathbf{R}/rR$. The first two terms in the expansion combine to yield an acceleration parallel to $\mathbf{R}$, so the lowest-order tide-raising potential per unit mass in m_0 due to M is given by the second harmonic term

$$W_2 = -\frac{GMr^2}{2R^3}\left[3\cos^2 S-1\right] \qquad [B.2]$$

Typically, $r/R \ll 1$ allows neglect of higher-order terms, but if a satellite is close to a primary (e.g., Phobos), higher-order terms contribute significantly to the total tidal interaction (Bills et al., 2005).

For a homogeneous sphere, Love (1944) showed that for a small distortion, the displacement of the surface due to the disturbing potential W_2 is a small factor ϵ_2 times a surface harmonic S_2 of the same degree:

$$\epsilon_2 S_2 = \frac{5}{2g}\frac{W_2(r=a_e)}{[1+19\mu/(2\rho g a_e)]} \qquad [B.3]$$

where g is the surface gravity acceleration on m_0, μ is the rigidity, ρ is the density, and a_e is the radius of the sphere. The potential per unit mass at $r > a_e$ due to this spherical surface harmonic mass distribution is then

$$-\frac{3}{5}\left(\frac{a_e}{r}\right)^3\epsilon_2 S_2 = -\frac{3}{2[1+19\mu/(\rho g a_e)]}\frac{GM}{2R^3}\times\frac{a_e^5}{r^3}\left(3\cos^2 S-1\right) \qquad [B.4]$$

such that if we place a mass m at $\mathbf{r}$,

$$V_T = -\frac{k_2 GMma_e^5}{2R^5r^5}\left[3(\mathbf{R}\cdot\mathbf{r})^2-r^2R^2\right] \qquad [B.5]$$

is the potential of m at $\mathbf{r}$ due to a tide raised by M on m_0 or the potential of M at $\mathbf{R}$ due a tide raised by m on m_0. The potential Love number $k_2 = 1.5/[1+19\mu/(2\rho g a_e)]$ refers to the body on which the tide is raised. For centrally condensed bodies, k_2 is

Author's personal copy

reduced. For example, the fluid Love number for the Earth is 0.934 instead of 1.5 (Lambeck, 1980, p. 29).

Now, consider M and m to be the same body. We model the effects of the dissipation by assuming the equilibrium response of m_0 to the tidal stress applied by m lags by some time Δt by analogy to the damped harmonic oscillator forced at a frequency that is small compared to the natural frequency of oscillation. Then, $\mathbf{R}(t)=\mathbf{r}(t-\Delta t)$ is the simplest representation of the effects of dissipation of tidal energy in m_0, that is, $\mathbf{R}$ is the position of m a short time in the past relative to a coordinate system fixed in m_0, where $\mathbf{r}$ is defined in the same coordinate system. The components of $\mathbf{R}=\mathbf{r}(t-\Delta t)$ are not to be included in differentiations of V_T in determining the tidal forces on m. We distinguish these coordinates with a * and write the coordinates of m considered the tide-raising body as

$$\mathbf{r}^*(t-\Delta t)=\mathbf{r}^*(t)-\left.\frac{d\mathbf{r}^*}{dt}\right|_t\Delta t=\mathbf{r}^*+\Delta\mathbf{r}^* \qquad [B.6]$$

If we substitute the right-hand side of eqn [B.6] into eqn [B.5] and expand to the first order in Δr^*, the results were

$$V_T=-k_2Gm^2a_e^5\left(\frac{1}{r^3r^{*3}}+\frac{3\mathbf{r}\cdot\Delta\mathbf{r}^*}{r^4r^{*4}}-\frac{6\mathbf{r}^*\cdot\Delta\mathbf{r}^*}{r^3r^{*5}}\right) \qquad [B.7]$$

where

$$\Delta\mathbf{r}^*=-\left.\frac{d\mathbf{r}^*}{dt}\right|_{body}\Delta t=-\Delta t\left(\frac{dx^{*'}}{dt}\mathbf{e}'_1+\frac{dy^{*'}}{dt}\mathbf{e}'_2+\frac{dz^{*'}}{dt}\mathbf{e}'_3\right) \qquad [B.8]$$

and the time derivative is relative to the $x'y'z'$ system of coordinates fixed in the rotating body m_0. With the generic relation for a vector $\vec{D}$, $\left(d\vec{D}/dt\right)_{body}=\left(d\vec{D}/dt\right)_{space}-\vec{\omega}\times\vec{D}$ and with $\vec{D}\to\mathbf{r}^*$ and $\vec{\omega}\to\vec{\psi}$, where $\vec{\psi}$ is the angular velocity of m_0, we can write $d\mathbf{r}^*/dt|_{body}=d\mathbf{r}^*/dt|_{space}-\vec{\psi}\times\mathbf{r}^*$ and

$$\Delta\mathbf{r}^*\cdot\mathbf{r}^*=-\mathbf{r}\cdot\dot{\mathbf{r}}\Delta t \qquad [B.9]$$

where the derivative on the right-hand side is now relative to inertial space.

As an example, we assume m_0 is rotating about the z' axis at angular velocity $\dot{\psi}$ relative to the xyz axes fixed in inertial space with the xy and $x'y'$ planes being coincident, so that $\psi=\psi_0+\dot{\psi}t$ is the angle between the x and x' axes. Then,

$$\Delta\mathbf{r}^*=-\left(\dot{\psi}\gamma+\dot{x}\right)\Delta t\mathbf{e}_1+\left(\dot{\psi}x-\dot{\gamma}\right)\Delta t\mathbf{e}_2-\dot{z}\Delta t\mathbf{e}_3 \qquad [B.10]$$

The force on m due to the tide it raises on m_0 is found from the negative gradient of V_T in eqn [B.7] with respect to the unstarred coordinates, after which we can set $r^*=r$ and substitute eqns [B.9] and [B.10] into the gradient so obtained to yield

$$\begin{aligned}
-\frac{\partial V_T}{\partial x}&=-k_2Gm^2a_e^5\left[\frac{3x}{r^8}+\frac{6\mathbf{r}\cdot\dot{\mathbf{r}}x\Delta t}{r^{10}}+\frac{3\left(\dot{\psi}\gamma+\dot{x}\right)\Delta t}{r^8}\right]\\
-\frac{\partial V_T}{\partial x}&=-k_2Gm^2a_e^5\left[\frac{3\gamma}{r^8}+\frac{6\mathbf{r}\cdot\dot{\mathbf{r}}\gamma\Delta t}{r^{10}}+\frac{3\left(-\dot{\psi}x+\dot{\gamma}\right)\Delta t}{r^8}\right]\\
-\frac{\partial V_T}{\partial z}&=-k_2Gm^2a_e^5\left[\frac{3z}{r^8}+\frac{6\mathbf{r}\cdot\dot{\mathbf{r}}z\Delta t}{r^{10}}+\frac{3\dot{z}\Delta t}{r^8}\right]
\end{aligned} \qquad [B.11]$$

The gradient $-\nabla V_T$ from eqn [B.11] is the force on m due to tides raised on m_0 by m. If the orbital angular velocity of m is always less than and in nearly the same direction as the spin angular velocity of m_0 (prograde), this force will increase m's orbital semimajor axis and eccentricity and decrease the orbital inclination relative to m_0's equatorial plane. Here, $k_2\to k_2^{m_0}$, the Love number for m_0; $a_e\to R_{m_0}$, the equatorial radius of m_0; $\dot{\psi}\to\dot{\psi}_{m_0}$, the angular velocity of m_0; and $\mathbf{r}\to\mathbf{r}_{m_0m}$, the vector from the center of m_0 to the center of m with corresponding components in inertial space of x, y, and z.

The force on m_0 due to tides raised on m by m_0 has the same form as eqn [B.11] with now $k_2\to k_2^m$, the Love number for m, $m\to m_0$; $a_e\to R_m$, the radius of m; and $\mathbf{r}=\mathbf{r}_{mm_0}=-\mathbf{r}_{m_0m}$. To convert this force to the force on m due to tides raised on m by m_0, the change in sign of the force expressions by Newton's third law is canceled by the change in the sign of the coordinates effected by setting $\mathbf{r}=\mathbf{r}_{\mathbf{m_0m}}$ instead of $-\mathbf{r}_{m_0m}$. So, eqn [B.11] expresses the force on m due to tides raised on m by m_0 with $k_2\to k_2^m$, $m\to m_0$, $a_e\to R_m$ and $\mathbf{r}\to\mathbf{r}_{m_0m}$.

The torque on m due to tides raised by itself on m_0 is $T_T=\mathbf{r}\times(-\nabla V_T)$, and the torque retarding m_0's spin is the negative of this, where $\mathbf{r}=\mathbf{r}_{m_0m}$. Similarly, the torque retarding m's spin due to tides raised on m by m_0 has the same form, but with $\mathbf{r}=\mathbf{r}_{mm_0}$ and $m\to m_0$, $a_e\to R_m$ and $k_2\to k_2^m$. This latter torque on m can be represented by

$$\boldsymbol{T}_T=\frac{3k_2Gm_0^2R^5}{r^6}(\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}_T)(\hat{\mathbf{r}}_T\times\hat{\mathbf{r}}) \qquad [B.12]$$

where we have omitted the sub- and superscripts but remember that the variables apply to m. $\hat{\mathbf{r}}$ is a unit vector toward the tide-raising body m_0, and $\hat{\mathbf{r}}_T$ is a unit vector toward the tidal maximum, which is the subprimary position on m a short time Δt in the past as in the preceding text. Equation [B.12] is valid for arbitrary inclination of the orbit and equatorial planes. Similar to the earlier-mentioned analysis,

$$\hat{\mathbf{r}}_T=\hat{\mathbf{r}}-\frac{d\hat{r}}{dt}\Delta t \qquad [B.13]$$

where the time derivative is relative to the body system of coordinates. Replacement of $\hat{\mathbf{r}}_T$ with eqn [B.13] yields

$$\boldsymbol{T}_T=\frac{3k_2Gm_0^2R^5\Delta t}{r^6}\hat{\mathbf{r}}\times\dot{\hat{\mathbf{r}}} \qquad [B.14]$$

where we have set $\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}_T=1$. So, with the generic relation between the derivatives of a vector relative to body and space coordinates described previously with $\hat{\mathbf{r}}=\cos f\mathbf{e}_1+\sin f\mathbf{e}_2$ and $\dot{\boldsymbol{\psi}}=\dot{\psi}(\sin i\sin\Omega'\mathbf{e}_1-\sin i\cos\Omega'\mathbf{e}_2+\cos i\mathbf{e}_3)+\dot{\Omega}'\mathbf{e}_3$, it is easy to obtain

$$\begin{aligned}
\hat{\mathbf{r}}\times\dot{\hat{\mathbf{r}}}=&\dot{\psi}\sin i\cos(f-\Omega')(-\sin f\mathbf{e}_1+\cos f\mathbf{e}_2)\\
&+\left(\dot{f}-\dot{\Omega}'-\dot{\psi}\cos i\right)\mathbf{e}_3
\end{aligned} \qquad [B.15]$$

where i and Ω' are inclination and longitude of the ascending node of the $x'y'$ (equator) plane on the xy (orbit) plane. Equation [B.15] is substituted into eqn [B.14] and averaged over the orbit period. Useful averages are

$$\begin{aligned}
\left\langle\frac{a^6}{r^6}\dot{f}\right\rangle&=n\left(1+\frac{15}{2}e^2+\frac{45}{8}e^4+\frac{5}{16}e^6\right)/(1-e^2)^6=nf_1(e)\\
\left\langle\frac{a^6}{r^6}\right\rangle&=\left(1+3e^2+\frac{3}{8}e^4\right)/(1-e^2)^{9/2}=f_2(e)\\
\left\langle\frac{a^6}{r^6}\cos 2f\right\rangle&=\left(\frac{3}{2}e^2+\frac{1}{4}e^4\right)/(1-e^2)^{9/2}=f_3(e)\\
\left\langle\frac{a^6}{r^6}\cos^2 f\right\rangle&=\frac{f_2(e)+f_3(e)}{2}\\
\left\langle\frac{a^6}{r^6}\sin^2 f\right\rangle&=\frac{f_2(e)-f_3(e)}{2}
\end{aligned} \qquad [B.16]$$

Author's personal copy

where $\dot{f} = n\left(a^2\sqrt{1-e^2}\right)/r^2$ has been used in the first of eqn [B.16] with a and e being the orbital semimajor axis and eccentricity, respectively. The averaged tidal torque on m due to tides raised by m_0 on m is thus

$$\langle \boldsymbol{T}_T \rangle = \frac{3k_2 G m_0^2 R^5 \Delta t}{a^6}\left\{-\dot{\psi}\sin i \sin\Omega'\left(\frac{f_2(e)-f_3(e)}{2}\right)\boldsymbol{e}_1 + \dot{\psi}\sin i\cos\Omega'\left(\frac{f_2(e)+f_3(e)}{2}\right)\boldsymbol{e}_2 + \left[nf_1(e) - f_2(e)\dot{\psi}\cos i\right]\boldsymbol{e}_3\right\} \quad [B.17]$$

where we have neglected $\dot{\Omega}'$ compared to n and $\dot{\psi}$.

As an example, consider the retardation of the rotation of a satellite in a circular ($e=0$) and equatorial ($i=0$) orbit from the tidal torque. The time rate of change of the angular momentum of the satellite $C\left(\mathrm{d}\dot{\psi}/\mathrm{d}t\right) = \langle T_T \rangle$, where C is the moment of inertia of m about its spin axis and where only the $\mathbf{e}_3$ term remains in eqn [B.17]. So,

$$\frac{\mathrm{d}\dot{\psi}}{\mathrm{d}t} = -\frac{3k_2 G m_0^2 R^5}{Ca^6}\Delta t\left(\dot{\psi}-n\right) = -\frac{15}{2}k_2\frac{m_0}{m}\frac{R^3}{a^3}\frac{n}{Q_0}\left(\dot{\psi}-n\right) \quad [B.18]$$

where $\Delta t = 1/(Q_0 n)$, $C = 2mR^2/5$, and $n^2 = Gm_0/a^3$ have been used. The expression for Δt follows from the discussion in the succeeding text with Q_0 being the dissipation function at frequency n. The solution of eqn [B.18] is a decaying exponential for $\dot{\psi} - n$ (if n is nearly constant) with time constant

$$\tau_T = \frac{2}{15}\frac{1}{k_2}\frac{m}{m_0}\frac{a^3}{R^3}\frac{Q_0}{n} \quad [B.19]$$

where $k_2 = 0.05$ and $Q_0 = 100$, $\tau_T \approx 4300$ years for Ganymede and 57 000 years for Callisto at their current distances from Jupiter. Note that for $e=0$, the asymptotic tidal state is $\dot{\psi} = n$ (i.e., rotation synchronous with the orbital mean motion), but for $e \neq 0$, the asymptotic tidal state is $\dot{\psi} > 0$ determined by the vanishing of $\langle T_T \rangle$ in eqn [B.17]. Satellites with $e \neq 0$ are able to maintain synchronous rotation if the torque on the permanent axial asymmetry dominates $\langle T_T \rangle$ and forces the axis of minimum moment of inertia to librate about the direction to the primary whenever the satellite is at periapse.

The relationship between the dissipation function Q and Δt follows from a simple example. The dissipation parameter Q for a system oscillating at frequency ω is defined by (e.g., Lambeck, 1980, p. 14)

$$\frac{1}{Q} = \frac{\oint \frac{\mathrm{d}E}{\mathrm{d}t}\mathrm{d}t}{2\pi E^*} = \omega\Delta t \quad [B.20]$$

where the numerator is the energy ΔE dissipated during a complete cycle of oscillation and E^* is the maximum energy stored during the cycle. For a tidally distorted, nearly spherical body with the disturbing body in a circular equatorial orbit, a cycle would consist of half a rotation of the distorted body relative to the body causing the tide. For a complex tide generated by a noncircular, nonequatorial orbit of the disturbing body, each periodic term in a Fourier decomposition of the tide would have its own maximum stored energy and dissipation over a complete cycle of oscillation. The response of an oscillator with forcing function $F = A'\sin\omega t$ when $\omega \ll \omega_0$, with ω_0 being the lowest frequency of free oscillation, can be represented by $x = B'\sin\omega(t-\Delta t)$, where Δt is the phase lag in the response due to the dissipation as was assumed in the preceding text for the tidal response. The rate at which the forcing function does work is $\mathrm{d}E/\mathrm{d}t = F\dot{x} = A'B'\omega\sin\omega t\cos\omega(t-\Delta t)$. Then,

$$E(t) = \int_{t_1}^{t} F\dot{x}\,\mathrm{d}t = A'B'\omega\left[\frac{-\cos(2\omega t - \omega\Delta t)}{4\omega} + \frac{\sin\omega\Delta t}{2}t\right]_{t_1}^{t} \quad [B.21]$$

The first term in eqn [B.21] is the periodic storage of energy in the oscillator, and the second is the secular loss of energy. The energy dissipated in a period $2\pi/\omega$ is $\Delta E = A'B'\pi\sin\omega\Delta t$. The peak energy stored is found by integrating $\mathrm{d}E/\mathrm{d}t$ from a value of t where $x=0$ to a value of t where $x = x_{max}$. Since $x \propto \cos(\omega t - \omega\Delta t)$ and $\dot{x} \propto -\omega\sin(\omega t - \omega\Delta t)$, $x=0$ when $\omega t - \omega\Delta t = (2n-1)\pi/2$ and $x = x_{max}$ when $\dot{x} = 0$ or $\omega t - \omega\Delta t = n\pi$. If we choose the limits spanning $t=0$, then the interval of integration is $-\pi/2 + \omega\Delta t \leq \omega t \leq \omega\Delta t$, where $x=0$ at the lower limit and $x = x_{max}$ at the upper limit. Applying these limits to the first term in eqn [B.21] to get the stored energy yields

$$E^* = A'B'\left[\frac{\cos\omega\Delta t}{4} - \frac{-\cos\omega\Delta t}{4}\right] = \frac{A'B'}{2}\cos\omega\Delta t \quad [B.22]$$

Substitution of the expressions for E^* and ΔE into eqn [B.20] yields $1/Q = \tan\omega\Delta t \approx \omega\Delta t$ when the dissipation is relatively small (large Q). Q is then proportional to $1/\omega$ for the assumption in this tidal model that Δt is constant.

The tidal effective Q of the solid Earth has been estimated from satellite observations to be 280 ± 70 (Ray et al., 2001), which is close to the value obtained by the attenuation of normal modes of the Earth (Widmer et al., 1991). For Mars, $Q \approx 100$ from the secular acceleration of the satellite Phobos (Bills et al., 2005; Smith and Born, 1976). For Mars, the dominant frequency is $\dot{\omega} = 2\left|\dot{\psi} - n_P\right|$, where $\dot{\psi}$ is Mars' rotational angular velocity and n_P is Phobos' mean orbital motion. The factor 2 follows from there being two tidal cycles for each synodic period. With a rotation period of 24 h 37 min 23 s and an orbital period of 0.319 days, $\omega = 3.14 \times 10^{-4}$ rad s^{-1} leading to $\Delta t \approx 32$ s for $Q = 100$. For Mercury, the fundamental tidal period is the orbit period, since Mercury rotates 180° relative to the Sun for each orbit. For the attenuation of seismic waves in the Earth and from laboratory experiments on Earth-like materials, Q is proportional to frequency raised to a small, positive fractional (0.2–0.4) power with nominal values near $Q = 100$ (e.g., Shito et al., 2004), but to avoid an infinite discontinuity in the torque whenever a tidal frequency passes through zero while changing sign, an alternative behavior must be invoked (see the discussion on Andrade and Maxwell tidal models).

For energy dissipation within a satellite, the approach of Peale and Cassen (1978) is simplified by Wisdom (2004). The rate at which work is done on a satellite by the tide is given by

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -\int_{\mathrm{body}} \rho\mathbf{v}\cdot\nabla W_2\,\mathrm{d}V \quad [B.23]$$

where W_2 is the previously mentioned tide-raising potential and $\mathbf{v}$ is the velocity of a volume element $\mathrm{d}V$ of density ρ. If we

Author's personal copy

can assume the satellite is incompressible, then $\nabla\cdot\mathbf{v}=0$, and from $\nabla\cdot W_2\mathbf{v}=\mathbf{v}\cdot\nabla W_2+W_2\nabla\cdot\mathbf{v}$, we can write

$$\frac{dE}{dt}=-\int_{body}\rho\nabla\cdot(W_2\mathbf{v})\,dV=-\rho\int_{surface}W_2\mathbf{v}\cdot n\,dS \qquad [B.24]$$

where the last form follows from divergence theorem. In eqn [B.24], $\mathbf{n}$ is the normal to a surface element dS, and we have taken ρ outside the integral by assuming a uniform density.

The rate at which the height of the surface changes is $\mathbf{v}\cdot\mathbf{n}=d\Delta R/dt\approx h_2 dW_2'/dt/g$ (eqn [B.3]), where $h_2=5\,k_2/3$ is the second degree displacement Love number (Love, 1944) and g is the surface gravity acceleration of the satellite. The prime on W_2 indicates the time delay in the response of the body to the forcing function. Then,

$$\frac{dE}{dt}=-\frac{\rho h_2}{g}\int_{surface}W_2\frac{d}{dt}(W_2')\,dS \qquad [B.25]$$

If higher-order terms are necessary in the tidal potential for close satellites, eqn [B.25] would be a sum of terms involving the Love numbers h_3, h_4, etc.

If we consider the case where the satellite equatorial and orbital planes are coincident, $\mathbf{r}=(r\cos f,\ r\sin f,\ 0)$, and $\mathbf{R}=\left(R\cos\left(\phi+\dot\psi t\right),\ R\sin\left(\phi+\dot\psi t\right),\ 0\right)$, where R and ϕ are the ordinary spherical radial and azimuthal coordinates of a point on the surface in a coordinate system fixed in the satellite, and the $\dot\psi t$ accounts for the rotation of the satellite relative to an inertial system. For small eccentricities, we can write $a/r\approx 1+e\cos nt$, $\cos f\approx\cos nt+e(\cos 2nt-1)$ and $\sin f=\sin nt+e\sin 2nt$ to the first order in eccentricity e (e.g., Murray and Dermott, 1999). For the time-lagged potential W_2', t is replaced by $t-\Delta t$ and for a satellite rotating synchronously with its orbital motion, $\dot\psi=n$. With these substitutions, the average rate of energy dissipation is found by performing the surface integral and averaging over an orbit period:

$$\frac{dE}{dt}=-\frac{42}{5}\frac{\pi\rho h_2}{g}\frac{G^2m_p^2R^6}{a^6}e^2n\sin n\Delta t\approx-\frac{42}{19}\frac{\pi\rho^2n^5R^7e^2}{\mu Q} \qquad [B.26]$$

where $h_2\approx 5\rho gR/(19\mu)$ from the definition of k_2, $n^2=Gm_p/a^3$ has been used, and $\sin n\Delta t\to n\Delta t=1/Q$. The last form in eqn [B.26] agrees with that in Peale et al. (1979) as corrected in Peale (2003). Equation [B.26] is the lowest-order approximation for the tidal dissipation in a synchronously rotating satellite with orbital eccentricity e. For very large eccentricities, higher-order terms must be included.

The tidal dissipation in a synchronously rotating satellite will tend to reduce the eccentricity as $e\neq 0$ is the cause of the dissipation. The spin angular momentum of the satellite is conserved because of the synchronous lock. The specific orbital angular momentum $[G(m_p+m_s)a(1-e^2)]^{1/2}$ can thus not gain angular momentum from the satellite spin. The orbital energy $-Gm_pm_s/2a$ must decrease if energy is dissipated in the satellite, and a must thereby decrease. But the conserved angular momentum means e must decrease if a decreases. At the same time, the tide raised on the planet by the satellite tends to increase the eccentricity if the spin angular velocity $\dot\psi_p$ exceeds a value near $1.5n$ for small eccentricity (depending on the tidal model). The rate of change of the eccentricity of a satellite orbit due to tides raised on the primary by the satellite can be determined from the Lagrange planetary equations as expressed in terms of the perturbing accelerations S, T, and W (radial, tangential, and perpendicular to the orbit) (e.g., Plummer, 1960 (1918), p. 151). For the tidal model adopted herein, S, T, and W can be determined from the xyz components given in eqn [B.11]. Substitution of the required S and T components into the Lagrange equation for de/dt averaged over an orbit period yields

$$\frac{de}{dt}=\frac{k_{2p}}{Q_p}\frac{m_s}{m_p}\frac{R_p^5}{a^5}\frac{ne}{(1-e^2)^{13/2}}\left[\frac{\dot\psi_p(1-e^2)^{3/2}}{n}\left(\frac{33}{2}+\frac{99e^2}{4}+\frac{33e^4}{16}\right)-\left(27+\frac{405e^2}{4}+\frac{405e^4}{8}+\frac{135e^6}{64}\right)\right] \qquad [B.27]$$

where Q_p is the value of Q corresponding to frequency of the orbital mean motion n and the subscripts s and p refer to the satellite and planet, respectively. The value of the planetary angular velocity $\dot\psi_p$ above which $de/dt>0$ increases with eccentricity, which follows because the planet has to be rotating sufficiently faster than the orbital motion at periapse to give it the kick that increases the eccentricity. Otherwise, tides raised on the primary will reduce the eccentricity. In the limit of small eccentricity, $\dot\psi_p\gtrsim 1.64n$ for $de/dt>0$. In a model where the tidal potential is separated into its component frequencies with individual phase lags, Goldreich (1963) found that $\dot\psi_p\gtrsim 1.5n$ for positive de/dt.

The magnitude of the rate of decrease in the eccentricity due to tides raised on the satellite can be found in the same way by interchanging the satellite and planetary parameters as discussed in eqn [B.11], where now $\dot\psi_s\equiv n$ leads to $de/dt<0$ for a satellite rotating synchronously with its orbital mean motion for all values of e. The variation in the eccentricity to lowest order in e from tidal dissipation in the planet and the satellite is thus

$$\frac{de}{dt}=\frac{k_{2p}}{Q_p}\frac{m_s}{m_p}\frac{R_p^5}{a^5}e\left(\frac{33}{2}\dot\psi_p-27n\right)-\frac{21}{2}\frac{k_{2s}}{Q_s}\frac{m_p}{m_s}\frac{R_s^5}{a^5}ne \qquad [B.28]$$

where $\dot\psi_p$ and n are the spin angular velocity and orbital mean motion, respectively, and subscripts p and s refer to planet and satellite. (This equation replaces the erroneous eqn [B.21] of this chapter in the first edition of this volume of the *Treatise on Geophysics*.) Whether eccentricity of a satellite orbit increases or decreases as the orbit evolves under tidal evolution depends on the relative magnitude of the two terms in eqn [B.28].

The reader should be aware of the uncertainty and somewhat arbitrariness of tidal models. Our analysis of tidal dissipation using a constant time lag in the response of a planet or satellite to the forcing function leads to $Q\propto 1/\omega$, where ω is a tidal frequency, is the simplest. It has the most straightforward development while yielding relatively intuitive tidal evolution. It has the advantage of being applicable for arbitrary eccentricities, whereas most other models require an expansion in eccentricity (e.g., Kaula, 1964), which must be truncated. Each of the periodic terms in the Kaula expansion is lagged in phase from the same term in the tide-raising potential with the phase lag having the same sign as the frequency of the term. The magnitude of the phase lag is $1/Q$ and one can assume various frequency dependencies for the value of Q to represent different tidal models. Our constant Δt model would have $Q\propto 1/\omega$ where ω is the frequency of the term. A tidal

Author's personal copy

model often used is Q = constant, which has the nonphysical characteristic that each term in the expansion changes sign abruptly without reducing its magnitude as the frequency of that term passes through zero. Other frequency dependencies for Q can be substituted into the Kaula expansion to represent other assumptions about the dissipative properties of the material.

The major disadvantage of the constant Δt model for solid bodies is that it does not account for creep or residual distortion after the perturbing force is removed. A discussion of how creep is handled is given in Karato's (2008) book, where quite complicated frequency dependencies for Q may result. Some researchers now favor the empirically determined Andrade tidal model (Castillo-Rogez et al., 2011), but it mimics the $Q \propto 1/\omega$ model at very low frequencies and thereby no longer represents real materials. There is some evidence favoring the Maxwell model at very low frequencies, and Efroimsky (2012) had formulated a tidal model with the empirical Andrade model used for normal frequencies but transitioning to the Maxwell model for very low frequencies. All of these manipulations are for solid materials. Although it is usually inferred that partially molten bodies, as in our discussion of the Earth–Moon system, have a high dissipation compared to the completely solid or completely fluid bodies, there is no detailed tidal model to represent this state. The frustration of model dependence of all conclusions regarding the outcome of tidal dissipation persists.

We have seen in the preceding text that nominal values of Q (or Q_0 in the model detailed earlier in the text) are near 100 for Mars and for the damping of seismic waves in the Earth's mantle and near 300 for satellite determinations of the solid Earth response to the Moon and for the damping of seismically excited normal modes of the Earth. $Q = 100$ is an often adopted value representing Earth-like materials. However, the damping of seismic waves in the Moon's upper layers leads to $Q \approx 4000$ at 4 Hz and $Q \approx 7000$ at 8 Hz (Nakamura and Koyama, 1982), but the monthly tidal $Q \approx 37$ and the annual $Q \approx 60$ (Williams et al., 2001). The widely varying values of Q reflect the nature of the materials where the dissipation is taking place, where the dryness of the surface layers of the Moon may lead to the high values of Q, and partial melts in the interior may yield the low tidal Qs. Efroimsky (2012) pointed out the uncertainties in comparing seismic Qs with tidal Qs. Finally, a totally different procedure is needed for the unsolved problem of determining effective Qs of gaseous planets (and stars), where resonances of tidal frequencies with normal modes of oscillation of the gaseous body become important. In summary, the uncertainty of the proper tidal model to use in various circumstances persists, but the simplest tidal theory described previously in the text can be used to determine qualitative constraints on the origin of the natural satellites from their inferred evolutions.

Acknowledgments

We are grateful to David Stevenson for commenting on the Earth–Moon section. The support of SJP by the Planetary Geology and Geophysics Program under grant NNX08AL76G is gratefully acknowledged.

References

Agnor CB, Canup RM, and Levison HF (1999) On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus* 100: 307–325.

Agnor CB and Hamilton DP (2004) Satellite capture via binary exchange reactions: Application to Triton. *Bulletin of the American Astronomical Society* 36: 1169.

Agnor CB and Hamilton DP (2006) Neptune's capture of its moon Triton in a binary-planet gravitational encounter. *Nature* 441: 192–194.

Alfven H (1963) The early history of the Moon and the Earth. *Icarus* 1: 357–363.

Alfvén H and Arrhenius G (1969) Two alternatives for the history of the Moon. *Science* 165: 11–17.

Alfvén H and Arrhenius G (1972) Origin and evolution of the Earth-Moon system. *Moon* 5: 210–230.

Alfvén H and Arrhenius G (1976) *Evolution of the Solar System*. Washington, DC: NASA Scientific and Technical Information Office, NASA SP-345.

Alibert Y, Mordasini C, Benz W, and Winisdoerffer C (2005a) Models of giant planet formation with migration and disc evolution. *Astronomy and Astrophysics* 434: 343–353.

Alibert Y and Mousis O (2006) Structure and evolution of Saturn's subnebula – Implications for the formation of Titan. In: *Lunar and Planetary Science XXXVII*, Abstract 1141.

Alibert Y, Mousis O, and Benz W (2005b) Modeling the Jovian subnebula I. Thermodynamic conditions and migration of proto-satellites. *Astronomy and Astrophysics* 439: 1205–1213.

Anderson JD, Jacobson RA, McElrath TP, Moore WB, Schubert G, and Thomas PC (2001) Shape, mean radius, gravity field and interior structure of Callisto. *Icarus* 153: 157–161.

Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998a) Distribution of rock, metals and ice in Callisto. *Science* 280: 1573–1576.

Anderson JD, Scubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998b) Europa's differentiated internal structure: Inferences from four Galileo encounters. *Science* 281: 2019–2022.

Andert TP, Rosenblatt P, Pätzold M, et al. (2010) Precise mass determination and nature of Phobos. *Geophysical Research Letters* 37: L09202, 4 pp.

Andrews-Hanna JC, Asmar SW, Head JW, et al. (2013) Ancient igneous intrusions and early expansion of the Moon revealed by GRAIL gravity gradiometry. *Science* 339: 675–678.

Artymowicz P and Lubow SH (1996) Mass flow through gaps in circumbinary disks. *Astrophysical Journal Letters* 467: L77–L80.

Asphaug E and Reufer A (2013) Late origin of the Saturn system. *Icarus* 223: 544–565.

Balbus S (2003) Enhanced angular momentum transport in accretion disks. *Annual Review of Astronomy and Astrophysics* 41: 555–597.

Banfield D and Murray N (1992) A dynamical history of the inner Neptunian satellites. *Icarus* 99: 390–401.

Barr AC and Canup RM (2008) Constraints on gas giant satellite formation from the interior states of partially differentiated satellites. *Icarus* 198: 163–177.

Barr AC, Citron RI, and Canup RM (2010) Origin of a partially differentiated Titan. *Icarus* 209: 858–862.

Bate MR, Lubow SH, Ogilvie GI, and Miller KA (2003) Three-dimensional calculations of high-and low-mass planets embedded in protoplanetary discs. *Monthly Notices of the Royal Astronomical Society* 341: 213–229.

Belbruno E and Gott JR (2005) Where did the Moon come from? *Astronomical Journal* 129: 1724–1745.

Bell KR and Lin DNC (1994) Using Fu Orionis outbursts to constrain self-regulated protostellar disk models. *Astrophysical Journal* 474: 987–1004.

Benner LAM and McKinnon WB (1995) Orbital behavior of captured satellites: The effect of solar gravity on Triton's postcapture orbit. *Icarus* 114: 1–20.

Benz W, Slattery WL, and Cameron AGW (1986) The origin of the moon and the single-impact hypothesis I. *Icarus* 66: 515–535.

Benz W, Slattery WL, and Cameron AGW (1987) The origin of the moon and the single-impact hypothesis II. *Icarus* 71: 30–45.

Bills BG, Neumann GA, Smith DE, and Zuber MT (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004.

Borderies N, Goldreich P, and Tremaine S (1984) Excitation of inclinations in ring-satellite systems. *Astrophysical Journal* 284: 429–434.

Borderies N and Yoder CF (1990) Phobos' gravity field and its influenced on its orbit and physical librations. *Astronomy and Astrophysics* 233: 235–251.

Borg LE, Connelly JN, Boyer M, and Carlson RW (2011) Chronological evidence that the Moon is either young or did not have a global magma ocean. *Nature* 477: 70–74.

Author's personal copy

Boss AP and Peale SJ (1986) Dynamical constraints on the origin of the Moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 59–101. Houston, TX: Lunar and Planetary Institute.

Boué G and Laskar J (2010) A collisionless scenario for Uranus tilting. *Astrophysical Journal* 712: L44–L47.

Bryden G, Chen X, Lin DNC, Nelson RP, and Papaloizou JCB (1999) Tidally induced gap formation in protostellar disks: Gap clearing and suppression of protoplanetary growth. *Astrophysical Journal* 514: 344–367.

Bryden G, R'ozyczka M, Lin DNC, and Bodenheimer P (2000) On the interaction between protoplanets and protostellar disks. *Astrophysical Journal* 540: 1091–1101.

Buie MW, Grundy WM, Young EF, Young LA, and Stern SA (2006) Orbits and photometry of Pluto's satellites: Charon, S/2005 P1 and S/2005 P2. *Astronomical Journal* 132: 290–298.

Buie MW, Tholen DJ, and Grundy WM (2012a) The orbit of Charon is circular. *Astronomical Journal* 144, 19 pp.

Buie MW, Tholen DJ, and Grundy WM (2012b) Two body orbits for the satellites of Pluto. In: *Asteroids, Comets, Meteors*, Proceedings of the Conference held in Niigata, Japan, May 16–20. LPI Contribution 1667, id.6249.

Burns JA (1986) Some background on satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 1–38. Tucson, AZ: University of Arizon Press.

Burns JA (2003) Smashed into orbit, a review of *The Big Splat, or How Our Moon Came to Be*, by Dana MacKenzie. *Nature* 425: 664–665.

Cameron AGW and Ward WR (1976) The origin of the Moon (abstract). In: *Lunar Science VII*, pp. 120–122. Houston, TX: Lunar Science Institute.

Canup RM (2004a) Dynamics of lunar formation. *Annual Review of Astronomy and Astrophysics* 42: 441–475.

Canup RM (2004b) Simulations of a late lunar-forming impact. *Icarus* 168: 433–456.

Canup RM (2005) A giant impact origin of Pluto–Charon. *Science* 307: 546–550.

Canup RM (2008) Lunar-forming collisions with pre-impact rotation. *Icarus* 196: 518–538.

Canup RM (2010) Origin of Saturn's rings and inner moons by mass removal from a lost Titan-sized satellite. *Nature* 468: 943–946.

Canup RM (2011) On a giant impact origin of Charon, Nix, and Hydra. *Astronomical Journal* 141, 9 pp.

Canup RM (2012) Forming a Moon with an Earth-like composition via a giant impact. *Science* 338: 1052–1055.

Canup RM (2013) Modification of the rock content of the inner Saturnian satellites by an outer solar system LHB. *44th Lunar and Planetary Science Conference*. Contribution No. 1719, p. 2298.

Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.

Canup RM and Esposito LW (1996) Accretion of the Moon from an impact-generated disk. *Icarus* 119: 427–446.

Canup RM and Righter K (2000) *Origin of the Earth and Moon*. Tucson, AZ: University of Arizona Press.

Canup RM and Ward WR (2000) A possible impact origin of the Uranian satellite system. *Bulletin of the American Astronomical Society* 32: 1105.

Canup R and Ward WR (2001) Steady-state conditions in a Jovian protosatellite accretion disk. Paper presented at the *Jupiter Conference*, Boulder, CO, June 25–30.

Canup R and Ward WR (2002) Formation of the Galilean satellites: Conditions of accretion. *Astronomical Journal* 124: 3404–3423.

Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839.

Canup RM and Ward WR (2009) Origin of Europa and the Galilean satellites. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 59–83. Tucson, AZ: University of Arizona Press.

Canup RM, Levison HF, and Stewart GR (1999) Evolution of a terrestrial multiple-moon system. *Astronomical Journal* 117: 603–620.

Canup RM, Barr AC, and Crawford DA (2013) Lunar-forming impacts: High-resolution SPH and AMR-CTH simulations. *Icarus* 222: 200–219.

Carlson RW (1999) A tenuous carbon dioxide atmosphere on Jupiter's moon Callisto. *Science* 283: 820–821.

Carruba V, Burns JA, Nicholson PD, and Gladman BJ (2002) On the inclination distribution of the Jovian irregular satellites. *Icarus* 158: 434–449.

Carruba V, Nesvorný D, Burns JA, and Ćuk M (2003) Kozai resonators among distant moons of the Jovian planets. *Bulletin of the American Astronomical Society* 35: 1044.

Carruba V, Nesvorný D, Burns JA, Ćuk M, and Tsiganis K (2004) Chaos and the effects of planetary migration on the orbit of S/2000 S5 Kiviuq. *Astronomical Journal* 128: 1899–1915.

Cassen P and Moosman A (1981) On the formation of protostellar disks. *Icarus* 48: 353–376.

Castillo-Rogez JC, Efroimsky M, and Lainey V (2011) The tidal history of Iapetus: Spin dynamics in the light of a refined dissipation model. *Journal of Geophysical Research* 116: E09008, 29 pp.

Chambers JE (2001) Making more terrestrial planets. *Icarus* 152: 205–224.

Charnoz S, Crida A, Castillo-Rogez JC, et al. (2011) Accretion of Saturn's mid-sized moons during the viscous spreading of young massive rings: Solving the paradox of silicate-poor rings versus silicate-rich moons. *Icarus* 216: 535–550.

Charnoz S, Dones L, Esposito LW, Estrada PR, and Hedman MM (2009) Origin and evolution of Saturn's ring system. In: Dougherty MK, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini–Huygens*, pp. 537–575. The Netherlands: Springer.

Charnoz S, Salmon J, and Crida A (2010) Origin of Saturn's small moons and F ring: Recent accretion at the ring's outer edge. *Nature* 465: 752–754.

Cheng WH (2011) *Tidal Evolution of Pluto–Charon and the Implication for the Origin of the Satellites Nix and Hydra*. Masters Thesis, University of Hong Kong.

Cheng WH, Lee MH, and Peale SJ (2014a) Complete tidal evolution of Pluto–Charon. *Icarus* 233: 242–258.

Cheng WH, Peale SJ, and Lee MH (2014b) On the origin of Pluto's small satellites by resonant transport. *Icarus* (In press).

Colombo G and Franklin FA (1971) On the formation of the outer satellite groups of Jupiter. *Icarus* 15: 186–189.

Colwell JE and Esposito LW (1992) Origins of the rings of Uranus and Neptune. I – Statistics of satellite disruptions. *Journal of Geophysical Research* 97E: 10227–10241.

Conway BA (1982) On the history of the lunar orbit. *Icarus* 51: 610–622.

Costillo-Rogez JC, Efroimsky M, and Lainey V (2011) The tidal history of Iapetus: Spin dynamics in the light of a refined dissipation model. *Journal of Geophysical Research* 116: E09008, 29 pp.

Craddock RA (1994) The origin of Phobos and Deimos. In: *Lunar and Planetary Science Conference XXV*, pp. 293–294.

Craddock RA (2011) Are Phobos and Deimos the result of a giant impact? *Icarus* 211: 1150–1161.

Crida A and Charnoz S (2012) Formation of regular satellites from ancient massive rings in the solar system. *Science* 338: 1196–1199.

Ćuk M and Burns JA (2004) Gas-drag-assisted capture of Himalia's family. *Icarus* 167: 369–381.

Ćuk M and Gladman BJ (2005) Constraints on the orbital evolution of Triton. *Astrophysical Journal* 626: L113–L116.

Ćuk M and Stewart ST (2012) Making the Moon from a fast-spinning Earth: A giant impact followed by a resonant despinning. *Science* 338: 1047–1052.

Cuzzi JN, Burns JA, Charnoz S, et al. (2010) An evolving view of Saturn's dynamic rings. *Science* 327: 1470–1475.

Cuzzi JN, Hogan RC, Paque JM, and Dobrovolskis AR (2001) Size selective concentration of chondrules and other small particles in protoplanetary nebula turbulence. *Astrophysical Journal* 546: 496–508.

Cuzzi JN, Hogan RC, and Shariff K (2008) Toward planetesimals: Dense chondrule clumps in the protoplanetary nebula. *Astrophysical Journal* 687: 1432–1447.

d'Angelo G, Kley W, and Henning T (2003) Orbital migration and mass accretion of protoplanets in three-dimensional global computations with nested grids. *Astrophysical Journal* 586: 540–561.

Daisaka H, Tanaka H, and Ida S (2001) Viscosity in a dense planetary ring with self-gravitating particles. *Icarus* 154: 296–312.

Darwin GH (1879) On the precession of a viscous spheroid, and on the remote history of the Earth. *Philosophical Transactions of the Royal Society* 170: 447–530.

Darwin GH (1880) On the secular changes in the elements of the orbit of a satellite revolving about a tidally distorted planet. *Philosophical Transactions of the Royal Society* 171: 713–891.

Dahl TW and Stevenson DJ (2010) Turbulent mixing of metal and silicate during planet accretion—-And interpretation of the Hf-W chronometer. *Earth and Planetary Science Letters* 295: 177–186.

Dickey JO, Bender PL, Faller JE, et al. (1994) Lunar laser ranging: A continuing legacy of the Apollo Program. *Science* 265: 482–490.

Dobrovolskis AR (1982) Internal stresses in Phobos and other triaxial bodies. *Icarus* 52: 136–148.

Dones L (1991) A recent cometary origin for Saturn's rings? *Icarus* 92: 194–203.

Dones L and Tremaine S (1993) On the origin of planetary spins. *Icarus* 103: 67–92.

Duncan MJ, Levison HF, and Lee MH (1998) A multiple time step symplectic algorithm for integrating close encounters. *Astronomical Journal* 116: 2067–2077.

Duxbury TC and Callahan JD (1981) Pole and prime meridian expressions for Phobos and Deimos. *Astronomical Journal* 86: 1722–1727.

Duxbury TC and Callahan JD (1989) Phobos and Deimos control networks. *Icarus* 77: 275–286.

Duxbury TC and Veverka J (1978) Deimos encounter by Viking – Preliminary imaging results. *Science* 201: 812–814.

Author's personal copy

Dyczmons V (1978) Formation of planetary and satellite systems. *Astrophysics and Space Science* 58: 521–523.

Efroimsky M (2012) Tidal dissipation compared to seismic dissipation in small bodies, Earths and super-Earths. *Astrophysical Journal* 746, 20 pp.

Elkins-Tanton LT, Burgess S, and Yin Q-Z (2011) The lunar magma ocean: Reconciling the solidification process with lunar petrology. *Earth and Planetary Science Letters* 304: 326–336.

Esposito LW (1986) Structure and evolution of Saturn's rings. *Icarus* 67: 345–357.

Estrada PR and Mosqueira I (2006) A gas-poor planetesimal capture model for the formation of giant planet satellite systems. *Icarus* 181: 486–509.

Estrada PR, Mosqueira I, Lissauer JJ, D'Angelo G, and Cruikshank DP (2009) Formation of Jupiter and conditions of accretion of the Galilean satellites. In: Pappalardo RT, McKinnon WB, and Khurana K (eds.) *Europa*, pp. 27–58. Tucson, AZ: University of Arizona Press.

Farinella P, Milani A, Nobili AM, and Valsecchi GB (1980) Some remarks on the capture of Triton and the origin of Pluto. *Icarus* 44: 810–812.

Fleshman BL, Delamere PA, and Bagenal F (2010) The source of Saturn's extended neutral cloud. *AGU Fall Meeting*.

Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars Pathfinder. *Science* 278: 1749–1752.

Fortney JJ, Marley MS, and Barnes JW (2007) Planetary radii across five orders of magnitude in mass and stellar insolation. *Astrophysical Journal* 659: 1661–1672.

French RS and Showalter MR (2012) Cupid is doomed: An analysis of the stability of the inner Uranian satellites. *Icarus* 220: 911–921.

Fujii YI, Okuzumi S, and Inutsuka S (2011) A fast and accurate calculation scheme for ionization degrees in protoplanetary and circumplanetary disks with charged dust grains. *Astrophysical Journal* 743: 53–62.

Gao P and Stevenson DJ (2012) How does nonhydrostaticity affect the determination of icy satellites' moment of inertia? In: *43rd Lunar and Planetary Science Conference*, 1701.

Garrick-Bethell I, Wisdom J, and Zuber MT (2006) Evidence for a past high eccentricity lunar orbit. *Science* 313: 652–655.

George RB, Halliday AN, Schauble EA, and Reynolds BC (2007) Silicon in the Earth's core. *Nature* 447: 1102–1106.

Gerstenkorn H (1955) Uber Gezeitenreibuing beim Zweikorperproblem. *Zeitschrift für Astrophysik* 36: 245–274.

Gerstenkorn H (1969) The earliest past of the Earth-Moon system. *Icarus* 11: 189–207.

Gladman BJ, Kavelaars JJ, Holman M, et al. (2000) The discovery of Uranus XIX, XX, and XXI. *Icarus* 147: 320–324, erratum (2000) 148, 320.

Gladman B, Kavelaars JJ, Holman M, et al. (2001) Discovery of 12 satellites of Saturn exhibiting orbital clustering. *Nature* 412: 163–169.

Gladman BJ, Nicholson PD, Burns JA, et al. (1998) Discovery of two distant irregular moons of Uranus. *Nature* 392: 897–899.

Goldreich P (1963) On the eccentricity of satellite orbits in the solar system. *Monthly Notices of the Royal Astronomical Society* 126: 257–268.

Goldreich P (1965) Inclination of satellites orbits about an oblate precessing planet. *Astronomical Journal* 70: 5–9.

Goldreich P (1966) History of the lunar orbit. *Reviews of Geophysics* 4: 411–439.

Goldreich P, Lithwick Y, and Sari R (2002) Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* 420: 643–646.

Goldreich P, Murray N, Longaretti PY, and Banfield D (1989) Neptune's story. *Science* 245: 500–504.

Goldreich P and Soter S (1966) Q in the solar system. *Icarus* 5: 375–389.

Goldreich P and Tremaine S (1979) Precession of the epsilon ring of Uranus. *Astronomical Journal* 84: 1638–1641.

Goldreich P and Tremaine S (1982) The dynamics of planetary rings. *Annual Review of Astronomy and Astrophysics* 20: 249–283.

Goldreich P and Ward WR (1973) The formation of planetesimals. *Astrophysical Journal* 183: 1051–1062.

Gomes RS, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic late heavy bombardment period of the terrestrial planets. *Nature* 435: 466–469.

Greenzweig Y and Lissauer JJ (1990) Accretion rates of protoplanets. *Icarus* 87: 40–77.

Greenzweig Y and Lissauer JJ (1992) Accretion rates of protoplanets. II – Gaussian distributions of planetesimal velocities. *Icarus* 100: 440–463.

Halliday AH (2004) Mixing, volatile loss and compositional change during impact driven accretion of the Earth. *Nature* 427: 505–509.

Hamilton DP and Burns JA (1991) Orbital stability zones about asteroids. *Icarus* 92: 118–131.

Hamilton DP and Krivov AV (1997) Dynamics of distant moons of asteroids. *Icarus* 128: 241–249.

Harris AW (1977) The effect of tidal friction on the origin and thermal evolution of the Moon, (abstract). In: *Lunar Science VIII*, pp. 401–402. Houston, TX: Lunar Science Institute.

Harris AW (1978) Satellite formation II. *Icarus* 34: 128–145.

Harris AW and Kaula WM (1975) A co-accretion model of satellite formation. *Icarus* 24: 516–524.

Hartmann WK and Davis DR (1975) Satellite-sized planetesimals and lunar origin. *Icarus* 24: 504–515.

Hartmann WK, Phillips RJ, and Taylor GJ (1986) *Origin of the Moon*. Houston, TX: Lunar and Planetary Institute.

Heppenheimer TA and Porco CC (1977) New contributions to the problem of capture. *Icarus* 30: 385–401.

Hibbitts CA, Klemaszewski JE, McCord TB, Hansen GB, and Greeley R (2002) CO_2-rich impact craters on Callisto. *Journal of Geophysical Research (Planets)* 107: 14.1–14.12.

Hibbitts CA, McCord TB, and Hansen GB (2000) Distributions of CO_2 and SO_2 on the surface of Callisto. *Journal of Geophysical Research* 430: 22541–22558.

Hibbitts CA, Pappalardo RT, Hansen GB, and McCord TB (2003) Carbon dioxide on Ganymede. *Journal of Geophysical Research (Planets)* 108: 2.1–2.22.

Hollenbach DJ, York HW, and Johnstone D (2000) Disk dispersal around young stars. In: Manning V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, p. 401. Tucson, AZ: University of Arizona Press.

Holman MJ, Kavelaars JJ, Grav T, et al. (2004) Discovery of five irregular moons of Neptune. *Nature* 430: 865–867.

Horedt GP (1976) Pre-capture orbits of the Moon. *Moon* 15: 439–443.

Hussmann H and Spohn T (2004) Thermal-orbital evolution of Io and Europa. *Icarus* 171: 391–410.

Ida S, Canup RM, and Stewart GR (1997) Lunar accretion from an impact generated disk. *Nature* 389: 353–357.

Ida S and Lin DNC (2004) Toward a deterministic model of planetary formation. I. A desert in the mass and semimajor axis distributions of extrasolar planets. *Astrophysical Journal* 604: 388–413.

Ida S and Lin DNC (2008) Toward a deterministic model of planetary formation. IV. Effects of Type I migration. *Astrophysical Journal* 673: 487–501.

Iess L, Rappaport NJ, Jacobson RA, et al. (2010) Gravity science at Titan, *EGU General Assembly,* Vienna, Austria, 2-7 May, P 11037.

Imaeda Y and Inutsuka S (2002) Shear flows in smoothed particle hydrodynamics. *Astrophysical Journal* 569: 501–518.

Jacobsen SB (2005) The Hf-W isotopic system and the origin of the Earth and Moon. *Annual Review of Earth and Planetary Sciences* 33: 531–570.

Jacobson RA (2010) The orbits and masses of the Martian satellites and the libration of Phobos. *Astronomical Journal* 139: 668–679.

Jaumann R, Clark RN, Nimmo F, et al. (2009) Icy satellites: Geological evolution and surface processes. In: Dougherty MK, Esposito LW, and Krimigis SM (eds.) *Saturn from Cassini–Huygens*, pp. 637–681. Dordrecht: Springer.

Jewitt D and Sheppard S (2005) Irregular satellites in the context of planet formation. *Space Science Reviews* 116: 441–455.

Johansen A, Klahr H, and Henning T (2006) Gravoturbulent formation of planetesimals. *Astrophysical Journal* 636: 1121–1134.

Johansen A, Klahr H, and Henning T (2011) High resolution simulations of planetesimal formation in turbulent protoplanetary discs. *Astronomy and Astrophysics* 529, A62, 16 pp.

Johansen A, Oishi J, Mac Low M-M, Klahr H, Henning T, and Youdin A (2007) Rapid planetesimal formation in turbulent circumstellar disks. *Nature* 448: 1022–1025.

Johansen A and Youdin A (2007) Protoplanetary disk turbulence driven by the streaming instability: Nonlinear saturation and particle concentration. *Astrophysical Journal* 662: 627–641.

Johansen A, Youdin AN, and Lithwick Y (2012) Adding particle collisions to the formation of asteroids and Kuiper belt objects via streaming instabilities. *Astronomy and Astrophysics* 537, A125, 17 pp.

Johansen A, Youdin A, and Mac Low M-M (2009) Particle clumping and planetesimal formation depend strongly on metallicity. *Astrophysical Journal* 704: L75–L79.

Jones JH and Palme H (2000) Geochemical constraints on the origin of the Moon. In: Canup RM and Richter J (eds.) *Origin of the Earth and Moon*, pp. 197–216. Tucson, AZ: University of Arizona Press.

Karato S-I (2008) *Deformation of Earth Materials: An Introduction to the Rheology of Solid Earth*. Cambridge: Cambridge University Press.

Kary DM and Lissauer JJ (1995) Nebular gas drag and planetary accretion. II. Planet on an eccentric orbit. *Icarus* 117: 1–24.

Kary DM, Lissauer JJ, and Greenzweig Y (1993) Nebular gas drag and planetary accretion. *Icarus* 106: 288–307.

Kaula WM (1964) Tidal dissipation by tidal friction and the resulting orbital evolution. *Reviews of Geophysics* 2: 661–685.

Kivelson MG, Khurana KK, Coroniti FV, et al. (1997) Magnetic field and magnetosphere of Ganymede. *Geophysical Research Letters* 24: 2155–2158.

Author's personal copy

Kivelson MG, Khurana KK, Stevenson DJ, et al. (1999) Europa and Callisto: Induced or intrinsic fields in a periodically varying plasma environment. *Journal of Geophysical Research* 104: 4609–4626.
Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.
Kleine T, Mezger KK, Palme H, and Münker C (2004) The W isotope evolution of the bulk silicate Earth: Constraints on the timing and mechanisms of core formation and accretion. *Earth and Planetary Science Letters* 228: 109–123.
Kleine T, Palme H, Mezger K, and Halliday AN (2005) Hf-W chronometry of lunar metals and the age and early differentiation of the Moon. *Science* 310: 1671–1674.
Kley W (1999) Mass flow and accretion through gaps in accretion discs. *Monthly Notices of the Royal Astronomical Society* 303: 696–710.
Kley W and Crida A (2008) Migration of protoplanets in radiative discs. *Astronomy and Astrophysics* 487: L9–L12.
Kley W, D'Angelo G, and Henning T (2001) Three-dimensional simulations of a planet embedded in a protoplanetary disk. *Astrophysical Journal* 547: 457–464.
Kokubo E, Ida S, and Makino J (2000) Evolution of a circumterrestrial disk and formation of a single Moon. *Icarus* 148: 419–436.
Kolyuka YF, Efimov AE, Kudryavtsev SM, Margorin OK, Tarasov VP, and Tikhonov VF (1990) Refinement of the gravitational constant of Phobos from Phobos-2 tracking data. *Soviet Astronomy Letters* 16: 168–170.
König S, Münker C, Hohl S, et al. (2011) The Earth's tungsten budget during mantle melting and crust formation. *Geochimica et Cosmochimica Acta* 75: 2119–2136.
Kuskov OL and Kronrod VA (2005) Internal structure of Europa and Callisto. *Icarus* 177: 550–569.
Lainey V, Arlot J-E, Karatekin Ö, and van Hoolst T (2009) Strong tidal dissipation in Io and Jupiter from astrometric observations. *Nature* 459: 957–959.
Lainey V, Karatekin Ö, Desmars J, et al. (2012) Strong tidal dissipation in Saturn and constraints on Enceladus' thermal state from astrometry. *Astrophysical Journal* 752: 14L.
Lambeck K (1980) *The Earth's Variable Rotation: Geophysical Causes and Consequences*. New York: Cambridge University Press.
Lambrechts M and Johansen A (2012) Rapid growth of gas-giant cores by pebble accretion. *Astronomy and Astrophysics* 544, A32, 13 pp.
Laskar J and Robutel P (1993) The chaotic obliquity of the planets. *Nature* 361: 608–612.
Lee MH and Peale SJ (2000) Making Hyperion. *Bulletin of the American Astronomical Society* 32: 1078.
Lee MH and Peale SJ (2006) On the orbits and masses of the satellites of the Pluto–Charon system. *Icarus* 184: 573–583.
Lin DCN and Papaloizou J (1985) On the dynamical origin of the solar system. In: Black D and Matthews MS (eds.) *Protostars and Planets II*, pp. 981–1072. Tucson, AZ: University of Arizona Press.
Lissauer JJ (1987) Timescales for planetary accretion and structure of the protoplanetary disk. *Icarus* 69: 249–265.
Lissauer JJ (1995) Urey Prize lecture: On the diversity of plausible planetary systems. *Icarus* 114: 217–236.
Lissauer JJ and Kary DM (1991) The origin of the systematic component of planetary rotation: 1. Planet on a circular orbit. Icarus 94: 126–159.
Lithwick Y and Wu Y (2008) On the origin of Pluto's minor moons Nix and Hydra. Astrophysics ArXiv:0802.2951v1.
Love AEH (1944) *A Treatise on the Mathematical Theory of Elasticity*. New York: Dover.
Lubow SH and Martin RG (2013) Dead zones in circumplanetary disks as formation sites for regular satellites. *Monthly Notices of the Royal Astronomical Society* 428: 2668–2673.
Lubow SH, Seibert M, and Artymowicz P (1999) Disk accretion onto high-mass planets. *Astrophysical Journal* 526: 1001–1012.
Lugmair GW and Shukolyukov A (1998) Early solar system timescales according to ^{53}Mn-^{53}Cr systematics. *Geochimica et Cosmochimica Acta* 62: 2863–2886.
Lunine JI and Nolan MC (1992) A massive early atmosphere on Triton. *Icarus* 100: 221–234.
Lunine J and Stevenson DJ (1982) Formation of the Galilean satellites in a gaseous nebula. *Icarus* 52: 14–39.
Lynden-Bell D and Pringle JE (1974) The evolution of viscous disks and the origin of the nebular variables. *Monthly Notices of the Royal Astronomical Society* 168: 603–637.
Lyttleton RA (1967) Dynamical capture of the Moon by the Earth. *Proceedings of the Royal Society of London* A296: 285–292.
Machida MN (2009) Thermal effects of circumplanetary disc formation around proto-gas giant planets. *Monthly Notices of the Royal Astronomical Society* 392: 514–524.
Machida MN, Kokubo E, Inutsuka S-I, and Matsumoto T (2008) Angular momentum accretion onto a gas giant plant. *Astrophysical Journal* 685: 1220–1236.
Machida MN, Kokubo E, Inutsuka S-I, and Matsumoto T (2010) Gas accretion onto a protoplanet and formation of a gas giant planet. *Monthly Notices of the Royal Astronomical Society* 405: 1227–1243.
MacKenzie D (2003) *The Big Splat, or How Our Moon Came to Be*. Hoboken, NJ: Wiley.
Margot J-L, Nolan MC, Benner LAM, et al. (2002) Binary asteroids in the near-Earth object population. *Science* 296: 1445–1448.
Marley MS, Fortney JJ, Hubickyj O, Bodenheimer P, and Lissauer JJ (2007) On the luminosity of young Jupiters. *Astrophysical Journal* 655: 541–549.
Martin RG and Lubow SH (2011) Tidal truncation of circumplanetary disks. *Monthly Notices of the Royal Astronomical Society* 413: 1447–1461.
Massett FS, Morbidelli A, Crida A, and Ferreira J (2006) Disk surface density transitions as protoplanet traps. *Astrophysical Journal* 642: 478–487.
McCord TB, Hansen GB, Clark RN, et al. (1998) Non-water-ice constituents in the surface material of the icy Galilean satellites from the Galileo near-infrared mapping spectrometer investigation. *Journal of Geophysical Research* 103: 8603–8626.
McKinnon WB (1984) On the origin of Triton and Pluto. *Nature* 311: 355–358.
McKinnon WB (1989) On the origin of the Pluto–Charon binary. *Astrophysical Journal* 344: L41–L44.
McKinnon WB (1997) Mystery of Callisto: Is it undifferentiated? *Icarus* 130: 540–543.
McKinnon WB (2006) On convection in ice I shells of outer Solar System bodies, with detailed application to Callisto. *Icarus* 183: 435–450.
McKinnon WB and Leith AC (1995) Gas drag and the orbital evolution of a captured Triton. *Icarus* 118: 392–413.
McKinnon WB, Lunine JI, and Banfield D (1995) Origin and evolution of Triton. In: Cruikshank DP (ed.) *Neptune and Pluto*, pp. 807–877. Tucson, AZ: University of Arizona Press.
Melosh HJ (2007) A hydrocode equation of state for SiO_2. *Meteoritics & Planetary Science* 42: 2079–2098.
Melosh HJ (2009) An isotopic crisis for the giant impact origin of the Moon? *Meteoritics and Planetary Science Supplement*. 72nd Meteoritical Society Meeting, July 13-18, 2009 in Nancy, France; p 5104.
Merline WJ, Weidenschilling SJ, Durda DD, Margot J-L, Pravec P, and Storrs AD (2002) Asteroids do have satellites. In: Bottke WF Jr., Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 289–312. Tucson, AZ: University of Arizona Press.
Mignard F (1979) The evolution of the lunar or orbit revisited, I. *Moon and Planets* 20: 301–315.
Mignard F (1981a) The lunar orbit revisited, III. *Moon and Planets* 24: 189–207.
Mignard F (1981b) Evolution of the Martian satellites. *Monthly Notices of the Royal Astronomical Society* 194: 365–379.
Mitchell TR and Stewart GR (2011) Photoevaporation as a truncation mechanism for circumplanetary disks. *Astronomical Journal* 142: 168–178.
Morbidelli A and Crida A (2007) The dynamics of Jupiter and Saturn in the gaseous protoplanetary disk. *Icarus* 191: 158–171.
Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435: 462–465.
Morbidelli A, Tsiganis K, Batygin K, Crida A, and Gomes R (2012) Explaining why the Uranian satellites have equatorial prograde orbit despite the large planetary obliquity. *Icarus* 219: 737–740.
Mosqueira I and Estrada PR (2003a) Formation of the regular satellites of giant planets in an extended gaseous nebula I: Subnebula model and accretion of satellites. *Icarus* 163: 198–231.
Mosqueira I and Estrada PR (2003b) Formation of the regular satellites of giant planets in an extended gaseous nebula II: Satellite migration and survival. *Icarus* 163: 232–255.
Mosqueira I and Estrada PR (2011) On the origin of the Saturnian Moon-Ring system, *42nd Lunar and Planetary Science Conference*. Contribution No. 1608, p2151.
Mosqueira I, Estrada P, and Turrini D (2010) Planetesimals and satellitesimals: Formation of the satellite systems. *Space Science Reviews* 153: 431–446.
Mousis O and Alibert Y (2006) Modeling the Jovian subnebula: II. Composition of regular satellite ices. *Astronomy and Astrophysics* 448: 771–778.
Mousis O, Pargamin J, Grasset O, and Sotin C (2002) Experiments in the NH_3-H_2O system in the [0, 1 GPa] pressure range – Implications for the deep liquid layer of large icy satellites. *Geophysical Research Letters* 29(45): 1–4.
Murray CD and Dermott SF (1999) *Solar System Dynamics*. Cambridge: Cambridge University Press.
Nagel K, Breuer D, and Spohn T (2004) A model for the interior structure, evolution, and differentiation of Callisto. *Icarus* 169: 402–412.
Nakajima M and Stevenson DJ (2013) Thermodynamic processes during the Moon-forming impact. In: *44th Lunar and Planetary Conference*, 2680.pdf.
Nakamura Y and Koyama J (1982) Seismic Q of the lunar upper mantle. *Journal of Geophysical Research* 87: 4855–4861.

Author's personal copy

Nesvorny D, Vokrouhlicky D, and Mobidelli A (2007) Capture of irregular satellites during planetary encounters. *Bulletin of the American Astronomical Society* 39: 475.
Nesvorny D, Youdin AN, and Richardson DC (2010) Formation of Kuiper Belt binaries by gravitational collapse. *Astronomical Journal* 140: 785–793.
Niemann HB, Atreya SK, Bauer SJ, et al. (2005) The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* 438: 779–784.
Niemann HB, Atreya SK, Demick JE, et al. (2010) Composition of Titan's lower atmosphere and simple surface volatiles as measured by the Cassini–Huygens probe gas chromatograph mass spectrometer experiment. *Journal of Geophysical Research* 115: E12006.
Nogueira E, Brasser R, and Gomes R (2011) Reassessing the origin of Triton. *Icarus* 214: 113–130.
O'Rourke JG and Stevenson DJ (2013) Stability of ice/rock mixtures with application to a partially differentiated Titan. *Icarus* 227: 67–77.
Ogihara M, Duncan MJ, and Ida S (2010) Eccentricity trap: Trapping of resonantly interacting planets near the disk inner edge. *Astrophysical Journal* 721: 1184–1192.
Ogihara M and Ida S (2012) N-body simulations of satellite formation around giant planets: Origin of orbital configuration of the Galilean moons. *Astrophysical Journal* 753: 60–77.
Öpik EJ (1969) The Moon's surface. *Annual Review of Astronomy and Astrophysics* 7: 473–526.
Öpik EJ (1971) Comments on lunar origin. *Irish Astronomical Journal* 10: 190–238.
Owen T and Niemann HB (2009) The origin of Titan's atmosphere: Some recent advances. *Philosophical Transactions of the Royal Society A* 367: 607–615.
Paardekooper S-J, Baruteau C, Crida A, and Kley W (2010) A torque formula for non-isothermal type I planetary migration –I. Unsaturated horseshoe drag. *Monthly Notices of the Royal Astronomical Society* 401: 1950–1964.
Paardekooper S-J and Mellema G (2006a) Dust flow in gas disks in the presence of embedded planets. *Astronomy and Astrophysics* 453: 1129–1140.
Paardekooper S-J and Mellema G (2006b) Halting type I planet migration in non-isothermal disks. *Astronomy and Astrophysics* 459: L17–L20.
Pahlevan K and Stevenson DJ (2007) Equilibrium in the aftermath of the lunar-forming impact. *Earth and Planetary Science Letters* 262: 438–449.
Pang KD, Pollack JB, Veverka J, Lane AL, and Ajello JM (1978) The composition of Phobos – Evidence for carbonaceous chondrite surface from spectral analysis. *Science* 199: 64–66.
Papaloizou JCB and Larwood JD (2000) On the orbital evolution and growth of protoplanets embedded in a gaseous disc. *Monthly Notices of the Royal Astronomical Society* 315: 823–833.
Papaloizou JCB and Nelson RP (2005) Models of accreting gas giant protoplanets in protostellar disks. *Astronomy and Astrophysics* 433: 247–265.
Papaloizou JCB and Terquem C (1999) Critical protoplanetary core masses in protoplanetary disks and the formation of short-period giant planets. *Astrophysical Journal* 521: 823–838.
Peale SJ (1977) Rotation histories of the natural satellites. In: *Planetary Satellites*, pp. 87–111. Tucson, AZ: University of Arizona Press.
Peale SJ (1986) Orbital resonances, unusual configurations and exotic rotation states among planetary satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 159–223. Tucson, AZ: University of Arizona Press.
Peale SJ (1988) Speculative histories of the Uranian satellite system. *Icarus* 74: 153–171.
Peale SJ (1993) The effect of the nebula on the Trojan precursors. *Icarus* 106: 308–322.
Peale SJ (1999) Origin and evolution of the natural satellites. *Annual Review of Astronomy and Astrophysics* 37: 533–602.
Peale SJ (2003) Tidally induced volcanism. *Celestial Mechanics and Dynamical Astronomy* 87: 129–155.
Peale SJ and Cassen PM (1978) Contribution of tidal heating to lunar thermal history. *Icarus* 36: 245–269.
Peale SJ, Cassen PM, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894.
Peale SJ and Lee MH (2002) A primordial origin of the Laplace relation among the Galilean satellites. *Science* 298: 593–597.
Pires dos Santos PM, Giuliatti-Winter SM, and Sfair R (2011) Gravitational effects of Nix and Hydra in the external region of the Pluto–Charon system. *Monthly Notices of the Royal Astronomical Society* 410: 273–279.
Pires dos Santos PM, Morbidelli A, and Nesvorny D (2012) Dynamical capture in the Pluto–Charon system. *Celestial Mechanics and Dynamical Astronomy* 114: 341–352.
Plummer HC (1960) *Introductory Treatise on Dynamical Astronomy*. New York: Dover, Reprinted from the Cambridge University Press edition (1918).
Pollack JB and Burns JA (1977) An origin by capture for the Martian satellites? *Bulletin of the American Astronomical Society* 9: 518.
Pollack JB, Burns JA, and Tauber ME (1979) Tauber gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* 37: 587–611.
Pollack JB, Hollenbach D, Beckwith S, Simonelli DP, Roush T, and Fong W (1994) Composition and radiative properties of grains in molecular clouds and accretion disks. *Astrophysical Journal* 421: 615–639.
Pollack JB, Veverka J, Pang KD, Colburn DS, Lane AL, and Ajello JM (1978) Multicolor observations of Phobos with the Viking lander cameras – Evidence for a carbonaceous chondritic composition. *Science* 199: 66–69.
Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active south pole of Enceladus. *Science* 311: 1393–1401.
Prandtl L (1952) *Essentials of Fluid Dynamics*. New York: Hafner, pp. 174–196.
Pritchard ME and Stevenson DJ (2000) Thermal aspects of a lunar origin by giant impact. In: Canup RM and Richter K (eds.) *Origin of the Earth and Moon*, pp. 179–196. Tucson, AZ: University of Arizon Press.
Ray R, Eanes RJ, and Lemoine FG (2001) Constraints on energy dissipation in the Earth's body tide from satellite tracking and altimetry. *Geophysical Journal International* 144: 471–480.
Reufer A, Meier MMM, Benz W, and Wieler R (2012) A hit and run giant impact scenario. *Icarus* 221: 296–299.
Ringwood AE (1970) Origin of the Moon: The precipitation hypothesis. *Earth and Planetary Science Letters* 8: 131–140.
Ringwood AE (1972) Some comparative aspects of lunar origin. *Physics of the Earth and Planetary Interiors* 6: 366–376.
Rosenblatt P (2011) The origin of the Martian moons revisited. *Astronomy and Astrophysics* 19: 44.
Rosenblatt P and Charnoz S (2012) On the formation of the Martian moons from a circum-Mars accretion disk. *Icarus* 221: 806–815.
Rubie DC, Melosh HJ, Reid JE, Liebske C, and Righter K (2003) Mechanisms of metal-silicate equilibration in the terrestrial magma ocean. *Earth and Planetary Science Letters* 205: 239–255.
Ruden SP and Pollack JB (1991) The dynamical evolution of the protosolar nebula. *Astrophysical Journal* 375: 740–760.
Ruskol EL (1961) The origin of the Moon I. *Soviet Astronomy-AJ* 4: 657–668.
Ruskol EL (1963) The origin of the Moon II. *Soviet Astronomy-AJ* 7: 221–227.
Ruskol EL (1972a) The origin of the Moon. III. Some aspects of the dynamics of the circumterrestrial swarm. *Soviet Astronomy-AJ* 15: 646–654.
Ruskol EL (1972b) Possible differences in the chemical composition of the Earth and Moon, for a Moon formed in the circumterrestrial swarm. *Soviet Astronomy-AJ* 15: 1061–1063.
Ruskol EL (1972c) On the initial distance of the Moon forming in the circumterrestrial swarm. *Moon* 5: 402–404.
Ruskol EL (1975) The origin of the Moon. In: *Proceedings of the Soviet-American Conference on Cosmo-chemistry of the Moon and Planets*, pp. 638–644. Moscow: Nauka (translated in English in NASA SP-370 (1977), pp. 815–822).
Safronov VS (1969) *Evolution of the Protoplanetary Cloud and the Formation of the Earth and Planets*. Moscow: Nauka Press, NASA TTF-677, 1972.
Safronov VS, Pechernikova GV, Ruskol EL, and Vitiazev AV (1986) Protosatellite swarms. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 89–116. Tucson, AZ: University of Arizona Press.
Saha P and Tremaine S (1993) The orbits of the retrograde Jovian satellites. *Icarus* 106: 549–562.
Salmon J and Canup RM (2012) Lunar accretion from a Roche-interior fluid disk. *Astrophysical Journal* 760: 18.
Salmon J, Charnoz S, Crida A, and Brahic A (2010) Long-term and large-scale viscous evolution of dense planetary rings. *Icarus* 209: 771–785.
Sasaki T, Stewart GR, and Ida S (2010) Origin of the different architectures of the Jovian and Saturnian satellite systems. *Astrophysical Journal* 714: 1052–1064.
Schubert G, Spohn T, and Reynolds RT (1986) Thermal histories compositions and internal structures of the moons in the solar system. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 224–292. Tucson, AZ: University of Arizona Press.
Schwarzschild M (1958) *Structure and Evolution of the Stars*. Princeton, NJ: Princeton University Press, Dover Publications, New York (1965), p. 62.
Sears FW and Salinger GL (1975) *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*. Reading, MA: Addison-Wesley, p. 286.
Sekine Y and Genda H (2012) Giant impacts in the Saturnian system: A possible origin of diversity in the inner mid-sized satellites. *Planetary and Space Science* 63–64: 133–138.
Shakura NI and Sunyaev RA (1973) Black holes in binary systems. Observational appearance. *Astronomy and Astrophysics* 24: 337–355.
Shapiro SL and Teukolsky SA (1983) *Black Holes, White Dwarfs, and Neutron Stars*. New York: Wiley, ch. 14.
Shenk P and Jackson MP (1993) Diapirism on Triton – A record of crustal layering and instability. *Geology* 21: 299–302.

Author's personal copy

Shito A, Karato S, and Park J (2004) Frequency dependence of Q in Earth's upper mantle inferred from continuous spectra of body waves. *Geophysical Research Letters* 31: L12603.
Shock EL and McKinnon WB (1993) Hydrothermal processing of cometary volatiles – Applications to Triton. *Icarus* 106: 464–477.
Shor VA (1975) The motion of the Martian satellites. *Celestial Mechanics* 12: 61–75.
Showalter MR, Hamilton DP, Stern SA, Weaver HA, Steffl AJ, and Young LA (2011) New satellite of (134340) Pluto: S/2011 (134340) 1, IAU Circulars, 9221, 1 (2011), Green DWE (ed.).
Showalter MR, Weaver HA, Stern SA, et al. (2012) New Satellite of (134340) Pluto S/2012 (134340)1. *International Astronomical Union*. Circular 9253, 1.
Shu F (1984) Waves in planetary rings. In: Greenberg R and Brahic A (eds.) *Planetary Rings*, pp. 513–561. Tucson, AZ: University of Arizona Press.
Shu F, Johnston D, and Hollenbach D (1993) Photoevaporation of the solar nebula and the formation of the giant planets. *Icarus* 106: 92–101.
Schubert G, Anderson JD, Spohn T, and McKinnon WB (2004) Interior composition, structure, and dynamics of the Galilean satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, vol. 1. New York: Cambridge University Press.
Shultz PH and Lutz-Garihan AB (1982) Forgotten satellites of Mars: A possible record from oblique angle impacts. In: *Lunar and Planetary Science XIII*, pp. 698–699 (abstract).
Sicardy B, Bellucci A, Gendron E, et al. (2006) Charon's size and an upper limit on its atmosphere from a stellar occultation. *Nature* 439: 52–54.
Sinclair AT (1978) The orbits of the satellites of Mars. *Vistas in Astronomy* 22: 133–137.
Singer SF (1968) The origin of the Moon and geophysical consequences. *Geophysical Journal of the Royal Astronomical Society* 15: 205–226.
Singer SF (1970) Origin of the Moon by capture and its consequences. *EOS, Transactions American Geophysical Union* 51: 637–641.
Singer SF (1972) Origin of the Moon by tidal capture and some geophysical consequences. *Moon* 5: 206–209.
Singer SF and Bandermann LW (1970) Where was the Moon formed? *Science* 170: 438–439.
Slattery WL, Benz W, and Cameron AGW (1992) Giant impacts on a primitive Uranus. *Icarus* 99: 167–174.
Smith JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.
Smith DE, Lemoine FG, and Zuber MT (1995) Simultaneous estimations of the masses of Mars, Phobos, and Deimos from distant spacecraft encounters. *Geophysical Research Letters* 22: 2171–2174.
Smith BA, Soderblom LA, Banfield D, et al. (1989) Voyager 2 at Neptune – Imaging science results. *Science* 246: 1422–1449.
Sohl F, Spohn T, Breuer D, and Nagel K (2002) Implications from Galileo observations on the interior structure and chemistry of the Galilean satellites. *Icarus* 157: 104–119.
Solomon SC (1977) The relationship between crustal tectonics and internal evolution in the Moon and Mercury. *Physics of the Earth and Planetary Interiors* 15: 135–145.
Solomon SC (1986) On the early thermal state of the Moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 435–452. Houston, TX: Lunar and Planetary Institute.
Soter S and Harris AW (1977) The equilibrium figures of Phobos and other small bodies. *Icarus* 30: 192–199.
Spohn T and Schubert G (2003) Oceans in the icy Galilean satellites of Jupiter? *Icarus* 161: 456–467.
Stephens DC and Noll KS (2006) Detection of six trans-neptunian binaries with NICMOS: A high fraction of binaries in the cold classical disk. *Astronomical Journal* 131: 1142–1148.
Stern SA, Bottke WF, and Levison HF (2003) Regarding the putative eccentricity of Charon's orbit. *Astronomical Journal* 125: 902–905.
Stern SA and McKinnon WB (2000) Triton's surface age and impactor population revisited in light of Kuiper Belt fluxes: Evidence for small Kuiper Belt objects and recent geological activity. *Astronomical Journal* 119: 945–952.
Stern SA, Parker JW, Duncan MJ, Snowdall JC Jr., and Levison HF (1994) Dynamical and observational constraints on satellites in the inner Pluto-Charon system. *Icarus* 108: 234–242.
Stern SA, McKinnon WB, and Lunine JI (1997) On the origin of Pluto, Charon, and the Pluto–Charon binary. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 605–663. Tucson, AZ: University of Arizona Press.
Stern SA, Weaver HA, Steffl AJ, et al. (2006) A giant impact origin for Pluto's small moons and satellite multiplicity in the Kuiper belt. *Nature* 439: 946–948.
Stevenson DJ (1984) The Uranus-Neptune dichotomy: The roll of giant impacts. In: *Lunar and Planetary Science XVII*, pp. 1011–1012.
Stevenson DJ (2001a) Origin of the Jovian Satellites. Paper presented at the *Jupiter Conference*, Boulder, CO, June 25–30.
Stevenson DJ (2001b) Jupiter and its moons. *Science* 294: 71–72.
Stevenson DJ, Harris AW, and Lunine JI (1986) Origin of satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 39–88. Tucson, AZ: University of Arizona Press.
Stone JM, Gammie CF, Balbus SA, and Hawley JF (2000) Transport processes in protostellar disks. In: Mannings V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, pp. 589–601. Tucson, AZ: University of Arizona Press.
Szebehely V and Evans RT (1980) On the capture of the Moon. *Celestial Mechanics* 21: 259–264.
Szeto AMK (1983) Orbital evolution and origin of the Martian satellites. *Icarus* 55: 133–168.
Takata T and Stevenson DJ (1996) Despin mechanism for protogiant planets and ionization state of protogiant planetary disks. *Icarus* 123: 404–421.
Takeda T and Ida S (2001) Angular momentum transfer in a protolunar disk. *Astrophysical Journal* 560: 514–533.
Tanaka H, Takeuchi T, and Ward WR (2002) Three-dimensional interactions between a planet and an isothermal disk. I. Corotation and Lindblad torques and planet migration. *Astrophysical Journal* 565: 1257–1274.
Tanaka H and Ward WR (2004) Three-dimensional interaction between a planet and an isothermal gaseous disk. II. Eccentricity waves and bending waves. *Astrophysical Journal* 602: 388–395.
Tanigawa T, Ohtsuki K, and Machida MN (2012) Distribution of accreting gas and angular momentum onto circumplanetary disks. *Astrophysical Journal* 747: 47–63.
Taylor SR and Jakes P (1974) The geochemical evolution of the moon. In: Proceedings of the 5th Lunar Science Conference, pp. 1287–1305. Houston, TX: Lunar Science Institute.
Taylor DJ, McKeegan KD, and Harrison TM (2009) Lu-Hf zircon evidence for rapid lunar differentiation. *Earth and Planetary Science Letters* 279: 157–164.
Testi L, Natta A, Shepherd DS, and Wilner DJ (2003) Large grains in the disk of CQ Tau. *Astronomy and Astrophysics* 403: 323–338.
Tholen DJ, Buie MW, and Grundy WM (2012) *Asteroids, Comets, Meteors*, Proceedings of the conference held 16–20 May 2012 in Niigata, Japan. LPI Contribution 1667, id.6327.
Thomas PC (1993) Gravity, tides, and topography on small satellites and asteroids – Application to surface features of the Martian satellites. *Icarus* 105: 326–344.
Thomas PC (2010) Sizes, shapes, and derived properties of the Saturnian satellites after the Cassini nominal mission. *Icarus* 208: 395–401.
Thompson C and Stevenson DJ (1988) Gravitational instability in two-phase disks and the origin of the moon. *Astrophysical Journal* 333: 452–481.
Tolson RH, Duxbury TC, Born GH, et al. (1978) Viking first encounter of Phobos – Preliminary results. *Science* 199: 61–64.
Touboul M, Klien T, Bourdon B, Palme H, and Weiler R (2007) Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature* 450: 1206–1209.
Touboul M, Puchtel IS, and Walker RJ (2012) ^{182}W evidence for long-term preservation of early mantle differentiation products. *Science* 335: 1065–1069.
Touma J and Wisdom J (1993) The chaotic obliquity of Mars. *Science* 259: 1294–1297.
Touma J and Wisdom J (1994) Evolution of the Earth-Moon system. *Astronomical Journal* 108: 1943–1961.
Touma J and Wisdom J (1998) Resonances in the early evolution of the Earth-Moon system. *Astronomical Journal* 115: 1653–1663.
Tsiganis K, Gomes R, Morbidelli A, and Levison H (2005) Origin of the orbital architecture of the giant planets of the solar system. *Nature* 435: 459–461.
Turner NJ, Lee MH, and Sano T (2014) Magnetic coupling in the disks around young gas giant planets. *Astrophysical Journal* 783: 14, arXiv:1306.2276v1 [astro-ph.EP].
Veras D and Armitage PJ (2004) Outward migration of extrasolar planets to large orbital radii. *Monthly Notices of the Royal Astronomical Society* 347: 613–624.
Vokrouhlicky D, Nesvodrny D, and Levison HF (2008) Irregular satellite by exchange reaction. *Astronomical Journal* 136: 1463–1476.
Wada K, Kokubo E, and Makino J (2006) High-resolution simulations of a Moon-forming impact and post-impact evolution. *Astrophysical Journal* 638: 1180–1186.
Walsh KJ, Richardson DC, and Michel P (2008) Rotational breakup as the origin of small binary asteroids. *Nature* 454: 188–191.
Ward WR (1979) Present obliquity oscillations of Mars – Fourth-order accuracy in orbital E and I. *Journal of Geophysical Research* 84: 237–241.
Ward WR (1981) Orbital inclination of Iapetus and the rotation of the Laplacian plane. *Icarus* 46: 97–107.
Ward WR (1993) Density waves in the solar nebula – Planetesimal velocities. *Icarus* 106: 274–287.
Ward WR (1997) Protoplanet migration by nebula tides. *Icarus* 126: 261–281.
Ward WR (2002) *Astrophysical Tides: The Effects in the Solar and Exoplanetary Systems*. Nanjing, China: IAU Colloquium.

Author's personal copy

Ward WR (2009) Particle filtering by a planetary gap. In: *40th Lunar and Planetary Conference*, 1477.
Ward WR (2012) On the vertical structure of the protolunar disk. *Astrophysical Journal* 744: 11.
Ward WR and Cameron AGW (1978) Disc evolution within the Roche limit. *Lunar and Planetary Science Conference* 9: 1205–1207.
Ward WR and Canup RM (2000) Origin of the Moon's orbital inclination from resonant disk interactions. *Nature* 403: 741–743.
Ward WR and Canup RM (2013) The evection resonance and the angular momentum of the Earth-Moon system. *44th Lunar and Planetary Science Conference,* March 18–22, LPI Contribution 1719; p 3029.
Ward WR and Canup RM (2006) Forced resonant migration of Pluto's outer satellites by an impact-produced Charon. *Science* 313: 1107–1109.
Ward WR and Canup RM (2010) Circumplanetary disk formation. *Astronomical Journal* 140: 1168–1193.
Ward WR and Hahn JM (1994) Damping of orbital inclinations by bending waves. *Icarus* 110: 95–108.
Ward WR and Hahn JM (2000) Disk-planet interactions and the formation of planetary systems. In: Mannings V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, pp. 1135–1158. Tucson, AZ: University of Arizona Press.
Weaver HA, Stern SA, Mutchler MJ, et al. (2006) Discovery of two new satellites of Pluto. *Nature* 439: 943–945.
Weidenschilling SJ (1977) Aerodynamics of solid bodies in the solar nebula. *Monthly Notices of the Royal Astronomical Society* 180: 57–70.
Weidenschilling SJ (1995) Can gravitation instability form planetesimals? *Icarus* 116: 433–435.
Wetherill GW (1985) Occurrence of giant impacts during the growth of the terrestrial planets. *Science* 228: 877–879.
Wetherill GW (1992) An alternative model for the formation of the asteroids. *Icarus* 100: 307–325.
Wetherill GW and Cox LP (1984) The range of validity of the two body approximation in models of terrestrial planet accumulation. I. Gravitational perturbations. *Icarus* 60: 40–55.
Wetherill GW and Cox LP (1985) The range of validity of the two body approximation in models of terrestrial planet accumulation. II. Gravitational cross sections and runaway accretion. *Icarus* 63: 290–303.
Wetherill GW and Stewart GR (1989) Accumulation of a swarm of small planetesimals. *Icarus* 77: 330–357.
Whipple FL (1964) The history of the solar system. *Proceedings of the National Academy of Sciences of the United States of America* 52: 565–594.
Widmer R, Masters G, and Gilbert F (1991) Spherically symmetric attenuation within the Earth from normal mode data. *Geophysical Journal International* 104: 541–553.
Wiechert U, Halliday AN, Lee D-C, Snyder GA, Taylor LA, and Rumble D (2001) Oxygen isotopes and the moon-forming giant impact. *Science* 294: 345–348.
Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27933–27968.
Willner K, Oberst J, Hussmann H, et al. (2010) Phobos control point network, rotation and shape. *Earth and Planetary Science Letters* 294: 541–546.
Wilner DJ, D'Alessio P, Calvet N, Claussen MJ, and Hartmann L (2005) Toward planetesimals in the disk around TW Hydrae: 3.5 Centimeter dust emission. *Astrophysical Journal* 626: L109–L112.
Wilson EH and Atreya SK (2004) Current state of modeling the photochemistry of Titan's mutually dependent atmosphere and ionosphere. *Journal of Geophysical Research* 109: E06002.
Winters RR and Malcuit RJ (1977) The lunar capture hypothesis revisited. *Moon* 17: 353–358.
Wisdom J (2004) Spin-orbit secondary resonance dynamics of Enceladus. *Astronomical Journal* 128: 484–491.
Wisdom J and Holman M (1991) Symplectic maps for the n-body problem. *Astronomical Journal* 102: 1528–1538.
Wood JA (1972) Fragments of terra rock in the Apollo 12 soil samples and a structural model of the moon. *Icarus* 16: 462–501.
Wood JA, Dickey JS, Marvin UB, and Powell BN (1970) Lunar anorthosites. *Science* 167: 602–604.
Yoder CF (1982) Tidal rigidity of Phobos. *Icarus* 49: 327–346.
Youdin AN and Goodman J (2005) Streaming instabilities in protoplanetary disks. *Astrophysical Journal* 620: 459–469.
Youdin A and Johansen A (2007) Protoplanetary disk turbulence driven by the streaming instability: Linear evolution and numerical methods. *Astrophysical Journal* 662: 613–626.
Youdin AM, Kratter KM, and Kenyon SJ (2012) Using Pluto's newest moon to constrain the masses of Nix and Hydra. *Astrophysical Journal* 755, Id 17, 11 pp.
Young EF and Binzel RP (1994) A new determination of radii and limb parameters for Pluto and Charon from mutual event lightcurves. *Icarus* 108: 219–224.
Yuan D-N, Sjogren WL, Konopliv AS, and Kucinskas AB (2001) Gravity field of Mars: A 75th degree and order model. *Journal of Geophysical Research* 106: 23377–23401.
Yung YL, Allen M, and Pinto JP (1984) Photochemistry of the atmosphere of Titan – Comparison between model and observations. *Astrophysical Journal Supplement Series* 55: 465–506.
Zhang J, Dauphas N, Davis AM, Leya I, and Fedkin A (2012) The proto-Earth as a significant source of lunar material. *Nature Geoscience* 5: 251–255.
Zolotov MY, Owen T, Atreya S, Niemann HB, and Shock EL (2005) An endogenic origin of Titan's methane. *American Geophysical Union*, *Fall Meeting 2005*, abstract #P43B-04.