The origin and early distribution of water ice and more volatile compounds in the outer solar system is considered. The origin of water ice during planetary formation is at least twofold: It condenses beyond a certain distance from the proto-Sun — no more than 5 AU but perhaps as close as 2 AU — and it falls in from the surrounding molecular cloud. Because some of the infalling water ice is not sublimated in the ambient disk, complete mixing between these two sources was not achieved, and at least two populations of icy planetesimals may have been present in the protoplanetary disk. Added to this is a third reservoir of water ice planetesimals representing material chemically processed and then condensed in satellite-forming disks around giant planets. Water of hydration in silicates inward of the condensation front might be a separate source, if the hydration occurred directly from the nebular disk and not later in the parent bodies. The differences among these reservoirs of icy planetesimals ought to be reflected in diverse composition and abundance of trapped or condensed species more volatile than the water ice matrix, although radial mixing may have erased most of the differences. Possible sources of water for Earth are diverse, and include Mars-sized hydrated bodies in the asteroid belt, smaller “asteroidal” bodies, water adsorbed into dry silicate grains in the nebula, and comets. These different sources may be distinguished by their deuterium-to-hydrogen ratio, and by predictions on the relative amounts of water (and isotopic compositional differences) between Earth and Mars.

**1. INTRODUCTION**

Water is, by number, the most important condensable in a cosmic composition soup of material. By mass, it rivals that of rock — depending upon the extent to which oxygen is also tied up in carbon monoxide and carbon dioxide (Prinn, 1993). And yet many workers fail to consider water ice as a “planet-building material” in the same way as silicates and metals are — in large part because we do not have samples of cometary material. Water ice is simply not stable on small bodies in the region of the solar system inhabited by Earth and Mars, inhabiting instead the polar and high-altitude regions of Earth and the poles and subcrustal reservoirs on Mars. Therefore, traditional meteorite studies ignore water ice in favor of the rocky and metallic phases.

Beyond the asteroid belt, water ice is abundant. It is a minor component of Jupiter’s moon Europa, but constitutes almost half the mass of Jupiter’s moons Ganymede and Callisto and Saturn’s moon Titan. It is the dominant, or at least key, constituent of the intermediate-sized moons of Saturn (e.g., Enceladus), the moons of Uranus, Neptune’s moon Triton, and the Kuiper belt object Pluto and its moon Chiron. It was almost certainly an important core-building material of the giant planets. Likewise, water ice is an important component of comets, ranging from being the major solid in fresh comets to a minor component of old comets in asteroid-like orbits. Thus, understanding the formation of planets and smaller bodies in our solar system requires consideration not only of the meteorite record but of icy bodies as well.

Unlike meteorites, icy bodies have been studied only by remote sensing (even the sample collection by the *Stardust* mission captures only, or primarily, atoms from the dust component of Comet 83P/Wild 2). Future opportunities to study icy bodies directly will come from the *Rosetta* mission, already launched, the *Deep Impact* mission to impact the surface of a comet, and a proposed Europa lander or penetrator. Ice in the high latitudes of Mars will be directly sampled and studied by the Project Phoenix Mars lander in 2008. Beyond these four opportunities, icy bodies in the outer solar system will continue to be studied primarily by remote sensing in the near future.

This brief chapter sketches the possible origins of condensed phases of water. Sources of water via direct condensation and infall from the surrounding molecular cloud are considered first, followed by consideration of how giant planet formation may have led to a chemically distinct class of water ice. The implications of the origin and distribution of water ice in the solar system for the source of Earth’s water is considered, and the chapter closes with a brief consideration of the density of outer solar system icy bodies for the existence of different reservoirs of condensed water.

**2. THE NEBULAR SNOWLINE**

The protoplanetary disk out of which the solar system formed (which is referred to in the meteoritical literature as the solar nebula, a term we will use as well here) is a natural outgrowth of the interaction between gravitationally driven collapse of a dense clump of molecular cloud gas,
and the conservation of angular momentum contained in the material. The disk is the physical medium of gas and solids through which much of the material out of which the Sun itself formed traveled, and dissipation in the disk transported mass inward to the center and angular momentum outward (Nelson et al., 1998). The dissipation is accomplished through net gravitational forces (torques) that one portion of the rotating disk will exert on another, generating waves of various types (Lin et al., 2000), through torques associated with a possible magnetic field embedded in the disk, or through the shear associated with the radial variation of Keplerian rotation and motion perpendicular to the disk (“vertical”) caused by convection of heat away from the disk midplane to the colder upper regions (Stevenson, 1990).

[An alternative birth site to a clump in a dense molecular cloud, offered recently on the basis of the apparent existence of live 60Fe in early solar system materials (Hester et al., 2004), in the midst of an assemblage of short-lived and massive high-mass stars, has potential implications for typical solar nebula disk models that have yet to be evaluated.]

In a disk with sufficient gas density to be optically thick (that is, with an optical depth — the product of scale length, material density, and absorption coefficient — in excess of unity), the temperature drop along the midplane is determined by the nature of the dissipative processes, and will decline with distance r from the center along the midplane as 1/r. For an optically thin (optical depth less than 1) disk, the temperature drop is determined by absorption of the Sun’s radiation by midplane material and drops as 1/r1/2.

In either case, this drop in density ensures that some sort of radial gradation in solid-forming material will occur within the disk. Refractory silicates such as corundum will be stable as solids closer to the proto-Sun than will the more-abundant magnesium silicates, and water ice will appear at even greater distances. The radial distance along the midplane at which water ice may first stably appear is referred to in the planetological literature as the snowline.

The simplest calculation of the snowline requires comparing the saturation vapor pressure — a function of the ambient temperature — to the partial pressure of water vapor at the midplane, as a function of distance outward from the center. The partial pressure of water vapor is the total nebular pressure times the mole (i.e., number) fraction of water, and the latter inward of the condensation front is simply the mole fraction based on cosmic abundance of the elements and the distribution of oxygen among several molecular species. Where the partial pressure exceeds the saturation vapor pressure it is thermodynamically possible for condensation to occur, and hence water ice (perhaps with the consistency of snow) will form there. A typical nebular temperature for the water condensation is between 160 and 170 K, with a radial distance that depends on nebular models and the stage of nebular evolution under consideration. The snowline radius, rsnow, has a rather large potential range between 1 AU and 5 AU, the former being an extreme value for older, dusty, cold disks (Sasselov and Lecar, 2000). Figure 1 shows the snowline position for several different models of the temperature of the solar nebula.

The warmest is purely schematic, in which the temperature at 5 AU is set to 160 K and that at Saturn’s orbit to 100 K, consistent with the volatiles seen in Saturn’s moon Phoebe, likely captured from solar orbit in the vicinity of Saturn. The other profiles are displaced downward and come from Sasselov and Lecar (2000), but have the same slope as the schematic model. Thus there is no clear preference based on nebular temperature profiles for a snowline at 1 vs. 5 AU, but both may obtain at different times in the history of the nebula as accretion ceases and temperatures decline.

The above calculation of the snowline radius assumes that the gas and the grain temperatures are identical, which is not necessarily the case. Sizes of grains falling into the nebula range from 0.1 to 10 µm or larger, although the larger grains will be fluffy aggregates of, and hence behave thermally as, the smaller particles (Weidenschilling and Ruzmaikina, 1994). Radiative properties of the grains depend on their composition but especially on their size, since the wavelength range over which the peak emission occurs is comparable to or larger than the particle size. Put simply, the smaller particles are poorer radiators and absorbers of thermal energy than are the larger ones (Lunine et al., 1991). The cross section for interaction between a grain of radius a and radiation of wavelength λ can be written as the prod-

![Fig. 1.](image-url)
uct of the geometric area of the grain and an efficiency factor $Q$; the latter is near unity if the grain radius (actually, the circumference $2\pi a$) is comparable to or larger than $\lambda$, and decreases rapidly as $2\pi a/\lambda$ drops progressively below unity corresponding to shrinking grain size (Podolak and Zucker, 2004).

The primary consequence of this analysis is that heating and cooling fluxes must be computed in a time-dependent fashion to determine the particle temperatures as a function of the gas temperature and the particle size and composition (icy vs. rocky). In practice this leads to little change in the position of the snowline at the nebular midplane in terms of temperature: Depending on the assumptions regarding heating and cooling rates, the snowline is between 150 K and 170 K, and the temperatures of the gas and icy grains are nearly equal to each other there. However, grain temperatures may be well below the nebular gas temperature inward of the snowline, where only silicates are stable. This is a consequence of the nebular midplane being optically thick, because then the gas is the only source of heating for the grains and the small size and low emissivity of the silicate grains makes the process inefficient (Podolak and Zucker, 2004). The situation is much more complicated at the optical surface, or photosphere, of the nebula, where the gas becomes (by definition) optically thin, and grains lofted from the midplane evaporate due both to the lower gas pressure and the direct radiation from the Sun. Although evaporation rates at 150–170 K are very small, dirty ice grains (those with darkening agents) may evaporate within the lifetime of a 10-m.y. disk. Clean ice grains have longer lifetimes (Fig. 2), and so the grain longevity may depend on what other materials are trapped in the water ice during grain formation, and how radiation from the proto-Sun will alter them. For example, methane and methanol may darken rather quickly and increase the grain evaporation rates, and these might be incorporated into icy grains at nebular distances as small as 5 AU (Hersant et al., 2004), or be transported inward by radial gas drag after formation (Cyr et al., 1998) — a complication to which we next turn.

Inward transport of ice particles by gas drag constitutes an important modification to the simple vapor-pressure-driven picture. Because the gaseous phase of the solar nebula is supported to a small extent by the pressure force, the gas molecules orbit slightly more slowly than the Keplerian speed associated with their radial distance from the disk center. Solid particles thus experience a wind as they move at the Keplerian orbital speed through the gas, which is a function of grain size (Weidenschilling, 1977). Very small grains — those less than the mean free path in the gas — act like gas molecules, embedded in the flow and affected only by the collisions with surrounding molecules; particles larger than the mean free path experience frictional drag and are slowed down, causing them to spiral inward toward the center of the disk. Because in the drag regime the ratio of surface area to mass (volume) of the particle determines the efficiency of the drag force, the largest particles will expe-

![Fig. 2. Lifetime of grains against evaporation for 0.1- (dashed) and 10-µm (solid) grain sizes; the smaller grains are poorer absorbers and radiators of photons and hence evaporate more slowly. The horizontal line corresponds to a disk (solar nebula) lifetime of 10 m.y. From Podolak and Zucker (2004).](image-url)
(Chick and Cassen, 1997). For the purely diffusive solution, the region of “enhanced” (over the depleted background) water vapor is about 1 AU (Cuzzi and Zahnle, 2004), and the nondiffusive effects cited above will tend to broaden this somewhat.

The snowline is undoubtedly more complicated than has been outlined here, and will not be a unique phenomenon in a nebula where other major species (silicates, sulfur-bearing compounds, ammonia, carbon dioxide) may condense out directly, although many molecular species such as methane and the noble gases are more likely to become trapped in the water ice as an adsorbate or clathrate hydrate guest molecule (Gautier et al., 2001). However, the basic idea that there is a particular distance at which water ice becomes stable and abundant, and which consequently alters the water vapor abundance inward of that point, is a robust one at any given time in the evolution of the nebula.

The snowline has been interpreted traditionally as setting the dividing line between icy (cometary) and rocky (asteroidal) small bodies, with water-rich (C-type) asteroids receiving their water inventory from hydration reactions occurring at temperatures too high for ice itself to be stable. Thus, a somewhat more poorly defined “hydration line” could be defined as extending inward some 1 or 2 AU from the snowline itself, based on the thermodynamic stability of hydrated silicates at temperatures below between 225 K and 250 K (Fegley, 2000), and hence defining the parent bodies of the carbonaceous chondrites. However, the kinetics of the hydration reaction between the water-laden gas and preexisting silicate grains are such that the time for conversion may greatly exceed the disk lifetime (Fegley, 2000). Others have argued that the process may be much faster on the basis of laboratory studies of dehydration and a model for the relationship between the forward and back reactions (Ganguly and Bose, 1995), or because of shock effects near the snowline itself (Ciesla et al., 2003).

Laboratory studies of carbonaceous chondrites suggest that the hydration reactions occurred inside meteorite parent bodies upon exposure to liquid water, rather than in the nebula itself (Clayton, 2003). This would implicate water ice within the asteroidal parent bodies, acquired perhaps as the snowline moved inward prior to the accretion of these objects, as the source of the water of hydration. The primary reservoir of water within the hydrated asteroids (or their parent bodies) would then be that of the snowline, with the consequent implication that icy bodies formed outside the snowline ought to have an isotopic composition, for the water, the same as that found in carbonaceous chondrites—in particular, a deuterium-to-hydrogen ratio with a mean value of standard mean ocean water (“SMOW”). If, instead, the silicates were hydrated from the nebula itself, inward of the snowline and hence in the absence of the formation of water ice itself, then the isotopic composition of the hydrated and icy material may not be related. This is an important issue that pertains to the origin of Earth’s water, which we will discuss in section 5.

### 3. INFALLING WATER-ICE GRAINS

Water ice is formed directly in interstellar clouds, mostly from direct adhesion of atoms of preexisting silicate grains at very low temperatures yielding amorphous solid-water phases (Irvine and Knacke, 1989), but these may be modified by sublimation and recondensation of such amorphous structures as they cycle in and out of hot cores in the clouds. Grains grow as they fall into a protoplanetary disk (Weidenschilling and Ruzmaikina, 1994), but sublimation of the water ice will occur both in gas dynamical heating associated with shocks (Lunine et al., 1991) and in the ambient disk within a certain radial distance from the Sun, due simply to the radial-temperature profile of the disk. This latter effect is the same as that which determines the nebular snowline distance, but here preexisting, infalling ice particles are considered rather than the formation of the ice particles from the nebula vapor phase.

As one might expect, the survival zone for infalling ice particles could be as close as the nebular snowline, when dynamical heating due to infall is assumed minimal, or could be much farther out if shocks play an important role as a heating mechanism (Chick and Cassen, 1997). The extent to which shocks heat the grains depends on the type of shock formed as accreting material accelerates toward the disk and then is decelerated by interaction with the disk itself. Chick and Cassen (1997) examined a range of nebular models and found water-ice grain survival could occur as close as 5 AU for an optically thick disk, moving inward to as close as 2 AU if an optically thin cavity forms. The ice stability line could have been as far as 30 AU from the proto-Sun for high accretion rates and warm disk temperatures. That the ice survival line should range from 2 to 30 AU simply reflects the wide range of possible conditions during the lifetime of the solar nebula disk, as well as uncertainties in the model parameters and the environment surrounding the disk. Most of the water vapor in the protoplanetary disk must originally have had its source in sublimated amorphous ice grains in the standard picture of a cold molecular cloud phase as a precursor of collapse of clumps to form stars and planets. However, birth in a more supernova-rich environment with strong UV erosion of a disk (Hester et al., 2004) could have implied a different history of the water in vapor and amorphous phases and led to very different patterns of ice survival, but these possibilities have yet to be quantified.

Ice that has been sublimated (or vaporized, if a surface layer of transient liquid forms first on the grain, although this has no measurable effect on the outcome) will recondense if the ambient gas temperature into which the vapor is mixed is low enough, i.e., if the final “resting place” of the vapor is beyond the snowline. And vapor inward of the snowline will eventually diffuse or be advected or convected into the snowline region, leading to the picture of a water-vapor-depleted inner zone and a water-rich outer zone, the boundary being the snowline. The practical effects of the
sublimation and recondensation are two-fold: (1) to convert amorphous ice into crystalline ice and (2) to redistribute substances more volatile than water ice from trapping sites in interstellar grains into the nebular gas, and hence into other phases in the reformed ice and silicate grains.

Effect (1) requires that amorphous ice formed at very low temperatures in molecular clouds be sublimated (at any temperature) and then recondensed at a temperature in excess of 130–140 K (Kouchi et al., 1994). If the nebular snowline is at a temperature of 150–160 K, driven by the solar abundance of available oxygen in the form of water vapor, then this requirement is satisfied. We therefore expect the solar nebula to have contained a mixture of amorphous and crystalline ice: the crystalline ice a result of the water vapor sublimated during nebular infall and then recondensed at the snowline, the amorphous representing the grain component that survived infall to the disk. Some protostellar regions show evidence for crystalline ice (Malfait et al., 1999), and it is possible that disk reprocessing is the source, although direct conversion from amorphous ice in the surrounding hot core should also be examined. Likewise, fresh comets should be a mixture of crystalline and amorphous ice, unless other heat sources (such as decay of the short-lived isotope $^{26}\text{Al}$) trigger conversion early in the history of the solar system. The low spin temperature seen in the hydrogen in some comets (Kawakita et al., 2001; Irvine et al., 2000) argues against wholesale reheating and in favor of the preservation of delicate amorphous grains.

Effect (2) is potentially much more complex. Interstellar grains are cold enough that virtually all elements and molecules may adsorb at some point on the grains; laboratory experiments down to 20–30 K have been conducted to confirm that the resulting abundances of (for example) noble gases are essentially proportional directly to the gas phase abundances, and it has been argued that this pattern is seen, for example, in elemental abundances in Jupiter (Owen et al., 1999). However, somewhere within 30 AU of the proto-Sun, water ice will be partially or completely sublimated during nebular infall, and the more volatile species released into the gas phase. These species in turn will be retrapped as icy grains reform, but at different pressures and temperatures than obtained in the original molecular cloud (Lunine et al., 1991). Indeed, formation of crystalline ice or clathrate hydrate — in which the volatiles are trapped in a regular crystal structure of water ice with a well-defined entropy and thermodynamic free energy — might have taken place in the region where Jupiter now exists (Gautier et al., 2001) or even further out (Hersant et al., 2004). The pattern of volatile enrichment in the ice depends, in this case, on the details of the reformation and cooling history of the ice.

One potential way of distinguishing between the two alternatives requires a definitive measurement of the oxygen abundance within Jupiter. Adsorption of volatiles on small ice grains at 20 K should be extremely efficient, leading to an enrichment of elemental oxygen (relative to the solar abundance) comparable to that of the noble gases. If, instead, the trapping occurred at a later stage in the nebula, in crystalline (possible clathrate hydrate) water ice, it would have been much less efficient, and the oxygen abundance correspondingly much higher. Specifically, the “primitive adsorption” model implies an oxygen abundance in Jupiter three times solar, while the sublimation and retrapping model implies an oxygen abundance at least nine times solar (Hersant et al., 2004).

### 4. WATER ICE IN THE REGION OF GIANT PLANET FORMATION

The environments around the giant planets during their formation were vastly different from that in the ambient solar nebula. Unfortunately, quantifying these conditions is extremely difficult, because the mechanisms by which Jupiter and Saturn (as well as extrasolar giant planets) formed remain controversial (Lunine et al., 2004). The traditional model relies on the so-called “nucleated instability,” in which the growth of a rock and ice core triggers rapid accretion, if not collapse, of a large gaseous envelope to make a giant planet (Pollack et al., 1996; Wuchterl et al., 2000). Direct collapse, a process dismissed many years ago based on preliminary mathematical models of protoplanetary disks, has been found to produce bodies of jovian mass given the right, marginally unstable, background disk conditions of low temperature and relative high density (Boss, 2001). The timescales of the two processes are very different — nucleated collapse requiring on the order of a million years, and direct collapse taking as little as hundreds of years (Mayer et al., 2002). Thus, the temperature and density perturbations of the background nebula must clearly have been different in the two cases. The nucleated instability formation of giant planets is associated with short-lived global perturbations on the entire protoplanetary disk, and so it is not useful to try to distinguish a special set of conditions near the collapsing giant planet that might produce a distinct reservoir of icy planetesimals. Indeed, the disk collapse into giant planets, and subsequent gravitational interactions among these planets leading to a subset being ejected, would likely have erased any radial gradations in icy grain properties through direct dynamical effects and changes in the radial temperature profile. These have yet to be evaluated quantitatively in the context of a detailed model of an unstable protoplanetary disk.

The longer timescales and more nearly quasistatic conditions during the nucleated-instability growth of the giant planets renders practical the modeling of nebular conditions in the vicinity of the growing giant planets. Optically thick disk models with temperature profiles determined by an adiabat are a possible end member for Jupiter and Saturn (Lunine and Stevenson, 1982), but an oversimplified one because such a disk does not take into account the transport of material through the disk into the giant planet itself. More recent models include disk transport, and find that an optically thick circumjovian disk is too hot to per-
mit a snowline at the orbits of Ganymede or Callisto — or anywhere within a reasonable migration distance. Disks may be cooler, less massive, and with a slower influx of material into proto-Jupiter (Canup and Ward, 2002) — limited by the 5–10-m.y. timescale over which Jupiter forms (Lunine et al., 2004). Such disks exhibit snowlines located so as to permit ice to exist where Ganymede and Callisto formed (with a small amount of inward migration), and water of hydration to occur at the orbit of Europa (Fig. 3). Disk pressures are elevated above that in the surrounding solar nebula, so that the snowline and hydration fronts are at correspondingly higher temperatures. The 200 K temperature appropriate for the snowline in the circumjovian disk almost certainly means that crystalline water ice condenses in preference to amorphous ice, and the trapping of more volatiles species depends on clathrate formation rather than adsorption in amorphous ice layers. Thus, the planetesimal population in the circumplanetary disks will differ in their physical properties and volatile compositions from those in the solar nebula, and hence constitute a distinct, third, reservoir of water ice.

5. DELIVERY OF WATER TO THE TERRESTRIAL PLANETS

Although the outer solar system contains many tens of Earth masses of water, located within the giant planets, their moons, and objects in the Kuiper belt and Oort cloud, it is the water in the terrestrial-planet region that has garnered the most persistent interest from the astrobiological point of view [despite the possibility of a subsurface ocean on Europa (Chyba et al., 1998)]. The roughly 0.001 M⊕ of water in the Earth’s crust and mantle is the critical factor in the existence of life — together with a surface temperature salubrious for liquid water. Evidence continues to accumulate that Mars once possessed a surface water budget equivalent to somewhere between 0.01 and 1 times that of Earth (Baker, 2001). The highly enriched deuterium abundance in the atmosphere of Venus suggests that this nearly Earth-sized planet once contained amounts of water comparable to Earth (Hunten et al., 1989). Thus, although the amounts of water are small relative to those in the outer solar system, the implications for the question of planetary habitability are large, and hence the question of the origin of water on the terrestrial planets deserves close attention.

It is possible, based on the temperature profiles shown in Fig. 1, that water was directly available in the form of ice or water of hydration in the nebula at or near 1 AU when grains were still small. However, temperature profiles low enough for pure ice to condense are extreme ones, and likely existed (if at all) in the very late stages of disk accretion, when silicate bodies were already large (kilometer-sized or larger). The constraints on the temperature profile are less severe for water of hydration, but as discussed in section 2, doubts have been raised as to whether hydration of silicates could have occurred in the nebula. Were hydration to be widespread, we might expect to find a chondritic composition for planetesimals in the 1 AU region, and hence for Earth itself. However, Earth itself cannot contain more than a few percent, perhaps only 1%, of chondritic material (Drake and Righter, 2002). An alternative is that small bodies migrated inward from a zone of hydration at 2–3 AU via gas drag in the disk and retained sufficient water to supply Earth’s inventory (Ciesla et al., 2004). Adsorption of water vapor onto dry silicate grains that have the isotopic and elemental composition of Earth is also possible, but the grains must have a highly fractal nature in order to adsorb sufficient water (Stimpfl et al., 2004). Both the adsorption and the inward migration model create constraints on the properties of planetesimals at the time of the growth of Earth (size distribution, porosity, ambient gas density) that have yet to be explored in detail.

Bringing water to Earth from more distant regions of the protoplanetary disk could also have been accomplished through gravitational scattering of icy and hydrated bodies onto highly eccentric orbits that extend from their colder regions of origin to the orbit of Earth. At first sight, this would seem to be a very slow and inefficient process, because planetesimals are initially on circular orbits (determined by the effects of gas drag), and grow rapidly in the presence or absence of gas to an “isolation mass” at which the mean separation between bodies is vastly larger than their cross sections for gravitational interaction (Goldreich et al., 2004; Raymond et al., 2004). That mass is somewhere between the mass of the Moon and Mars, and dramatically slows further accretion to timescales much longer than the
tens of millions of years for Earth constrained by radioisotopic data (Cameron, 2002). However, the growth of Jupiter provided a large perturbing mass that rapidly increased the eccentricity of these lunar- to Mars-sized “embryos,” particularly in what is now the outer asteroid belt, sending them on trajectories to the inner solar system and, in some cases, on Earth-crossing trajectories. Since the formation of Jupiter had to occur when gas was present (Wuchterl et al., 2000), this must have occurred within the several-million-year period of the existence of the nebular gas (Hartmann, 2000), and various models of giant planet formation yield formation times comparable to this (Ida and Lin, 2004) or much shorter (Boss, 2003; Mayer et al., 2002).

Numerical simulations of this process (Morbidelli et al., 2000; Chambers and Cassen, 2002; Raymond et al., 2004) produce Earth-sized planets in the region around 1 AU on timescales consistent with the geochemical constraints. All the simulations assume that the region of the asteroid belt contained at least as much rocky material, prior to the formation of Jupiter, as did the equivalent area of the disk in the terrestrial planet region — a reasonable assumption given the current architecture of the asteroid belt and no compelling mechanism for creating an early “pre-Jupiter” gap corresponding to the present-day dearth of material between 2 and 4 AU (Petit et al., 2001). In most of the outcomes of the planet-forming simulations from lunar- to Mars-sized embryos, the majority of material from which these bodies are accreted is local, and determines the geochemical composition of the final planet. But a fraction comes from what is now the asteroid belt, and a portion of that from the region beyond 2 or 2.5 AU where the relatively water-rich chondrites are believed to have originated. The amount of water delivered to Earth, or to any of the terrestrial planets, is highly variable from one simulation to another [ranging from less than an Earth ocean to well over 100 Earth oceans (Raymond et al., 2004)] because the number of embryos is small — on the order of $10^2$–$10^3$. Thus, small changes in the starting conditions can lead to very different results in terms of planet position, mass, and water abundance.

Even assuming a significant loss of water during impact of the embryos with the growing Earth (approximately 50% loss when the mass of the growing Earth is close to the final value, solely an “educated guess” absent detailed tracking of the water during giant impact simulations), the dynamical simulations often yield 1-M$_{\oplus}$ bodies near 1 AU with an amount of water that overlaps estimates for the surface and mantle — about 3 to 5 times the mass of the surface ocean of Earth today (Abe et al., 2001). The simulations will also produce very water-rich Earths in some cases, up to the wet primordial mantles envisioned by some workers (Dreibus and Wanke, 1987).

A potential problem with this scenario arises from the amount of chondritic material added by Earth — assumed to be the prototype for the water-rich embryos beyond 2–3 AU from the Sun — which must be less than 1–3% (depending on the timing of the addition relative to core formation) to satisfy (1) the abundance of siderophile elements in the mantle and (2) the oxygen-isotopic similarity between Earth and the Moon (Drake and Righter, 2002). Except for the lower end of the estimates of total water added to Earth, the dynamical model modestly violates (by a factor of 2 or 3) the two geochemical constraints. However, constraint (1) can be removed if the carbonaceous embryo that delivered the water was differentiated, or partially differentiated. In this case, its core, containing most of the siderophile elements, would not mix with Earth’s mantle, and thus a much larger fraction of carbonaceous material could have been delivered to Earth without exceeding the amount of siderophile in Earth’s mantle. Constraint (2) can be removed if there is a way to homogenize the oxygen-isotopic composition of Earth and the Moon soon after the giant impact that formed the Moon. More work must be done to evaluate these possibilities.

The model for adding water to Earth from the primordial asteroid belt has the virtue that the deuterium-to-hydrogen ratio (D/H) measured in Earth’s oceans, the so-called “SMOW,” is consistent with the most probable value of D/H computed from the broad range seen in carbonaceous chondrites (Robert, 2001; see also Robert, 2006). A cometary contribution of water would have a higher D/H: Three comets that fall into the “long-period” comet class, thought to be derived from the Oort cloud, all have D/H approximately twice that of SMOW (Meier et al., 1998a,b). If typical, this should then limit severely the amount of Earth’s water that could have been contributed by comets, because no plausible mechanism has been identified for reducing the D/H value after accretion (contact with a nebular composition atmosphere without loss of the water to space seems implausible). Indeed, dynamical calculations limit the amount of water contributed by comets (essentially, from icy bodies resident at and beyond the orbit of Jupiter) to 10% of the total brought in from sources in the primordial asteroid belt (Morbidelli et al., 2000).

Is it possible that the D/H measured in comets represents an alteration of an original value via phase changes in the cometary nucleus? The three comets analyzed have different amounts of exposure to sunlight, Halley having been around the Sun many times, with Hyakutake and Hale-Bopp appearing to be relatively “fresh” comets. This question can be addressed experimentally (R. H. Brown and D. S. Lauretta, personal communication, 2005) or by measurement of additional long- and short-period comets. Podolak et al. (2002) have proposed a mechanism for altering the measured coma value of D/H relative to that in the nucleus of a comet, although concerns have been raised about their model (Krasnopolsky, 2004).

Mars is an important test of the origin of water in the terrestrial planets because of its greater proximity to the snowline and its small mass, which suggests that it could be a planetary embryo left behind from the accretion process, either by chance (Lunine et al., 2003), or through the orbital damping effect of residual nebular gas (Kominami and Ida, 2002). In either case, the source of water can no longer be large embryos, but is a mixture of smaller bodies, some of which are comets. The proportionate amount of
water contributed by comets depends on the importance of various local and distal sources of water, i.e., adsorbed water on grains, hydrated or icy bodies brought in by gas drag, or small primordial asteroids whose orbits were gravitationally perturbed by Jupiter. Although the martian data are not yet precise enough to choose among these possibilities, in principle better constraints on the initial amount of water on Mars (currently two orders of magnitude uncertain) will allow such a test to be made. The contribution of high D/H water to the inner solar system is perhaps hinted at in the analysis of hydrated minerals in martian meteorites, with values tending toward twice SMOW (Leshin, 2000), although it appears to be lower in some samples (Boctor et al., 2003; Gillet et al., 2002). The complex situation associated with D/H on Mars is discussed in more detail in Robert (2006).

6. THE OUTER PLANET SATELLITES: CONSTRAINTS ON RESERVOIRS OF WATER ICE

The satellites of Jupiter, Saturn, and Uranus are a diverse group of objects that illustrate well the idiosyncrasies of the formation of secondary bodies around giant planets (Lunine et al., 2004). Satellite formation in the Jupiter environment encouraged, for whatever reason (Mosqueira and Estrada, 2003), the formation of four large moons with roughly comparable amounts of silicate — perhaps reflecting a limit on the silicate abundance (essentially, surface density of refractories) in the disk. The total mass of the satellites was then determined by the amount of ice added, and this was in turn determined by the strong temperature gradient in the circumjovian disk (Camap and Ward, 2002). Conditions at Io were too warm even for hydrated silicates — or tidal heating result in dehydration of the phyllosilicates and loss of water. Europa began as a hydrated-silicate body, with subsequent internal tidal and radiogenic heating leading to the generation of a thin water mantle atop the Io-mass silicate body. Ganymede and Callisto formed with a full complement of water in the disk, but the accretional energy released per unit mass toward the end of accretion was comparable to the latent heat of vaporization of water, and so accretion cooled off as water vaporization and loss occurred (Stevenson et al., 1986). Thus none of the jovian satellites give us the rock-to-ice ratio of the primordial circumjovian disk.

The saturnian system, on the other hand, consists of (excluding Ganymede-like Titan) intermediate-sized satellites with sizes and densities that are not systematic. The formation of this system remains enigmatic, but evidently less material was available than at Jupiter, and the process was not determined by a radial temperature gradient in the circumsaturnian disk. The mean density of the intermediate-sized saturnian satellites, mass weighted and excluding Titan, is determined from Voyager data to be 1.3 g/cm³ (Jacobson, 2004), well below the 1.9 g/cm³ of Ganymede, Callisto, and Titan, and indicative of a circumsatellite nebular chemistry distinct from the solar nebula. If the dominant form of carbon in the solar nebula was CO and not CH₄ — an assertion consistent with models of the gas chemistry of the disk — then the amount of oxygen available to make water ice implies a density for ice bodies somewhere around 2.3 g/cm³ — a number increased over previously published values by the elemental oxygen abundance in the Sun, which has recently been redetermined and substantially lowered relative to earlier studies (Asplund et al., 2004).

The one intermediate-sized satellite that is demonstrably a captured object (in a loose retrograde orbit), Phoebe, has a Cassini-determined density around 1.63 g/cm³, contains water ice (Clark et al., 2005), and is irregularly shaped. It is significantly denser than the icier saturnian satellites, yet is small enough to be a porous body. For a reasonable porosity of 10–20%, Phoebe’s material density would be essentially identical to the 2 g/cm³ measured for Pluto and Triton (Stern et al., 1997), and hence consistent with formation in solar orbit somewhere in the outer solar system. The inferred rock-to-ice ratios for Pluto and Triton imply a solar nebula whose carbon budget is largely (but not exclusively) in the form of carbon monoxide (Johnson and Lunine, 2005). The densities of the other saturnian satellites demand a circumsaturnian disk chemistry much richer in water and hence a carbon budget in which carbon monoxide is a minor or absent component, and methane (or other nonoxidized carbon species) dominate. This more reduced carbon budget is also chemically consistent with the circumstantial evidence for circumsaturnian ammonia (NH₃) as the original source of Titan’s atmosphere (Owen, 1982), suggested by the absence of nonradiogenic argon in measurements of Titan by the Cassini Ion and Neutral Mass Spectrometer (Waite et al., 2005). If ammonia were absent from the warmer jovian protoplanetary disk, it would explain why Ganymede and Callisto do not have Titan-like dense atmospheres.

The major uranian satellites have water ice on their surfaces (Brown and Clark, 1984) but a mean density determined by Voyager 2 higher than that for the saturnian satellites (Johnson et al., 1987), implying higher rock-to-ice ratios in their interiors. Either the circumuranian disk was not dense or hot enough for its composition to be altered from the solar nebula value, or the circumstances of satellite formation were affected by the impact that altered the Uranus obliquity (Korycansky et al., 1990). Finally, while Pluto and its moon Charon show water ice on their surfaces (Cruikshank et al., 1997), Neptune’s moon Triton does not (Brown et al., 1995). Since Triton and Pluto have similar densities and are thought to have similar origins (McKinnon et al., 1995), it is assumed that the water ice crust of Triton is buried beneath other ices.

7. SUMMARY

Water ice, and water of hydration, were major planet-building materials in the protoplanetary disk. Multiple sources of water ice are suggested by modeling of the evolution of the planet-forming disk, but little evidence of these sources can be gleaned from the isotopic and chemical evidence currently available. On the other hand, the densities of the icy bodies of the outer solar system — giant planet
moons and Pluto — do suggest that regions around the forming giant planets existed that were chemically distinct from the surrounding solar nebula. Further progress in elucidating the origin of water ice during the formation of the planets, its abundance distribution, and content of more volatile gases depend on measurements of the water abundance and isotopic composition of water in martian materials, more accurate satellite densities, and isotopic measurements of water in comets, among others. These are daunting goals, but important ones if we are to understand how water finds its way into both icy bodies and habitable worlds in planetary systems.

Acknowledgments. Helpful discussions with A. Morbidelli form a part of the discussion on the limits to the carbonaceous contributions to Earth’s water budget. Reviews by F. Robert, T. Owen, and A. Morbidelli are also gratefully acknowledged. The preparation of this chapter was supported by NASA's Planetary Atmospheres Research and Analysis Program.

REFERENCES


