We review our current understanding of the formation and early evolution of the solar nebula, the protoplanetary disk from which the solar system formed. Astronomical understanding of the collapse of dense molecular cloud cores to form protostars is relatively advanced, compared to our understanding of the formation processes of planetary systems. Examples exist of nearly all the phases of protostellar evolution, guiding and validating theoretical models of the star formation process. Astronomical observations of suspected protoplanetary disks are beginning to provide key insights into the likely conditions within the solar nebula, although usually only on much larger scales than the regions where the terrestrial and giant planets formed. The primary mechanism for driving the early evolution of the solar nebula, leading finally to the formation of the planetary system, is still highly uncertain: Magnetic fields, gravitational torques, and baroclinic instabilities are among the competing mechanisms. While the discovery of extrasolar gas giant planets has reinforced the general belief that planetary system formation should be a widespread process, the absence of any undisputed evidence for giant protoplanets in the process of formation makes it hard to decide between competing mechanisms for their formation. The situation regarding extrasolar terrestrial planets is even less constrained observationally, although this situation should improve tremendously in the next decade.

1. INTRODUCTION

The most primitive meteorites are believed to be relatively unaltered samples of the basic building blocks of the inner solar system. As such, the primitive meteorites, as well as comets, are believed to preserve parts of the record of the formational processes that led to the origin of our solar system. Presolar grains contained within them undoubtedly carry much of the history of galactic nucleosynthesis in their isotopic abundances (Nittler and Dauphas, 2006; Meyer and Zinner, 2006). A major portion of the motivation for observational and laboratory studies of meteorites and comets stems from the desire to use the results to constrain or otherwise illuminate the physical and chemical conditions in the solar nebula, the Sun’s protoplanetary disk, in the hopes of learning more about the processes that led to the formation of our planetary system. In addition to meteoritical studies, there are important lessons to be learned about the planet formation process from astrophysical observations of young stellar objects and their accompanying protoplanetary disks. Theoretical models then serve as the ultimate recipient of these constraints and clues, charged with the task of assembling a cohesive explanation for some or all of the data relevant to some particular aspect of the planet formation process. While much progress has been made in the last several decades, we are still far from having an agreed-upon, universal model of planetary system formation and early evolution. An excellent resource for learning more about these and other subjects is the review volume Protostars and Planets IV (Mannings et al., 2000). The present chapter is adapted, modified, and updated from two recent review articles about the solar nebula (Boss, 2003, 2004c).

2. FORMATION OF THE SOLAR NEBULA

The protosun and solar nebula were formed by the self-gravitational collapse of a dense molecular cloud core, much as we see new stars being formed today in regions of active star formation. The formation of the solar nebula was largely an initial value problem, i.e., given detailed knowledge of the particular dense molecular cloud core that was the presolar cloud, one could in principle calculate the flow of gas and dust subject to the known laws of physics and thus predict the basic outcome. Specific details of the outcome cannot be predicted, however, as there appears to be an inevitable amount of stochastic, chaotic evolution involved, e.g., in the orbital motions of any ensemble of gravitationally interacting particles. Nevertheless, we expect that at least the gross features of the solar nebula should be predictable from theoretical models of cloud collapse, constrained by astronomical observations of young stellar objects. For this
reason, the physical structure of likely precollapse clouds is of interest with regard to inferring the formation mechanism of the protosun and the structure of the accompanying solar nebula.

In the following we describe the current state of our knowledge about the nature of the precollapse, dense molecular cloud, and the collapse of such a cloud leading to the formation of a protostar surrounded by a protoplanetary disk, based on the results obtained from analytical modeling and astronomical observations.

2.1. Observations of Precollapse Clouds

Astronomical observations at long wavelengths (e.g., millimeter) are able to probe deep within interstellar clouds of gas and dust, which are opaque at short wavelengths (e.g., visible wavelengths). These clouds are composed primarily of molecular hydrogen gas, helium, and molecules such as carbon monoxide, hence the term molecular clouds. About one percent by mass of these clouds is in the form of submicrometer-sized dust grains, with about another one percent composed of gaseous molecules and atoms of elements heavier than helium. Regions of active star formation are located within molecular clouds and complexes ranging in mass from a few solar masses ($M_\odot$) to over $10^6 M_\odot$. This association of young stars with molecular clouds is the most obvious manifestation of the fact that stars form from these clouds. Many of the densest regions of these clouds were found to contain embedded infrared objects, i.e., newly formed stars whose light is scattered, absorbed, and reemitted at infrared wavelengths in the process of exiting the placental cloud core. Such cores have already succeeded in forming stars. Initial conditions for the collapse of the presolar cloud can be more profitably ascertained from observations of dense cloud cores that do not appear to contain embedded infrared objects, i.e., precollapse cloud cores.

Precollapse cloud cores are composed of cold molecular gas with temperatures in the range of about 7–15 K, and with gas densities of about 1000–100,000 molecules per cubic centimeter. Some clouds may be denser yet, but this is hard to determine because of the limited density ranges for which suitable molecular tracers are abundant (typically isotopes of carbon monoxide and ammonia). Masses of these clouds range from $\sim 1 M_\odot$ to $10^3 M_\odot$, with the distribution of clump masses fitting a power law such that most of the clumps are of low mass, as is also true of stars in general. In fact, one estimate of the mass distribution of precollapse clouds in Taurus is so similar to the initial mass function of stars that it appears that the stellar mass distribution may be determined primarily by processes occurring prior to the formation of precollapse clouds (Onishi et al., 2002). The cloud properties described below are used to constrain the initial conditions for hydrodynamical models of the collapse of cloud cores. Large radio telescopes have enabled high-spatial-resolution mapping of precollapse clouds and the determination of their interior density structure. While such clouds undoubtedly vary in all three space dimensions, typically the observations are averaged in angle to yield an equivalent, spherically symmetric density profile. These radial density profiles have shown that precollapse clouds typically have flat density profiles near their centers, as is to be expected for a cloud that has not yet collapsed to form a star (Bacmann et al., 2000), surrounded by an envelope with a steeply declining profile that could be fit with a power law. The density profile thus resembles that of a Gaussian distribution, or more precisely, the profile of the Bonnor-Ebert sphere, which is the equilibrium configuration for an isothermal gas cloud (Alves et al., 2001).

While precollapse clouds often have a complicated appearance, attempts have been made to approximate their shapes with simple geometries. Triaxial spheroids seem to be required in general (Kerton et al., 2003), although most lower-mass clouds appear to be more nearly oblate than prolate (Jones et al., 2001). On the larger scale, prolate shapes seem to give a better fit than oblate spheroids. Another study found that the observations could be fit with a distribution of prolate spheroids with axis ratios of 0.54 (Curry, 2002), and argued that the prolate shapes derived from the filamentary nature of the parent clouds. We shall see that the precollapse cloud’s shape is an important factor for the outcome of the protostellar collapse phase.

Precollapse clouds have significant interior velocity fields that appear to be a mixture of turbulence derived from fast stellar winds and outflows, and magneto-hydrodynamic waves associated with the ambient magnetic field. In addition, there may be evidence for a systematic shift in velocities across one axis of the cloud, which can be interpreted as solid-body rotation around that axis. When estimated in this manner, typical rotation rates are found to be below the level needed for cloud support by centrifugal force, yet large enough to result in considerable rotational flattening once cloud collapse begins. Ratios of rotational to gravitational energy in dense cloud cores range from $\sim 0.0001$ to $\sim 0.01$ (Goodman et al., 1993; Caselli et al., 2002). The presence of a net angular momentum for the cloud is essential for the eventual formation of a centrifugally supported circumstellar disk.

2.2. Onset of Collapse Phase

Dense cloud cores are supported against their own self-gravity by a combination of turbulent motions, magnetic fields, thermal (gas) pressure, and centrifugal force, in roughly decreasing order of importance. Turbulent motions inevitably dissipate over timescales that are comparable to or less than a cloud’s freefall time (the time over which an idealized, pressureless sphere of gas of initially uniform density would collapse to form a star), once the source of the turbulence is removed. For a dense cloud core, freefall times are on the order of 0.1 m.y. However, dense
clouds do not collapse on this timescale, because once turbulence decays, magnetic fields provide support against self-gravity.

2.2.1. Ambipolar diffusion and loss of magnetic support. Magnetic field strengths in dense clouds are measured by Zeeman splitting of molecular lines, and found to be large enough (about 10–1000 µG) to be capable of supporting dense clouds, provided that both static magnetic fields and magnetohydrodynamic waves are present (Crutcher, 1999). Field strengths are found to depend on the density to roughly the 1/2 power, as is predicted to be the case if ambipolar diffusion controls the cloud’s dynamics (Mouschovias, 1991). Ambipolar diffusion is the process of slippage of the primarily neutral gas molecules past the ions, to which the magnetic field lines are effectively attached. This process occurs over timescales of a few million years or more for dense cloud cores, and inevitably leads to the loss of sufficient magnetic field support such that the slow inward contraction of the cloud turns into a rapid, dynamic collapse phase, when the magnetic field is no longer in control. This is generally believed to be the process through which stars in regions of low-mass (less than ~8 M\(_\odot\)) star formation begin their life, the “standard model” of star formation (Shu et al., 1987).

2.2.2. Shock-triggered collapse and injection of radioactivity. This standard model has been challenged by evidence for short cloud lifetimes and a highly dynamic star-formation process driven by large-scale outflows (Hartmann et al., 2001). A case for a short lifetime of the protosolar cloud has also been made based on the isotopic records seen in early solar system solids (Russell et al., 2006).

In regions of high-mass (~8 M\(_\odot\) or more) star formation, where the great majority of stars are believed to form, quiescent star formation of the type envisioned in the standard model occurs only until the phase when high-mass stars begin to form and evolve. The process of high-mass star formation is less well understood than that of low-mass stars, but observations make it clear that events such as the supernova explosions that terminate the life of massive stars can result in the triggering of star formation in neighboring molecular clouds that are swept up and compressed by the expanding supernova shock front (Preibisch and Zinnecker, 1999; Preibisch et al., 2002). Even strong protostellar outflows are capable of triggering the collapse of neighboring dense cloud cores (Foster and Boss, 1996; Yokogawa et al., 2003). A supernova shock-triggered origin for the presolar cloud has been advanced as a likely source of the short-lived radioisotopes (e.g., 26Al) that existed in the early solar nebula (Cameron and Truran, 1977; Marhas et al., 2002). Detailed models of shock-triggered collapse have shown that injection of shock-front material containing the 26Al into the collapsing protostellar cloud can occur, provided that the shock speed is on the order of 25 km/s (Boss, 1995), as is appropriate for a moderately distant supernova or for the wind from an evolved red giant star. High-resolution calculations have shown that this injection occurs through Rayleigh-Taylor instabilities at the shock-cloud boundary, where dense fingers grow downward into the presolar cloud (Foster and Boss, 1997; Vanhala and Boss, 2000, 2002). However, these results assume an isothermal shock front, and it remains to be seen what happens when a more detailed thermodynamical treatment is employed.

2.3. Outcome of Collapse Phase

Once a cloud begins to collapse as a result of ambipolar diffusion or triggering by a shock wave, supersonic inward motions develop and soon result in the formation of an optically thick first core, with a size on the order of 10 AU. This central core is supported primarily by the thermal pressure of the molecular hydrogen gas, while the remainder of the cloud continues to fall onto the core. For a 1 M\(_\odot\) cloud, this core has a mass of about 0.0 M\(_\odot\) (Larson, 1969). Once the central temperature reaches about 2000 K, thermal energy goes into dissociating the hydrogen molecules, lowering the thermal pressure and leading to a second collapse phase, during which the first core disappears and a second, final core is formed at the center, with a radius a few times that of the Sun. This core then accretes mass from the infalling cloud over a timescale of about 1 m.y. (Larson, 1969). In the presence of rotation or magnetic fields, however, the cloud becomes flattened into a pancake, and may then fragment into two or more protostars. At this point, we cannot reliably predict what sort of dense cloud core will form in precisely what sort of star or stellar system, much less what sorts of planetary systems will accompany them, but certain general trends are evident.

2.3.1. Multiple fragmentation leading to ejection of single stars. The standard model pertains only to formation of single stars, whereas most stars are known to be members of binary or multiple star systems. There is growing observational evidence that multiple star formation may be the rule, rather than the exception (Reipurth, 2000). If so, then it may be that many (or most) single stars like the Sun are formed in multiple protostar systems, only to be ejected soon thereafter as a result of the decay of the multiple system into an orbitally stable configuration (Bate et al., 2002). In that case, the solar nebula would have been subject to strong tidal forces during the close encounters with other protostars prior to its ejection from the multiple system. This hypothesis has not been investigated in detail (but see Kobrick and Kaula, 1979). Detailed models of the collapse of magnetic cloud cores, starting from initial conditions defined by observations of molecular clouds, show that while initially prolate cores tend to fragment into binary protostars, initially oblate clouds form multiple protostar systems that are highly unstable and likely to eject single protostars and their disks (Boss, 2002a). Surprisingly, magnetic fields were found to enhance the tendency for a collapsing cloud to fragment by helping to prevent the formation of a single central mass concentration of the type assumed to form in the standard model of star formation.
2.3.2. Collapse leading to single star formation. In the case of nonmagnetic clouds, where thermal pressure and rotation dominate, single protostars can result from the collapse of dense cloud cores that are rotating slowly enough to avoid the formation of a large-scale protostellar disk that could then fragment into a binary system (Boss and Myhill, 1995). Alternatively, the collapse of an initially strongly centrally condensed (power-law), nonmagnetic cloud leads to the formation of a single central body (Yorke and Bodenheimer, 1999). In the more idealized case of collapse starting from a uniform density cloud core, fragmentation does not occur prior to reaching densities high enough that the cloud starts heating above its initial temperature if the cloud is initially nearly spherical (Tsuribe and Imotsuka, 1999a,b). Considering that most cloud cores are believed to be supported to a significant extent by magnetic fields, the applicability of all these results is uncertain. In the case of shock-triggered collapse, calculations have shown that weakly magnetic clouds seem to form single protostars when triggering occurs after the core has already contracted toward high central densities (Vanhala and Cameron, 1998).

In the case of the nonmagnetic collapse of a spherical cloud (Yorke and Bodenheimer, 1999), the protostar that forms is orbited by a protostellar disk with a similar mass. When angular momentum is transported outward by assumed gravitational torques, and therefore mass is transported inward onto the protostar, the amount of mass remaining in the disk (~0.5 M☉) at the end of the calculation is still so large that most of this matter must eventually be accreted by the protostar. Hence the disk at this phase must still be considered a protostellar disk, not a relatively late-phase, protoplanetary disk where any objects that form have some hope of survival in orbit. Thus even in the relatively simple case of nonmagnetic clouds, it is not yet possible to compute the expected detailed structure of a protoplanetary disk, starting from the initial conditions of a dense cloud core. Calculations starting from less-idealized initial conditions, such as a segment of an infinite sheet, suffer from the same limitations (Boss and Hartmann, 2001).

Because of the complications of multiple protostar formation, magnetic field support, possible shock-wave triggering, and angular momentum transport in the disk during the cloud infall phase, among others, a definitive theoretical model for the collapse of the presolar cloud has not yet emerged.

2.4. Observations of Star-forming Regions

Observations of star-forming regions have advanced our understanding of the star-formation process considerably in the last few decades. We can now study examples of nearly all phases of the evolution of a dense molecular cloud core into a nearly fully formed star (i.e., the roughly solar-mass T Tauri stars). As a result of being able to observe nearly all phases of the star-formation process, our basic understanding of star formation is relatively mature, at least compared to the planet-formation process, where there are few observations of the phases intermediate between protoplanetary disks and mature planets. Future progress in our understanding of star formation is expected to center on defining the role played by binary and multiple stars and on refining observations of the known phases of evolution. Most (70% or more) stars form in rich clusters of stars where high-mass stars eventually form and shut down the local star-formation process (Lada and Lada, 2003). Examples are the Orion nebula cluster and the Carina nebula (Fig. 1). The remaining stars form in less-crowded regions where only low-mass stars form, such as Taurus and Rho Ophiuchus (Fig. 2). One major question for the origin of the solar system is its birthplace: Was it a region of low- or high-mass star formation? As we shall see, this difference can have profound effects on the protoplanetary disk. Meteoritical evidence may hold the key to deciding between these two different modes of star formation, e.g., with respect to the injection of short-lived radioactivities and photolysis at the disk surfaces driven by ultraviolet radiation from nearby massive stars.

2.4.1. Protostellar phases: Class –I, 0, I, II, III objects. Protostellar evolution can be conveniently subdivided into six phases that form a sequence in time. The usual starting point is the precollapse cloud, which collapses to form the first protostellar core, which is then defined to be a Class –I object. The first core collapses to form the final, second core, or Class 0 object, which has a core mass less than that of the infalling envelope. Class I, II, and III objects (Lada and Shu, 1990) are defined in terms of their spectral energy distributions at midinfrared wavelengths, where the emission is diagnostic of the amount of cold, circumstellar dust. Class I objects are optically invisible, infrared protostars with so much dust emission that the circumstellar gas mass is on the order of 0.1 M☉ or more. Class II objects have less dust emission, and a gas mass of about 0.01 M☉. Class II objects are usually optically visible, T Tauri stars, where most of the circumstellar gas resides in a disk rather than in the surrounding envelope. Class III objects are weak-line T Tauri stars, with only trace amounts of circumstellar gas and dust. While these classes imply a progression in time from objects with more to less gas emission, the time for this to occur for any given object is highly variable: Some Class III objects appear to be only 0.1 m.y. old, while some Class II objects have ages of several million years, based on theoretical models of the evolution of stellar luminosities and surface temperatures. Evidence for dust disks has been found around even older stars, such as Beta Pictoris, with an age of about 10 m.y., although its disk mass is much smaller than that of even Class III objects. Stars with such “debris disks” are often classified as Class III objects.

Multiple examples of all these phases of protostellar evolution have been found, with the exception of the short-lived Class –I objects, which have not yet been detected. It is noteworthy that observations of protostars and young stars find a higher frequency (twice as high in some young clusters) of binary and multiple systems than is the case for mature stars, implying the orbital decay of many of these
Fig. 1. Suspected protoplanetary disks in the Carina nebula, a region of both low- and high-mass star formation. This mosaic image shows that young stars in the Carina nebula have disks that are similar to the “proplyds” (protoplanetary disks) in Orion, except for being even larger in scale — these disks are on the order of 500 AU in radius. The disks are subjected to an intense flux of ultraviolet radiation from the high-mass stars in Carina, which photoevaporates the outer layers of the disks and blows the debris gas and dust away in a comet-like tail (see also Fig. 3). Image courtesy of University of Colorado/NOAO/AURA/NSF.

Fig. 2. Hubble Space Telescope image of a typical protoplanetary disk orbiting a solar-type star in a region of low-mass star formation. This Wide Field and Planetary Camera 2 (WFPC2) image shows the edge-on disk orbiting Herbig-Haro 30 (HH 30). The edge-on, flared disk occults the central star, while bipolar outflows from the protostar emerge perpendicular to the disk. Light from the protostar illuminates the bowl-like surfaces of the disk. Region shown is about 400 AU across. Image courtesy of Chris Burrows (STScI), WFPC2 Science Team, and NASA.
young systems (Reipurth, 2000; Smith et al., 2000), leading to single stars as ejected members (section 2.3).

2.4.2. Ubiquity of bipolar outflows from earliest phases. A remarkable aspect of young stellar objects is the presence of strong molecular outflows for essentially all young stellar objects, even the Class 0 objects. This means that at the same time that matter is still accreting onto the protostar, it is also losing mass through a vigorous wind directed in a bipolar manner in both directions along the presumed rotation axis of the protostar/disk system (Fig. 2). In fact, the energy needed to drive this wind appears to be derived from mass accretion by the protostar, as observed wind momenta are correlated with protostellar luminosities and with the amount of mass in the infalling envelope (Andrè, 1997).

There are two competing mechanisms for driving bipolar outflows, both of which depend on magnetic fields to sling ionized gas outward and to remove angular momentum from the star/disk system. One mechanism is the X-wind model (Shu et al., 2000), where coronal winds from the central star and from the inner edge of the accretion disk join together to form the magnetized X-wind, launched from an orbital radius of a few stellar radii. The other mechanism is a disk wind (Königl and Pudritz, 2000), launched from the surface of the disk over a much larger range of distances, from less than 1 AU to as far away as 100 AU or so. In both mechanisms, centrifugal support of the disk gas makes it easier to launch this material outward, and bipolar flows develop in the directions perpendicular to the disk, because the disk forces the outflow into these preferred directions. Toroidal magnetic fields seem to be required in order to achieve the high degree of collimation of the observed outflows, as a purely poloidal magnetic field would launch the wind at a significant angle away from the rotation axis of the protostar/disk system. Because it derives from radii deeper within the star’s gravitational potential well, an X-wind is energetically favored over a disk wind.

However, observations show that during FU Orionis-type outbursts in young stars, when stellar luminosities increase by factors of 100 in a few years, mass is added onto the central star and from the inner edge of the accretion disk and infalling envelope is removed, however, there is a strong correlation between the amount of mass in the infalling envelope (Andrè, 1997).

All T Tauri stars are believed to experience FU Orionis outbursts [however, see Bell et al. (2000) for a contrary opinion], so disk winds may be the primary driver of bipolar outflows, at least in the early FU Orionis phase of evolution. There is a strong correlation between the amount of mass available for accretion onto the disks and the amount of momentum in the outflow (Bontemps et al., 1996), suggesting that disk mass accretion is directly related to outflow energetics. It is unclear at present what effect an X-wind or a disk wind would have on the planet-formation process, beyond being responsible for the loss of energy (and angular momentum in the latter case), as the winds are thought to be launched either very close to the protostar in the former case, or from the disk’s surface in the latter case. However, they can potentially provide the energy needed for the formation of some of the earliest-formed solids in the solar nebula, such as the refractory inclusions (Ca-Al-rich inclusions, or CAIs) and the chondrules that require locally high temperatures in the nebula (Shu et al., 1996; Connolly et al., 2006).

2.4.3. Lifetimes of circumstellar disks. The most robust constraints on the timescale for disk removal come from astronomical observations of young stars. Because of the limited spatial resolution of current interferometric arrays, molecular hydrogen gas or tracer species such as carbon monoxide cannot be mapped at scales of less than about 10 AU in most disks. Instead, the presence of the gaseous portion of a disk is inferred indirectly by the presence of ongoing mass accretion from the disk onto the star. This accretion leads to enhanced emission in the star’s Hα line, which is then a diagnostic of the presence of disk gas.

Observations of Hα emission in the Orion OB1 association have shown that Hα emission drops toward zero once the stars reach an age of a few 10⁶ yr (Briceno et al., 2001). (This does not mean that solids are not left in orbit around these stars: Particles much larger than 1 μm would be undetectable.) Haisch et al. (2001) studied a number of young clusters and found that the frequencies of disks around the young cluster stars dropped sharply after an age of about 3 × 10⁶ yr. In the Orion nebula cluster, ages for the outer disks are thought to be even shorter, as a result of photoevaporation caused by UV irradiation by newly formed, massive stars (Figs. 1 and 3). Although the inner disks (inside about 10 AU) would be largely unaffected by this process initially. Once the outer disk and infalling envelope is removed, however, there would be no further source of replenishment of the inner disk as it is accreted by the protostar. Hence inner disk lifetimes should be shorter in regions of high-mass star formation (e.g., Orion, Carina) than in regions of low-mass star formation (e.g., Taurus, Ophiuchus). Portions of some disks last for as long as 10⁷ yr; Beta Pictoris is about 10⁷ yr old and has a remnant dust disk that seems to be replenished by collisions between the members of an unseen population of orbiting bodies.

Constraints on the lifetime of the solar nebula have been derived from isotopic studies of early solar system objects (CAIs and chondrules) that are considered to be products of solar nebula processes. The prevalence of ⁶⁸⁹Al with nearly uniform initial abundance in CAIs (Maclepherson et al., 1995) and the much lower abundance of ⁶⁹⁹Al in chondrules have been interpreted as implying that CAI formation in the nebula preceded chondrule formation by at least 10⁶ yr (Russell et al., 1996; Kita et al., 2000; Huss et al., 2001). This interpretation has been bolstered by the U-Pb dating of CAIs and chondrules that suggests a difference of ~2 × 10⁶ yr in absolute ages between these two sets of
objects (Amelin et al., 2002). Furthermore, the presence of a spread in the initial $^{26}$Al abundance in chondrules also requires that the nebular processes leading to chondrule formation lasted more than $10^6$ yr. Although recent observations (Bizzarro et al., 2004) suggest that some chondrules in the Allende meteorite formed contemporaneously with CAIs, formation of other chondrules in the same meteorite continued for at least $1.5 \times 10^6$ yr. Assuming that both the CAIs and chondrules formed within the solar nebula [see Cameron (2003) for a contrary view], the isotopic evidence then implies that the solar nebula had a lifetime of at least a few $10^6$ yr (Russell et al., 2006), similar to the ages inferred for disks around young cluster stars. The simple picture of the solar nebula being removed by a spherically symmetric T Tauri wind has long since been supplanted by the realization that young stars have directed, bipolar outflows that do not sweep over most of the disk. However, mature stars like the Sun do have approximately isotropic winds, so there must be some transition phase where the bipolar star/disk wind evolves into a more spherically symmetric stellar wind. Presumably this enhanced stellar wind would eventually scour any gas and dust from the system left over from the disk accretion phase, when most of the disk gas and dust is accreted by the growing central protostar.

2.4.4. Star-forming regions. While formation as a single star in an isolated, dense cloud core is usually imagined for the presolar cloud, in reality there are very few examples of isolated star formation. Most stars form in regions of high-mass star formation, similar to Orion, with a smaller fraction forming in smaller clusters of low-mass stars, like Taurus or Ophiuchus. The radiation environment differs considerably between these two extremes, with Taurus being relatively benign, and with Orion being flooded with ultraviolet (UV) radiation once massive stars begin to form (Hollenbach et al., 2000). Even in Taurus, though, individual young stars emit UV and X-ray radiations at levels considerably greater than mature stars (Feigelson and Montmerle, 1999). Ultraviolet radiation from the protosun has been suggested as a means of removing the residual gas from the outermost solar nebula (i.e., beyond about 10 AU) through photoevaporation of hydrogen atoms (Shu et al., 1993), a process estimated to require about $10^7$ yr.

In regions of high-mass star formation, the low-mass stars must form first, because once the higher-mass stars begin forming, their intense UV radiation heats the remaining molecular gas and drives it away from the newly formed star cluster. This UV radiation also removes the outermost disk gas from any protoplanetary disks that pass close to the high-mass stars, thereby limiting the lifetimes of these disks and potentially the possibilities for planetary system formation. However, UV radiation from massive stars has been invoked in a positive sense, as a means to photoevaporate the gas in the outer solar nebula (beyond about 10 AU) and then to form the ice giant planets, by photoevaporating the gaseous envelopes of the outermost gas giant protoplanets (Boss et al., 2002). The Carina “proplyds” (Fig. 1) imply a considerably different evolution scenario in the outer disk than is the case for isolated disks in regions like Taurus (Fig. 2), where the disks appear to be classic cases of isolated, symmetric, circumstellar disks with perpendicular outflows, similar to that envisioned in simple theoretical models.

3. EARLY EVOLUTION OF THE SOLAR NEBULA

On theoretical grounds, even an initially highly centrally condensed (i.e., power-law density profile) cloud core is likely to collapse to form a protostar surrounded by a fairly massive (~0.5 $M_\odot$) protostellar disk and envelope. Currently available observations of disks around young stars (e.g., Dutrey et al., 2004) imply that at early ages, disk masses are not always a significant fraction (i.e., 10%) of the protostar’s mass. However, these observations are unable to probe the innermost regions (i.e., within 50 AU or so) because of limited spatial resolution, so the true amount of disk mass at early phases remains uncertain. Nevertheless, the expectation is that protostellar disks must somehow transport most of their mass inward to be accreted by the protostar, eventually evolving into protoplanetary disks, where planetary bodies should be able to form and survive their subsequent interactions with the disk. This process occurs even as collapse of the presolar cloud onto the growing disk continues, adding significant amounts of mass and angular momentum. Observational evidence is beginning to emerge for decreasing disk masses as protostars become older (Eisner and Carpenter, 2003). The transition point from a protostellar disk to a protoplanetary disk is not clear, and the physical mechanisms responsible for disk evolution in either of these two phases remain uncertain, although progress

Fig. 3. Schematic diagram of the process of photoevaporation of the protoplanetary disks in regions of high-mass star formation, such as the Orion nebula cluster and the Carina nebula (see Fig. 1). Image courtesy of Space Telescope Science Institute.
seems to have been made in ruling out several proposed mechanisms.

### 3.1. Angular Momentum Transport Mechanisms

The basic theory of the evolution of an accretion disk can be derived by assuming that there is some physical mechanism operating that results in an effective viscosity of the gas. Because the intrinsic molecular viscosity of hydrogen gas is far too small to have an appreciable effect on disk evolution in a reasonable amount of time (i.e., within $10^7$ yr, given observed disk lifetimes), theorists have sought other sources for an effective viscosity, such as turbulence. In a fully turbulent flow, the effective viscosity can be equal to the molecular viscosity multiplied by a large factor: the ratio of the Reynolds number of the disk (about $10^{10}$) to the critical Reynolds number for the onset of turbulence (about $10^3$), or a factor of about $10^7$. (The Reynolds number is the dimensionless number equal to the product of a mean distance times a mean velocity, divided by the kinematic viscosity of the system.) Under very general conditions, it can be shown (Lynden-Bell and Pringle, 1974) that a viscous disk will evolve in such a manner as to transport most of its mass inward, thereby becoming more tightly gravitationally bound, and minimizing the total energy of the system. In order to conserve angular momentum, this means that angular momentum must be transported outward along with a small fraction of the mass, so that the accretion disk expands outside some radius. The loss of significant angular momentum by centrifugally launched winds somewhat relieves this need for the accretion disk to expand; this additional angular momentum sink was not recognized when the theory was first developed (note, however, that in the case of an X-wind, relatively little angular momentum can be lost by the X-wind). While the basic physics of a viscous accretion disk is fairly well developed, the physical mechanism(s) responsible for disk evolution remain contentious.

#### 3.1.1. Hydrodynamic turbulence

Given the high Reynolds number of a protoplanetary disk, one might expect that a turbulent cascade of energy would occur and result in an effective turbulent viscosity that might be sufficient to drive disk evolution. However, because of the strong differential rotation in a Keplerian disk, a high Reynolds number is not a sufficient condition for fully developed turbulence. Instead, the Rayleigh criterion, which applies to rotating fluids but is not strictly applicable to the solar nebula, suggests that Keplerian disks are stable with respect to turbulence. (The Rayleigh criterion deals with the stability of rotating, incompressible, inviscid fluids to turbulence.)

**3.1.1.1. Vertical convectively-driven turbulence**

While differential rotation may inhibit convective motions in the radial direction in a disk, motions parallel to the rotation axis are relatively unaffected by rotation. In a disk where heat is being generated near the midplane, such as by dissipation of turbulent motions involved in driving mass accretion onto the protostar, and where dust grains are the dominant source of opacity, the disk is likely to be unstable to convective motions in the vertical direction, which carry the heat away from the disk’s midplane and deposit it close to the disk’s surface, where it can be radiated away. Convective instability was conjectured to lead to sufficiently robust turbulence for the resulting turbulent viscosity to be large enough to drive disk evolution (Lin and Papaloizou, 1980), a seemingly attractive, self-consistent scenario that has motivated much of the work on viscous evolution of the solar nebula. However, three-dimensional hydrodynamical models of vertically convectively unstable disks have shown that the convective cells that result are sheared by differential rotation to such an extent that the net transport of angular momentum is very small, and may even be in the wrong direction (see Stone et al., 2000). As a result, convectively-driven disk evolution does not seem to be a major driver. In addition, heating of the surface of the disk by radiation from the central protostar will also act to suppress vertical convection. However, strong evidence of the importance of convection for cooling protoplanetary disks was presented by Boss (2004a).

**3.1.1.2. Rotational shear-induced turbulence**

It has also been suggested that finite amplitude (nonlinear) disturbances to Keplerian flow could result in a self-sustaining shear instability that would produce significant turbulence (Dubrulle, 1993). However, when three-dimensional hydrodynamical models were again used to investigate this possibility, it was found that the initially assumed turbulent motions decayed rather than grew (Stone et al., 2000). Evidently purely hydrodynamical turbulence can neither grow spontaneously nor be self-sustained upon being excited by an external perturbation. However, it has been claimed recently that there may exist other possible paths for hydrodynamic shear to produce turbulence (Chagelishvili et al., 2003), such as the bypass concept, where finite-amplitude initial perturbations undergo transient growth until they are large enough to produce positive feedback.

**3.1.1.3. Global baroclinic instability-driven turbulence**

In spite of these discouraging results for hydrodynamical turbulence, another possibility remains and is under investigation (Klahr and Bodenheimer, 2003; Klahr, 2004), that of a global baroclinic instability. In this mechanism, turbulence results in essence from steep temperature gradients in the radial direction, which then drive centrifugal effects head-on. Three-dimensional hydrodynamical models imply that this mechanism can drive inward mass transport and outward angular momentum transport, as desired. However, similar models by H. Klahr (personal communication, 2003) with a different numerical code (3D-ZEUS) have reached different conclusions, so the situation regarding this mechanism is unclear.

**3.1.1.4. Rossby waves**

Rossby waves occur in planetary atmospheres as a result of shearing motions and can produce large-scale vortices such as the Great Red Spot on Jupiter. Rossby waves have been proposed to occur in the solar nebula as a result of Keplerian rotation coupled with a source of vortices. While prograde rotation (cylonic) vor-
vortices are quickly dissipated by the background Keplerian flow, retrograde (anti-cyclonic) vortices are able to survive for longer periods of time (Godon and Livio, 1999). Rossby waves have been advanced as a significant source of angular momentum transport in the disk (Li et al., 2001). Rossby vortices could serve as sites for concentrating dust particles, but the difficulty in forming the vortices in the first place, coupled with their eventual decay, makes this otherwise attractive idea somewhat dubious (Godon and Livio, 2000). In addition, the restriction of these numerical studies to thin, two-dimensional disk models, where refraction of the waves away from the midplane is not possible, suggests that in a fully three-dimensional calculation, Rossby waves may be less vigorous than in the thin disk calculations (Stone et al., 2000). Recent numerical work (Davis, 2002) suggests that vortices have little long-term effect on the disk.

3.1.2. Magneto-rotational-driven turbulence. While a purely hydrodynamical source for turbulence has not yet been demonstrated, the situation is much different when magnetohydrodynamical (MHD) effects are considered in a shearing, Keplerian disk. In this case, the Rayleigh criterion for stability can be shown to be irrelevant: Provided only that the angular velocity of the disk decreases with radius, even an infinitesimal magnetic field will grow at the expense of the shear motions, a fact that had been noted by Chandrasekhar (1961) but was largely ignored until it was rediscovered some 30 years later.

3.1.2.1. Balbus-Hawley instability: Balbus and Hawley (1991) pointed out that in the presence of rotational shear, even a small magnetic field will grow on a very short timescale. The basic reason is that magnetic field lines can act like rubber bands, linking two parcels of ionized gas. The parcel that is closer to the protosun will orbit faster than the other, increasing its distance from the other parcel. This leads to stretching of the magnetic field lines linking the parcel, and so to a retarding force on the forward motion of the inner parcel. This force transfers angular momentum from the inner parcel to the outer parcel, which means that the inner parcel must fall farther inward toward the protosun, increasing its angular velocity, and therefore leading to even more stretching of the field lines and increased magnetic forces. Because of this positive feedback, extremely rapid growth of an infinitesimal seed field occurs. Consequently, the magnetic field soon grows so large and tangled that its subsequent turbulent evolution must be computed with a fully nonlinear, multidimensional MHD code.

Three-dimensional MHD models of a small region in the solar nebula (Hawley et al., 1995) have shown that, as expected, a tiny seed magnetic field soon grows and results in a turbulent disk where the turbulence is maintained by the magnetic instability. In addition, the magnetic turbulence results in a net outward flow of angular momentum, as desired. The magnetic field grows to a quasi-steady-state value and then oscillates about that mean value, depending on the assumed initial field geometry, which is large enough to result in relatively vigorous angular momentum transport. While promising, these studies of the magneto-rotational instability (MRI) are presently restricted to small regions of the nebula, and the global response of the disk to this instability remains to be determined.

3.1.2.2. Ionization structure and layered accretion: Magneto-rotational instability is a powerful phenomenon, but is limited to affecting nebula regions where there is sufficient ionization for the magnetic field, which is coupled only to the ions, to have an effect on the neutral atoms and molecules. The MRI studies described above all assume ideal MHD, i.e., a fully ionized plasma, where the magnetic field is frozen into the fluid. At the midplane of the solar nebula, however, the fractional ionization is expected to be quite low in the planetary region. Both ambipolar diffusion and resistivity (ohmic dissipation) are effective at limiting magnetic field strengths and suppressing MRI-driven turbulence, but a fractional ionization of only about 1 ion per $10^{12}$ atoms is sufficient for MRI to proceed in spite of ambipolar diffusion and ohmic dissipation. Close to the protosun, disk temperatures are certainly high enough for thermal ionization to create an ionization fraction greater than this, and thus to maintain full-blown MRI turbulence. Given that a temperature of at least 1400 K is necessary, MRI instability may be limited to the innermost 0.2 AU or so in quiescent phases, or as far out as about 1 AU during rapid mass accretion phases (Boss, 1998; Stone et al., 2000).

At greater distances, disk temperatures are too low for thermal ionization to be effective. Cosmic rays were thought to be able to ionize the outer regions of the nebula, but the fact that bipolar outflows are likely to be magnetically driven means that cosmic rays may have a difficult time reaching the disk midplane (Dolginov and Stepinski, 1994). However, the coronae of young stars are known to be prolific emitters of hard X-rays, which can penetrate the bipolar outflow and reach the disk surface at distances of about 1 AU or so, where they are attenuated (Glassgold et al., 1997). As a result, the solar nebula is likely to be a layered accretion disk (Gammie, 1996), where MRI turbulence results in inward mass transport within thin, lightly ionized surface layers, while the layers below the surface do not participate in MRI-driven transport. Thus the bulk of the disk, from just below the surface to the midplane, is expected to be a magnetically dead zone (Fleming and Stone, 2003). Layered accretion is thought to be capable of driving mass inflow at a rate of $\sim 10^{-8} M_\odot$/yr, sufficient to account for observed mass accretion rates in quiescent T Tauri stars (Calvet et al., 2000). Mass accretion rates can vary from less than $10^{-7} M_\odot$/yr to greater than $10^{-6} M_\odot$/yr for stars with ages from $10^6$ to $10^7$ yr, with some evidence for mass accretion rates tapering off for the older stars.

3.1.3. Gravitational torques in a marginally unstable disk. The remaining possibility for large-scale mass transport in the solar nebula is gravitational torques. The likelihood that much of the solar nebula was a magnetically dead zone where MRI transport was ineffective leads to the suggestion that there might be regions where inward MRI mass transport would cease, leading to a local pile-up of mass, which might then cause at least a local gravitational insta-
bility of the disk (Gammie, 1996). In addition, there is observational and theoretical evidence that protostellar disks tend to start their lives with sufficient mass to be gravitationally unstable in their cooler regions, leading to the formation of nonaxisymmetric structure and hence the action of gravitational torques, and that these torques may be the dominant transport mechanism in early phases of evolution (Yorke and Bodenheimer, 1999).

In order for gravitational torques to be effective, a protostellar disk or the solar nebula must be significantly nonaxisymmetric, e.g., threaded by clumps of gas, or by spiral arms, much like a spiral galaxy. In that case, trailing spiral structures, which form inevitably as a result of Keplerian shear, will result in the desired outward transport of angular momentum. This is because in a Keplerian disk, an initial bar-shaped density perturbation will be sheared into a trailing spiral arm configuration. The inner end of the bar rotates faster than the outer end and therefore moves ahead of the outer end. Because of the gravitational attraction between the inner end and the outer end, the inner end will have a component of this gravitational force in the backward direction, while the outer end will feel an equal and opposite force in the forward direction. The inner end will thus lose orbital angular momentum, while the outer end gains this angular momentum. As a result, the inner end falls closer to the protosun, while the outer end moves farther away, with a net outward transport of angular momentum.

3.1.3.1. Rapid mass and angular momentum transport: Models of the growth of nonaxisymmetry during the collapse and formation of protostellar disks show that large-scale bars and spirals can form with the potential to transfer most of the disk angular momentum outward on timescales as short as $10^3$–$10^5$ yr (Boss, 1989), sufficiently fast to allow protostellar disks to transport the most of their mass inward onto the protostar and thereby evolve into protoplanetary disks.

Early numerical models of the evolution of a gravitationally unstable disk (e.g., Cassen et al., 1981) suggested that a disk would have to be comparable in mass to the central protostar in order to be unstable, i.e., in order to have the Toomre Q disk parameter be close to unity (Toomre, 1964). $Q$ is proportional to the product of the disk sound speed times the angular velocity, divided by the surface density of the gas; low Q means that the disk is massive enough to overcome the stabilizing effects of thermal gas pressure and shearing motions (Keplerian rotation). Gravitational instability could then occur in protostellar, but not protoplanetary disks. Analytical work on the growth of spiral density waves implied that for a 1 M$_\odot$ star, gravitational instability could occur in a disk with a mass as low as 0.19 M$_\odot$ (Shu et al., 1990). Three-dimensional hydrodynamical models have shown that vigorous gravitational instability can occur in a disk with a mass of 0.1 M$_\odot$ or even less, in orbit around a 1 M$_\odot$ star (Boss, 2000), because of the expected low midplane temperatures (about 30 K) in the outer disk implied by cometary compositions (Kawakita et al., 2001) and by observations of disks (D’Alessio et al., 2001). Similar models with a complete thermodynamical treatment (Boss, 2002b), including convective transport and radiative transfer, show that a marginally gravitationally unstable solar nebula develops a robust pattern of clumps and spiral arms, persisting for many disk rotation periods, and resulting in episodic mass accretion rates onto the central protosun that vary from $\sim 10^{-7}$ M$_\odot$/yr to as high as $\sim 10^{-3}$ M$_\odot$/yr. The latter rates are more than high enough to account for FU Orionis outbursts (Bell et al., 2000).

Because angular momentum transport by a strongly gravitationally unstable disk is so rapid, it is unlikely that protostellar or protoplanetary disks are ever strongly gravitationally unstable, because they can probably evolve away from such a strongly unstable state faster than they can be driven into it by, e.g., accretion of more mass from an infalling envelope or radiative cooling. As a result, it is much more likely that a disk will approach gravitational instability from a marginally unstable state (Cassen et al., 1981). Accordingly, models of gravitationally unstable disks have focused almost exclusively on marginally gravitationally unstable disks (e.g., Boss, 2000), where primarily the outer disk, beyond about 5 AU, participates in the instability. Inside about 5 AU, disk temperatures appear to be too high for an instability to grow there, although these inner regions may still be subject to shock fronts driven by clumps and spiral arms in the gravitationally unstable, outer region. One-armed spiral density waves can propagate right down to the stellar surface.

3.1.3.2. Global process vs. local viscosity: Gravitational forces are intrinsically global in nature, and their effect on different regions of the nebula can be expected to be highly variable in both space and time. On the other hand, turbulent viscosity is a local process that is usually assumed to operate more or less equally efficiently throughout a disk. As a result, it is unclear if gravitational effects can be faithfully modeled as a single, effective viscosity capable of driving disk evolution in the manner envisioned by Lynden-Bell and Pringle (1974). Nevertheless, efforts have been made to try to quantify the expected strength of gravitational torques in this manner (Lin and Pringle, 1987). Three-dimensional models of marginally gravitationally unstable disks imply that such an effective viscosity is indeed large and comparable to that in MRI models (Laughlin and Bodenheimer, 1994).

3.2. Models of Solar Nebula Evolution

Given an effective source of viscosity, in principle the time evolution of the solar nebula can be calculated in great detail, at least in the context of the viscous accretion disk model. The strength of an effective viscosity is usually quantified by the $\alpha$ parameter. $\alpha$ is often defined in various ways, but typically $\alpha$ is defined to be the constant that, when multiplied by the sound speed and the vertical scale height of the disk (two convenient measures of a typical velocity and length scale), yields the effective viscosity of the disk (Lynden-Bell and Pringle, 1974). Three-dimensional MHD
models of the MRI imply an typical MRI $\alpha$ of about 0.005–0.5 (Stone et al., 2000). Similarly, three-dimensional models of marginally gravitationally unstable disks imply an $\alpha$ of about 0.03 (Laughlin and Bodenheimer, 1994). Steady mass accretion at the low rates found in quiescent T Tauri stars requires an $\alpha$ of about 0.01 (Calvet et al., 2000), in rough agreement with these estimates. Once planets have formed and become massive enough to open gaps in their surrounding disks, their orbital evolution becomes tied to that of the gas. As the gaseous disk is transported inward by viscous accretion, these planets must also migrate inward. The perils of orbital migration for planetary formation and evolution are addressed in Ward and Hahn (2000) and Lin et al. (2000). Here we limit ourselves to considering the evolution of dust and gas prior to the formation of planetary-sized bodies.

3.2.1. Viscous accretion disk models. The generation of viscous accretion disk models was an active area of research during the period when convective instability was believed to be an effective source of viscosity. Ruden and Pollack (1991) constructed models where convective instability was assumed to control the evolution, so that in regions where the disk became optically thin and thus convectively stable, the effective viscosity vanished. Starting with an $\alpha$ of about 0.01, they found that disks evolved for about $10^6$ yr before becoming optically thin, often leaving behind a disk with a mass of about 0.1 $M_{\odot}$. Midplane temperatures at 1 AU dropped precipitously from about 1500 K initially to about 20 K when convection ceased and the disk was optically thin at that radius. Similarly dramatic temperature drops occur throughout the disk in these models, and the outer regions of the models eventually became gravitationally unstable as a result. Given that convective instability is no longer considered to be a possible driver of disk evolution, the Ruden and Pollack (1991) models are interesting, but not likely to be applicable to the solar nebula. Unfortunately, little effort has gone into generating detailed viscous accretion models in the interim: The theoretical focus seems to have been more on the question of determining which mechanisms are contenders for disk evolution than on the question of the resulting disk evolution. In particular, the realization that the MRI mechanism is likely to have operated only in the magnetically active surface layers of the disk, and not in the magnetically dead bulk of the disk, presents a formidable technical challenge for viscous accretion disk models, which have usually been based on the assumption that the nebula can be represented by a thin, axisymmetric disk (e.g., Ruden and Pollack, 1991), greatly simplifying the numerical solution. The need for consideration of the vertical as well as the radial structure of the disk, and possibly the azimuthal (nonaxisymmetric) structure as well, points toward the requirement of a three-dimensional magnetohydrodynamical calculation of the entire disk. Such a calculation has not been performed, and even the MRI calculations performed to date on small regions of a disk can only be carried forward in time for a small fraction of the expected lifetime of the disk.

Some progress has been made in two-dimensional hydrodynamical models of a thick disk evolving under the action of a globally defined $\alpha$ viscosity, representing the effects of torques in a marginally gravitationally unstable disk (Yorke and Bodenheimer, 1999), but in these models the evolution eventually slows down and leaves behind a fairly massive protostellar disk after $10^7$ yr, with a radius on the order of 100 AU.

One aspect of particular interest about viscous accretion disk models is the evolution of solid particles, both in terms of their thermal processing and in terms of their transport in the nebula. Interstellar dust grains are small enough (sub-micrometer-sized) to remain well-coupled to the gas, so they will move along with the gas. During this phase, the gas and dust may undergo trajectories that are outward at first, as the disk accretes matter from the infalling envelope and expands by outward angular momentum transport, followed by inward motion once accretion stops and the disk continues to accrete onto the protostar (Cassen, 1996). Once collisional coagulation gets underway and grain growth begins, solid particles begin to move with respect to the gas, suffering gas drag and additional radial migration as a result (Weidenschilling, 1988, 2004; Weidenschilling and Cuzzi, 2006; Cuzzi and Weidenschilling, 2006), with the peak velocity with respect to the gas occurring for roughly meter-sized solids. However, this does not mean that all growing solids will be lost to monotonic inward migration toward the protosun, as the mixing and transport processes in a marginally gravitationally solar nebula are sufficiently robust to mix solids over distances of several AU within $10^3$ yr (Boss, 2004b). A strong source of generic $\alpha$ turbulent viscosity would contribute to this mixing. Particles that are centimeter-sized or smaller will be effectively tied to the gas during this process. In addition, in a disk with spiral arms, the meter-sized solids are forced by gas drag to move toward the centers of the arms rather than toward the protosun, thereby enhancing their chances for further growth to sizes large enough to become completely decoupled from the gas (Haghighipour and Boss, 2003).

3.2.1.1. Volatility patterns and transport in the inner solar system: The bulk compositions of the bodies in the inner solar system show a marked depletion of volatile elements compared to the solar composition. Cassen (2001) has shown that these volatile depletions can be explained as a result of the condensation of hot gases and coagulation of the resulting refractory dust grains into solids that are decoupled from the gas through the rapid growth of kilometer-sized planetesimals. The volatile elements remain in gaseous form at these temperatures, and so avoid being incorporated into the planetesimals that will eventually form the terrestrial planets and asteroids. In order for this process to work, significant regions of the nebula must have been hot enough at the midplane to keep volatiles in the gaseous form, a situation that would characterize the nebula when mass accretion rates were on the order of $10^{-7} M_{\odot}$/yr or less. The volatile gases would then be removed from the terrestrial planet region by viscous accretion onto the proto-
sun. The postulated rapid growth from dust grains to kilometer-sized bodies in \( \sim 10^5 \) yr required by this scenario appears to be possible (Woolum and Cassen, 1999). Cuzzi and Weidenschilling (2006) argue, however, that such rapid growth may not be possible in a nebula with even a low level of generic turbulence.

3.2.2. Clump formation in a marginally gravitationally unstable disk. Given the apparent limitation of MRI-driven accretion to the surfaces of protoplanetary disks, it would appear that gravitational torques may have to be responsible for the evolution in the bulk of the disk. In addition, there are strong theoretical reasons why gravitational torques may be effective, including the difficulty in forming the gas giant planets by the conventional means of core accretion. The standard model for Jupiter formation by core accretion envisions a nebula that has a surface density of solids high enough for a solid core to form within a few \( 10^6 \) yr through runaway accretion (Pollack et al., 1996) and then accrete a massive gaseous envelope. However, the gas in such a nebula is likely to be marginally gravitationally unstable, a situation that could result in the rapid formation of gas giant planets in \( \sim 10^3 \) yr by the formation of self-gravitating clumps of gas and dust (Boss, 1997, 2000, 2002b). Deciding between these two differing modes of giant planet formation is difficult based on present observations — dating the epoch of giant-planet formation would help to constrain the problem. The ongoing census of extrasolar planetary systems seems to be showing that Jupiter-mass planets are quite common (Marcy et al., 2000), so the giant planet formation mechanism appears to be a robust one. One extrasolar planet survey has found evidence that roughly 15% of nearby stars like the Sun are orbited by Jupiter-mass planets with orbital periods less than 3 yr, while another 25% show evidence for having longer-period, Jupiter-like companions (A. Hatzes, personal communication, 2004). Thus at least 40% of solar-type stars might be orbited by gas giant planets.

In their pioneering study of a marginally unstable disk, Laughlin and Bodenheimer (1994) found strong spiral arm formation but no clumps, presumably in large part as a result of the limited spatial resolution that was computationally possible at the time (up to 25,000 particles in a smoothed particle hydrodynamics code). Boss (2000) has shown that when \( 10^6 \) or more grid points are included in a finite-differences calculation, three-dimensional hydrodynamic models of marginally gravitationally unstable disks demonstrate the persistent formation of self-gravitating clumps, although even these models do not appear to have sufficient spatial resolution to follow the high-density clumps indefinitely in time. Regardless of whether or not such disk instability models can lead to gas-giant-planet formation, the likelihood that the solar nebula was at least episodically marginally gravitationally unstable has important implications for cosmochemistry.

3.2.2.1. Chondrule formation in nebular shock fronts: Perhaps the most well-known, unsolved problem in cosmochemistry is the question of the mechanism whereby dust grain aggregates were thermally processed to form chondrules, which are abundant in all the chondritic meteorites, and the somewhat rare CAIs, present primarily in primitive meteorites. Chondrule compositions and textures in particular require rapid heating and somewhat slower cooling for their explanation; a globally hot nebula is inconsistent with these requirements (Cassen, 2001). A wide variety of mechanisms has been proposed and generally discarded, but theoretical work seems to suggest that chondrule formation in nebular shock fronts is a very plausible mechanism (Desch and Connolly, 2002; Connolly et al., 2006).

In a marginally gravitationally unstable nebula, clumps and spiral arms at about 8 AU will drive one-armed spiral arms into the inner nebula, which at times results in shock fronts oriented roughly perpendicular to the orbits of bodies in the asteroidal region (Boss and Durisen, 2005; Fig. 4). Because of the tendency toward co-rotation in self-gravitating structures, this will lead to solids encountering a shock front at speeds as high as 10 km/s, sufficiently high to result in postshock temperatures of about 3000 K. Detailed one-dimensional models of heating and cooling processes in such a shock front have shown that shock speeds around 7 km/s are optimal for matching chondrule cooling rates and
therefore textures (Desch and Connolly, 2002). Shocks sufficiently strong to melt chondrule precursors are transient in nature and it remains to be seen how often the required conditions are met at 2.5 AU; continuous shock processing is not desired to explain chondrule properties. The chondrules would be transported both inward and outward by the chaotic spiral density waves (Boss, 2004b). Once Jupiter forms, it will continue to drive shock fronts in the asteroidal region capable of forming chondrules, for as long as the inner disk gas and dust remains. The meteoritical evidence (Russell et al., 2006) is that chondrule formation and thus the existence of the inner disk persisted for a couple of million years.

3.2.2.2. Mixing processes in gravitationally unstable disks: If disk evolution near the midplane is largely controlled by gravitational torques rather than by a turbulent process such as MRI or convection, then mixing processes might be profoundly different as a result. Gravitational torques could potentially result in matter flowing through the disk without being rapidly homogenized through mixing by turbulence. As a result, spatially heterogeneous regions of the disk might persist for some amount of time, if they were formed in the first place by processes such as the triggered injection of shock-wave material (Vanhala and Boss, 2000, 2002) or the spraying and size-sorting of solids processed by an X-wind onto the surface of the nebula (Shu et al., 1996). However, because convective motions appear to play an important part in cooling the disk midplane in recent models of disk instability (Boss, 2002b), it is unclear if gravitational torques could act in isolation without interference from convective motions or other sources of turbulence. At any rate, spatially heterogeneous regions might only last for short fraction of the nebular lifetime, requiring rapid coagulation and growth of kilometer-sized bodies if evidence of this phase is to be preserved.

4. CONCLUSIONS

While the meteoritical record provides a rich resource for understanding the early evolution of the solar nebula, eventually one wishes for a more universal theory of protoplanetary disk evolution. Our current understanding of protoplanetary disks in nearby star-forming regions is severely hampered by our inability to resolve processes occurring on the scale of a few AU at distances of 150 pc or more. The advent of the Atacama Large Millimeter Array (ALMA) toward the end of this decade should help alter this situation, as ALMA will have sufficient resolution to begin the probe the innermost regions of protoplanetary disks. ALMA will answer basic questions such as determining the density, temperature, and rotational profiles of the molecular gas in these disks over length scales of a few AU. These studies will be important for determining the astrophysical context for many of the processes inferred to have occurred from the meteoritical record. At the other end of the mass spectrum, we will also be learning in the next decade about the prevalence of Earth-like planets, through the ambitious space telescopes planned by NASA to search for Earth-like planets: the Kepler mission, the Space Interferometry mission, and the Terrestrial Planet Finder. These missions will tell us something about the cosmic efficiency of the collisional accumulation process that produced the planets of the inner solar system, yet failed so spectacularly in the asteroid belt, thereby preserving the window to our own past provided by studies of the most primitive meteorites.

Acknowledgments. This work was partially supported by the NASA Planetary Geology and Geophysics Program under grant NAG 5-10201.

REFERENCES

Boss A. P. (2002b) Evolution of the solar nebula. V. Disk insta-
Tsuribe T. and Inutsuka S.-I. (1999a) Criteria for fragmentation


