The Astrophysical Journal, 609:1045–1064, 2004 July 10

THE EVOLUTION OF GRAVITATIONALLY UNSTABLE PROTOPLANETARY DISKS: FRAGMENTATION AND POSSIBLE GIANT PLANET FORMATION

Lucio Mayer,[1] Thomas Quinn,[2] James Wadsley,[3] and Joachim Stadel[1]

Received 2003 October 22; accepted 2004 March 24

ABSTRACT

We carry out a large set of very high resolution, three-dimensional, smoothed particle hydrodynamics simulations describing the evolution of gravitationally unstable gaseous protoplanetary disks. We consider a broad range of initial disk parameters. Disk masses out to 20 AU range from 0.075 to 0.125 $M_{\odot}$, roughly consistent with the high end of the mass distribution inferred for disks around T Tauri stars. Minimum outer temperatures range from 30 to 100 K, as expected from studies of the early protosolar nebula and suggested by the modeling of the spectra of protoplanetary disks. The mass of the central star is also varied, although it is usually assumed to be equal to that of the Sun. Overall, the initial disks span minimum Q-parameters between 0.8 and 2, with most models having $Q \sim 1.4$. The disks are evolved assuming either a locally isothermal equation of state or an adiabatic equation of state with varying γ. Heating by (artificial) viscosity and shocks is included when the adiabatic equation of state is used. When condensations above a specific density threshold appear as a result of gravitational instability in a locally isothermal calculation, the equation of state is switched to adiabatic to account for the increased optical depth. We show that when a disk has a minimum Q-parameter less than 1.4, strong trailing spiral instabilities, typically three- or four-armed modes, form and grow until fragmentation occurs along the arms after about 5 mean disk orbital times. The resulting clumps contract quickly to densities several orders of magnitude higher than the initial disk density, and the densest of them survive even under adiabatic conditions. These clumps are stable to tidal disruption and merge quickly, leaving two to three protoplanets on fairly eccentric orbits (the mean eccentricity being around 0.2) after $\sim 10^3$ yr. Fragmentation is not strongly dependent on whether the disk starts from a marginally unstable state or gradually achieves it; we show that if the disk is allowed to grow in mass from a very light, very stable state over tens of orbital times, it still fragments at roughly the same mass and temperature as in the standard disk models. We show that the first stages of the instability, until the appearance of the overdensities, can be understood in terms of the maximum unstable Toomre wavelength and the local Jeans length. A high mass and force resolution are needed to correctly resolve both scales and follow the fragmentation process appropriately. Varying disk mass and temperature affects such physical scales and hence the typical masses of the protoplanets that form. Objects smaller than Saturn or a couple of times bigger than Jupiter can both be produced by fragmentation. Their final masses will then depend on the subsequent interactions and mergers with other clumps and on the accretion of disk material. The accretion rate depends on the disk thermodynamics and is negligible with adiabatic conditions. After $\sim 10^3$ yr the masses range from just below $1 M_{\rm Jup}$ to more than $7 M_{\rm Jup}$, well in agreement with those of detected extrasolar planets.

Subject headings: accretion, accretion disks — hydrodynamics — methods: *n*-body simulations — planetary systems: formation — planetary systems: protoplanetary disks — solar system: formation

On-line material: color figures

1. INTRODUCTION

The rapid formation of gas giant planets by gravitational instabilities in a protoplanetary disk (Kuiper 1951; Cameron 1978; Boss 1997) is an appealing alternative to the conventional scenario of accretion of gas onto preexisting large rocky cores formed by accumulation of planetesimals (Wetherill 1990). The latter scenario seems to require timescales well in excess of disk survival times in dense, highly irradiated environments, such as the Orion Nebula, where most of the stars in our galaxies are born (Throop et al. 2001), making giant planet formation a rare occurrence. Protoplanetary disks in lower density environments have lifetimes at most marginally consistent with the few millions of years required to form a Jupiter-sized planet in the core-accretion scenario at several AU from their star (Pollack et al. 1996; Hubickyj et al. 2002; Briceño et al. 2001; Haisch et al. 2001). However, even in the most favorable scenario, it is difficult to imagine how planets with masses as large as several Jupiter masses, such as many of the observed extrasolar planets (Marcy et al. 2000; Mayor et al. 1999), might also be produced in only a few million years. The problem is not simply that the planet would migrate inward faster than it could accrete enough mass (Nelson et al. 2000b; Bate et al. 2003). Indeed, inward type II migration might be stopped or reversed because of either corotational or Lindblad torques once more realistic disks with profiles that are not simple power laws are considered (Masset & Papaloizou 2003; P. Artymowicz & W. Peplinski 2004, in preparation), but the mass doubling time of a Jupiter-sized

[1] Institute of Theoretical Physics, University of Zürich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; lucio@physik.unizh.ch.

[2] Department of Astronomy, University of Washington, Seattle, WA 98195; trq@astro.washington.edu.

[3] Department of Physics and Astronomy, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4M1, Canada; wadsley@physics.mcmaster.ca.

planet after a gap has been opened is of the order of 1 Myr, even in significantly viscous disks (Artymowicz et al. 1998).

Very recent calculations that solve the magnetohydrodynamic equations on a grid suggest that such gaps might be even deeper, and hence the accretion of gas through the gap might be further slowed down, if magnetic turbulence is the main source of viscosity in disks (Nelson & Papaloizou 2003). Therefore, the disk might be dissipated well before a planet can grow up to the many Jupiter masses of some of the extrasolar planets (Bate et al. 2003). In addition, the evidence for inner rocky cores within the solar system giants, an inevitable prediction of the core-accretion mechanism, is weakening, since Jupiter might not have a solid core at all (Guillot 1999a, 1999b). In addition, for the transiting extrasolar giant HD 209458b, for which the planetary radius and mass are known (Charbonneau et al. 2000), models of the planet's interior are generally consistent with the absence of a core (e.g., Guillot & Showman 2002). The overall amount of metals in Jupiter and Saturn is significantly higher than solar, but this does not necessarily reflect the initial metal content of the planets (Boss 1998).

In any case, current models of the core-accretion mechanism need a surface density of solids 3–4 times in excess of the minimum-mass solar nebula model (Weidenschilling 1977) for the rocky cores of giant planets to form before 10 million yr (Lissauer 1993); if augmented by 100 times more mass in molecular hydrogen, such a protosolar nebula model will indeed be marginally gravitationally unstable. Gravitational instability will then become the prevailing formation mechanism because it takes as little as 1000 yr (Boss 1997, 2000, 2001). On the other hand, gravitational instabilities in a protoplanetary disk are difficult to treat correctly because of both various numerical pitfalls that can arise in the simulations and the difficulty of accounting properly for all the cooling and heating mechanisms present in real protoplanetary disks (Pickett et al. 1998, 2000a, 2000b, 2003).

Because of the complexity of the problem, the natural first step is to adopt a simple thermodynamic description of the disk and use a very high resolution simulation to probe in great detail the highly nonlinear dynamics associated with gravitational instability. In Mayer et al. (2002) we showed that, under certain conditions, a system of gas giants can arise whose properties are reminiscent of those of known extrasolar planetary systems. Essential to this result was the ability to achieve very high spatial and mass resolution thanks to the fast parallel Tree+SPH code GASOLINE (Wadsley et al. 2004). Here we describe the results of a much larger suite of simulations, exploring a wide parameter space in terms of both disk structural properties and thermodynamics, as well as addressing in detail the various numerical aspects of the calculations and how these can affect the final outcome of the gravitational instability. We also discuss the reliability of the initial conditions used in the simulations and the structural properties of the protoplanets formed in some of the runs. In a forthcoming paper we will discuss the effects of both irreversible heating and radiative cooling in high-resolution disk simulations.

2. INITIAL DISK MODELS

The self-gravitating disk models are a collection of 200,000 or 1 million smoothed particle hydrodynamics (SPH) particles rotating around a central star. Disks have masses ranging between 0.075 and 0.125 $M_{\odot}$, comparable to the most massive among T Tauri disks (Beckwith et al. 1990; Dutrey et al. 1996) and 3–4 times more massive than the minimum-mass solar nebula (Weidenschilling 1977). However, these masses are still lower than those expected in the early stages of the formation of a star-disk system from infalling molecular cloud material, as suggested by hydrodynamic simulations (Pickett et al. 1998, 2000a). To generate the initial conditions we use a modified version of the algorithm written by Hernquist (1993) to construct equilibrium N-body realizations of multicomponent collisionless and gaseous systems under their mutual gravitational potential (see also Quinn et al. 1993). The positions of gas particles are initialized randomly according to a prescribed three-dimensional disk density profile. The three-dimensional density profile is specified by choosing a form for the surface density profile and then assuming vertical hydrostatic equilibrium for a given temperature profile $T(r)$. Disks extend from $r_{\rm in} = 4$ AU to $r_{\rm out} = 20$ AU and have a surface density profile of the form

$$\Sigma(r) = \Sigma_c \exp\left[-\left(\frac{r_{\rm in}}{r}\right)^2\right] r^{-1} \exp\left[-\left(\frac{r}{r_{\rm out}}\right)^2\right],$$

where Σ_c is a normalization constant that is set once the disk mass is chosen. For most of the disk extent, the slope of the adopted profile is similar to the standard $\Sigma(r) \sim r^{-1.5}$ profile assumed in models of the protosolar nebula (Weidenschilling 1977); however, we opted for this more complicated form with exponential cutoffs to avoid too-sharp discontinuities at the boundaries and aid numerical stability. Gas particles are set in pure rotation around the central star. The central star is a softened point mass and can wobble in response to the time-dependent disk potential. The mass of the star M_s is usually equal to 1 $M_{\odot}$ (see Table 1), and its softening is set equal to 2 AU. We use a spline kernel softening (Hernquist & Katz 1989) for both the central star and the gas particles; with a spline softening the potential of a particle converges to the Keplerian value at two softening lengths. The rotational velocities of gas particles are calculated assuming centrifugal equilibrium in the combined potential of the central softened point mass and the disk. The nearly Keplerian velocity profile resulting from this procedure is still not in perfect equilibrium, since it needs to be adjusted to account for pressure forces. Therefore, we modify the initial disk velocity profile until we converge to a model as close as possible to equilibrium for the desired set of parameters. We do this by performing trial runs with very light disks (this way self-gravity plays a negligible role in the evolution and changes in disk structure are due to deviations from perfect equilibrium in the initial conditions). We do not apply any boundary conditions, a difference from most previous works on the subject (e.g., Boss 1998, 2002a; Nelson et al. 1998, 2000a), so disks are free to expand and, at the same time, material can accrete onto the central star because of transfer of angular momentum resulting from artificial viscosity and eventual nonaxisymmetric instabilities. The disks are not seeded with nonaxisymmetric perturbations as normally done in grid-based simulations (e.g., Boss 1998); Poisson noise is indeed present in an SPH simulation even at fairly high resolution (at a level of 0.1% in our 1 million particle runs, comparable to the seeds in Boss's models) because of the discrete representation of the system.

The softening of gas particles, ϵ_s, is constant with time in our runs, while in some of the previous SPH simulations of disk instability it was evolving (typically decreasing) with the smoothing length (e.g., Nelson et al. 1998, 2000a). Our

TABLE 1
PARAMETERS OF THE SIMULATIONS

Run (1)	M_d/M_s (2)	EOS (3)	α (4)	β (5)	N_p (6)	$T_{\min}$ (K) (7)	Softening (AU) (8)	M_s ($M_\odot$) (9)	Q (10)	Clumps (11)
DISL1.....	0.1	ISO	1	2	2×10^5	56	0.12	1	1.38	Yes
DISL1b.....	0.1	ISO	1	2	2×10^5	56	0.18	1	1.38	No
DISL1c.....	0.1	ISO	1	2	2×10^5	56	0.6	1	1.38	No
DISL1d.....	0.1	ISO	1	2	2×10^5	56	0.06	1	1.38	Yes
DISL1e.....	0.1	ISO	1	2.5	2×10^5	56	0.12	1	1.38	Yes
DISL1f.....	0.1	ISO	1	3	2×10^5	56	0.12	1	1.38	No
DISL1g.....	0.1	ISO	1	6	2×10^5	56	0.12	1	1.38	No
DISL1h.....	0.1	ISO	1	0.5	2×10^5	56	0.12	1	1.38	Yes
DISL1i.....	0.1	ISO	0	0.5	2×10^5	56	0.12	1	1.38	Yes
DISL2.....	0.1	ISO	1	2	2×10^5	100	0.06	1	2	No
DISH1	0.1	ISO	1	2	10^6	56	0.06	1	1.38	Yes
DISH2	0.08	ISO	1	2	10^6	56	0.06	1	1.65	No
DISH2b	0.08	ISO	0	0.5	10^6	56	0.06	1	1.65	No
DISH2c.....	0.08	ISO	0	0.5	10^6	56	0.006	1	1.65	No
DISH3	0.085	ISO	1	2	10^6	36	0.06	1	1.3	Yes
DISH3b	0.085	ISO	1	2	10^6	50	0.06	1	1.5	No
DISgr.....	0.0085	ISO	1	2	2×10^5	30(gr)	0.06	1	$Q(t)$	Yes
DISH4	0.075	ISO	1	2	10^6	56	0.06	1	1.9	No
DISH4b	0.075	ISO	1	2	10^6	56	0.006	1	1.9	No
DISL3.....	0.075	ISO	1	2	2×10^5	56	0.06	1	1.9	No
DISL4.....	0.075	ISO	1	2	2×10^5	56	0.06	0.5	1.38	Yes
DISLad1	0.1	1.4	1	2	2×10^5	56	0.06	1	1.38	No
DISLad2	0.125	1.4	1	2	2×10^5	20	0.06	1	0.8	No
DISLad3	0.125	1.3	1	2	2×10^5	20	0.06	1	0.8	Transient
DISLad4	0.125	1.2	1	2	2×10^5	20	0.06	1	0.8	Yes
DISLad5	0.1	1.4	1	0.5	2×10^5	56	0.06	1	1.38	No
DISLad6	0.1	1.4	0	0.5	2×10^5	56	0.06	1	1.38	No

NOTES.—Col. (1): Run name. Col. (2): Initial disk mass as a fraction of stellar mass. Col. (3): Equation of state (ISO for isothermal; otherwise γ is indicated). Col. (4): α-parameter in artificial viscosity. Col. (5): β-parameter in artificial viscosity. Col. (6): Number of disk particles. Col. (7): Minimum temperature. Col. (8): Softening of disk particles. Col. (9): Mass of central star. Col. (10): Toomre Q-parameter. Col. (11): Outcome of the simulations in terms of clump formation ("Yes" for formation of gravitationally bound clumps, "No" for no fragmentation at all, or "Transient" for transient clumps).

implementation is preferred in most astrophysical problems, as it avoids the unphysical situation of having particles with varying gravitational potential energy. A good choice for the softening is one that allows a gravitational force resolution initially close to the resolution of pressure forces; hence, $\epsilon_s \sim h$, where h is the SPH smoothing length calculated over 32 neighbors (Bate & Burkert 1997). In the latter case the softening scales with the mass of the gas particles m_g as $m_g^{1/3}$, as does the SPH smoothing length. As a result, we use smaller softenings in simulations with higher mass resolution. The latter type of scaling between softening and mass is widely adopted and has been repeatedly shown to be reliable in hydrodynamic simulations of structure formation (e.g., Hernquist & Katz 1989; Thacker et al. 2000). As explained in Bate & Burkert (1997), however, the gravitational softening must be small enough in order to resolve the local Jeans mass even in high-density regimes if collapse or fragmentation is the aim of our calculation. The Jeans mass will decrease if the local density increases because of gravitational instabilities, and the SPH smoothing length will also decrease correspondingly. Therefore, in order to leave room for better force resolution in the later stages of the simulations, we choose $\epsilon_s \sim 0.5h$ (where h is the initial smoothing length) in the majority of our simulations. However, in a few runs we explore the effect of choosing a bigger or smaller softening, and the results are discussed in § 4.3. Gravitational forces between distant particles are calculated using the hierarchical tree method with opening angle parameter $\theta = 0.7$ (Barnes & Hut 1986). In particular, GASOLINE uses a binary tree, and multipole expansions are carried out up to the hexadecapole order (see Stadel 2001 for details). The code uses multistepping to advance particles in an efficient way with a leapfrog integrator (see Wadsley et al. 2004).

In most of the simulations the disk is initialized with a mass and temperature profile so as to obtain a desired minimum Toomre Q-parameter in the outer, colder part of the disk, similar to what was done in Boss (2001, 2002a, 2002b). We recall that $Q = \Omega v_s/\pi G\Sigma$ for a thin disk in Keplerian rotation (Toomre 1964), where Ω is the angular velocity, Σ is the disk surface density, v_s is the sound speed (which is proportional to $T^{1/2}$, where T is the temperature), and G is the gravitational constant. Figure 1 shows examples of Q-profiles in some of our models. We also performed a run in which the disk approaches a given minimum Toomre parameter, $Q_{\min}$, from a very stable state (high Q), starting from a very low initial mass (run DISLgr in Table 1) that is grown over time by an order of magnitude. This simulation allows us to test the dependence of our results on the way the standard initial conditions are set up. The shape of the temperature profile is the same for each model (Fig. 2) and is similar to that used by Boss (1998, 2001). The temperature depends only on radius; thus, there is no difference between the midplane and an atmosphere. While Boss (1998) uses a combination of different power laws for the temperature profile, each of them holding in a particular

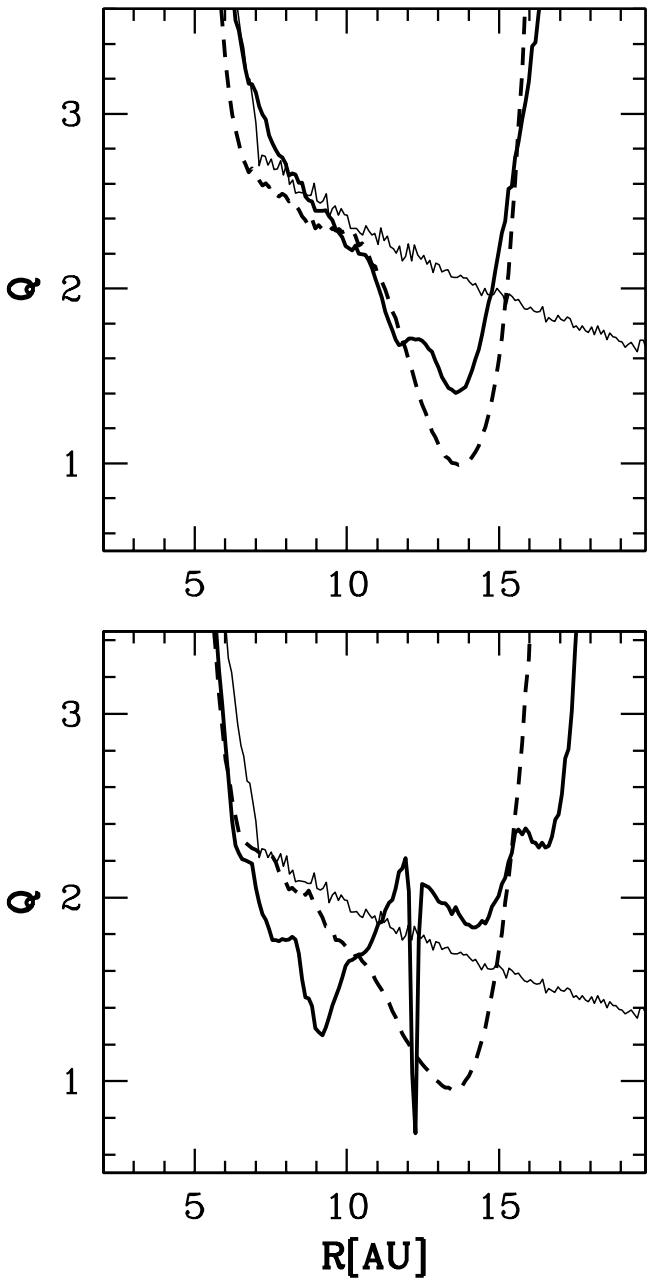

FIG. 1.—Evolution of Q-profiles. *Top*: Model DISH2. *Bottom*: Model DISH1. We show the profiles at $t = 0$ (*thin solid line*), 160, (*dashed line*), and 240 yr (*thick solid line*). Fragmentation occurs between 160 and 240 yr in model DISH1, while model DISH2 develops only strong spiral arms.

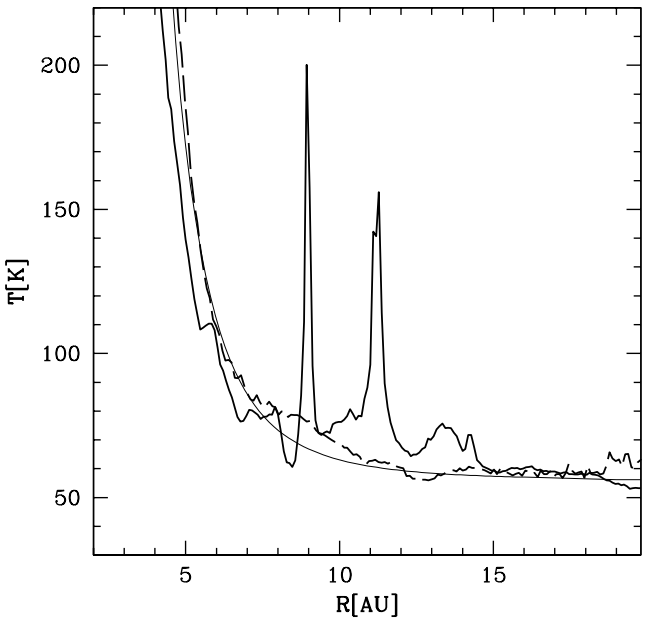

FIG. 2.—Evolution of the temperature profile. The initial profile (*thin solid line*) for an outer temperature of 56 K is shown; this profile was used for both model DISH1 and DISH2 (see Table 1). Model DISH1 undergoes clump formation, and its profile is shown at $T = 320$ yr (*thick solid line*), after the equation of state has been switched to adiabatic (see text). The peaks correspond to regions where bound clumps are (a single peak contains more than one clump because of limited bin size). Model DISH2 only forms spiral arms; its profile (*thick dashed line*) is also shown at $T = 320$ yr (the equation of state in this case is locally isothermal throughout the evolution, so the small changes of the temperature with radius are due to radial diffusion of particles only).

radial zone of the disk, we use a single smooth function over the entire range of radii for better numerical stability (sharp variations in the temperature profile are difficult to handle with a numerical technique such as SPH, which is based on averaging fluid properties over a given volume). We adopt the functional form $T(r) = T_0 \exp(-r/r_{\rm in})^{3/2} + C$, where $T_0 = 750$ K and C is a constant that is set equal to the minimum temperature chosen for each individual model. Between 5 and 10 AU the temperature scales as $\sim r^{-1/2}$, which resembles the slope obtained if viscous accretion onto the central star is the key driver of disk evolution (Boss 1993). Between 4 and 5 AU the temperature profile rises more steeply, being partially determined by irradiation from the central stellar source in agreement with the three-dimensional radiative transfer calculations of Boss (1996), while it smoothly flattens out for $R > 10$ AU and reaches a constant minimum temperature. The minimum temperature ranges between 35 and 100 K (typically it is around 50 K). It is implicitly assumed that the disk temperature is related to the temperature of the embedding molecular cloud core from which the disk would be accreting material (Boss 1996). Note that, at least for the protosolar nebula, 50 K is probably a conservative upper limit for the characteristic temperature at $R > 10$ AU based on the chemical composition of comets in the solar system (temperatures as low as 20 K are suggested in the recent study by Kawakita et al. 2001). Outer temperatures between 30 and 70 K are found also for several T Tauri disks by modeling their spectral energy distribution assuming a mixture of gas and dust and including radiative transfer (D'Alessio et al. 2001).

In addition to the Q-parameter, another important measure of the susceptibility of a disk to gravitational instabilities is provided by the X_m-parameter, $X_m = \Omega^2 R/2m\pi G\Sigma$, m being of the order of the unstable mode and R being the disk radius. Extensive numerical experiments conducted for both collisionless and gaseous disks have shown that X_m, coupled with Q, provides a good measure of the susceptibility of the disk to the swing amplification of a given mode. In swing amplification a leading wave is amplified into a higher amplitude trailing wave, and if the latter can be turned back into a leading wave, a feedback loop is initiated that can produce a disturbance whose amplitude is orders of magnitude greater than that of the initial wave (Binney & Tremaine 1987). In Figures 1 and 3 we show the Q-profile and X_m-profiles ($m = 2$ and 3) for two of our disk models; other models (see Table 1) differ only in the value of T or Σ, and hence their profiles can

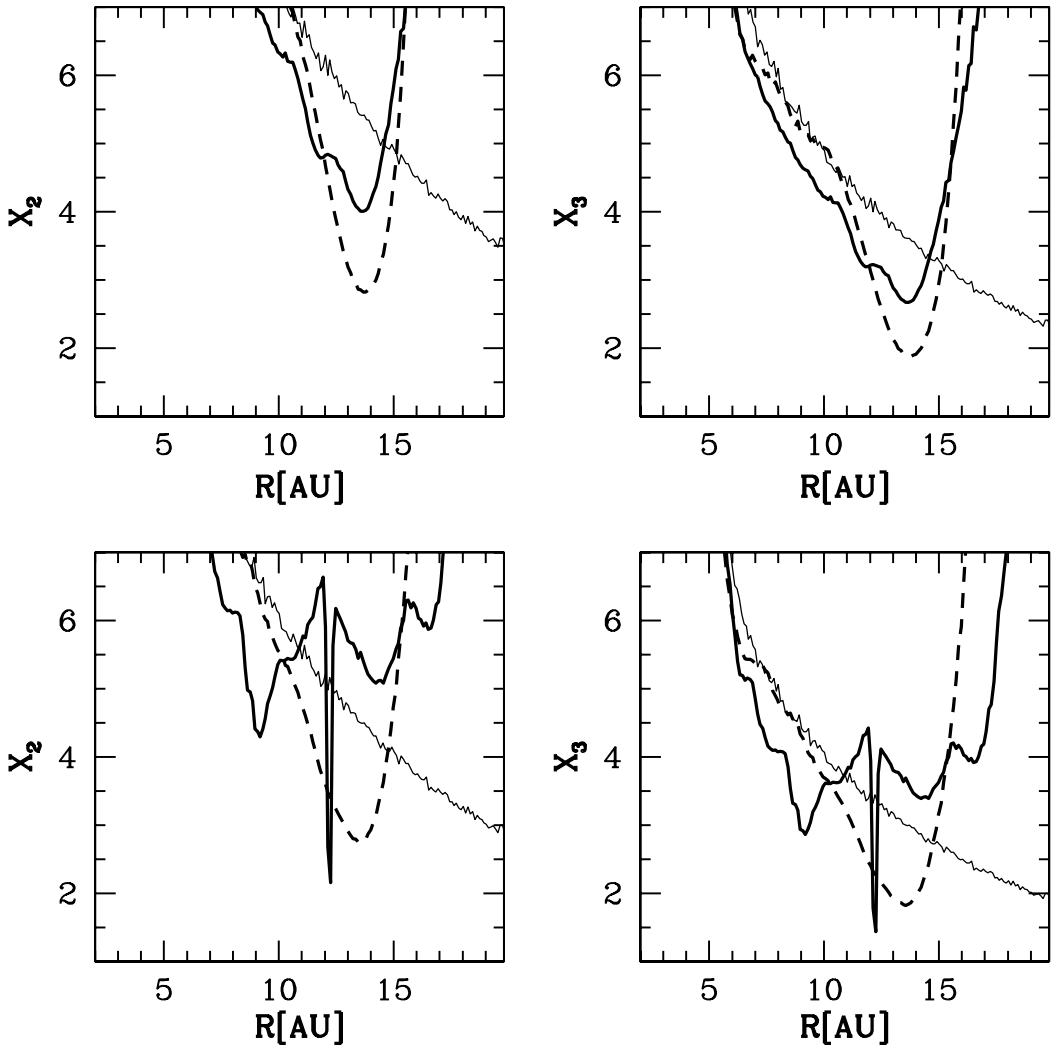

FIG. 3.—Evolution of the X_2 (*left panels*) and X_3 (*right panels*) profiles. *Top panels*: Model DISH2. *Bottom panels*: Model DISH1. Profiles are shown at $t = 0$ (*thin solid line*), 160 (*dashed line*), and 240 yr (*thick solid line*). See also Fig. 1 on the Q-profiles for the same models.

be easily recalculated (they differ only in the normalization, not in the shape). Strong swing amplification typically requires $X_m < 3$ and $Q < 3$ in some region of the disk (Binney & Tremaine 1987). Both conditions are marginally satisfied only at the very edge of the disk for $m = 2$, 3, and 4 in most of our models (e.g., DISL1 and DISH1; see Fig. 3), while they are definitely satisfied in model DISL4 at $R > 15$ AU (see Table 1; this model has a central star whose mass M_s is half of the standard value, and as $\Omega^2 \sim M_s$, $X_m \sim M_s$; hence, X_m is a factor of 2 lower for any m compared to other disk models with the same mass). In § 4.1.1 we discuss the role that the initially sharp outer edge of the disks might have in swing amplification. Note that from the definition of X_m, it follows that $X_m \to \infty$ for $m \to 0$. However, pressure also comes into play in determining which modes will actually grow. Modes whose scale lengths are smaller than the typical scale length associated with pressure, the disk scale height, will be inhibited, and this will occur preferentially for high-order, short-wavelength modes (see also § 4.1).

3. MODELING OF THERMODYNAMICS

We performed both locally isothermal and adiabatic runs. The gas is always assumed to be purely in the form of molecular hydrogen; hence, we assume $\mu = 2$ for the molecular weight. In what we call the adiabatic runs, we solve a thermal energy equation that also includes heating from artificial viscosity (see below). In particular, the quadratic term in the latter accounts for irreversible shock heating (Monaghan & Gingold 1983). The equation of state has the form $P = (\gamma - 1)\rho u$, where P is the pressure, ρ is the density, u is the specific internal energy of the gas, and γ is the ratio between the specific heats. We assume $\gamma = 1.4$ in most of the simulations (appropriate for molecular hydrogen with rotovibrational transitions), but we also consider a few cases with γ in the range 1–1.4. The locally isothermal equation of state stands on the assumption that any heating is instantaneously radiated away. This is simply written as $P = \rho u$, where u is now constant (therefore, no thermal energy equation is solved in this case) and is given by $u = k_B T/\mu$, where T is the temperature and k_B is the Boltzmann constant. In our simulations, the gas is isothermal in a Lagrangian sense, i.e., the thermal energy of a given particle is assigned based on its initial distance from the star and does not vary, irrespective of its subsequent motion through the disk. We note that, although some radial mixing is expected to occur because of the development of nonaxisymmetric instabilities, the initial temperature of the disk is constant by construction throughout most of the outer disk, where the strongest instabilities should

develop. We verified that radial motions of particles do not have any significant impact on the outer disk temperature profile; therefore, the hypothesis of local isothermality is self-consistent in the regions of interest (Fig. 2). Values of γ smaller than the canonical 1.4 yield smaller pressure and are expected to produce results closer to those of the locally isothermal runs. When the local density grows beyond 10 times the initial local value, the gas should behave nearly adiabatically according to the radiative transfer calculations of Boss (2002a) because of the increase in opacity (we recall that our disks have initial surface densities roughly identical to those in Boss's models), and therefore the locally isothermal calculations should not be self-consistent anymore. Therefore, in most of the locally isothermal runs we switch the equation of state to adiabatic once the critical density threshold is achieved. The change occurs throughout the entire disk in the calculations described in this paper.

All our simulations include artificial viscosity in the standard Monaghan formulation (Monaghan & Gingold 1983) plus the Balsara correction term (Balsara 1995) to reduce unwanted viscosity in purely shearing flows (see Wadsley et al. 2004 for details). Minimizing shear viscosity is crucial in simulations of self-gravitating disks, since if the shear flow is artificially weakened by viscosity, fragmentation could be driven in otherwise stable disks. Attempts to compare grid and SPH codes in fragmentation simulations do suggest that this might be a problem for SPH (Durisen et al. 2000). On the other hand, the effectiveness of the Balsara correction in reducing shear viscosity has been thoroughly investigated (Thacker et al. 2000). Artificial viscosity appears in both the momentum and the thermal energy equation. In SPH codes artificial viscosity is introduced for a number of reasons, primarily to avoid particle interpenetration and reduce postshock oscillations in high Mach number flows. The magnitude of the viscosity terms becomes smaller with decreasing smoothing length, and hence with increasing resolution. In most of the runs the linear and quadratic coefficients of artificial viscosity are set to, respectively, $\alpha = 1$ and $\beta = 2$. Although this choice is standard in three-dimensional SPH calculations (see, e.g., Hernquist & Katz 1989; Navarro & Benz 1991; Thacker et al. 2000), it is still borne out of classic tests such as the shock tube and the isothermally collapsing gas cloud, and it is not guaranteed to be optimal in more complex systems such as those considered here. In particular, whereas the quadratic term is needed to properly follow the shocks that will eventually develop during the gravitational instability, the linear term mainly damps the velocities of particles, reducing the noise inherent to the SPH technique. One worry is that numerical viscosity, by acting as an effective pressure, might smear out even physical small-scale features in the velocity field (e.g., in a region that is about to collapse because of gravitational instability) and generate spurious angular momentum transfers (Thacker et al. 2000). Heat generated by artificial viscosity when the thermal energy equation is solved might also affect the disk evolution. We investigate how artificial viscosity affects our results in both isothermal and adiabatic calculations by varying the value of the α- and β-coefficients (see Table 1, § 4.4).

4. RESULTS

In what follows we describe the results of our large suite of simulations. The various setups are indicated in Table 1. In the same table we also indicate whether clumps (protoplanets) are formed or not in a given run. Disks are typically followed for about 12–15 orbital times, where we define the characteristic orbital time as that at $R = 10$ AU (halfway between the center and the edge of the disk), $T_{\rm orb} = 2\pi(GM_s/R^3)^{1/2} = 28$ yr. Disks that undergo fragmentation are generally evolved for the shortest timescale (12 orbital times) because the growing local overdensities require increasingly smaller time steps to be accurately followed, with the result of slowing down considerably the computation. We use up to 400,000 time steps for the most expensive calculations. Two simulations (DISL1 and DISgr) were carried out for much longer timescales, about 20 and 30 orbital times, respectively, and their results are described in § 4.5, together with the structural and orbital properties of the formed protoplanets.

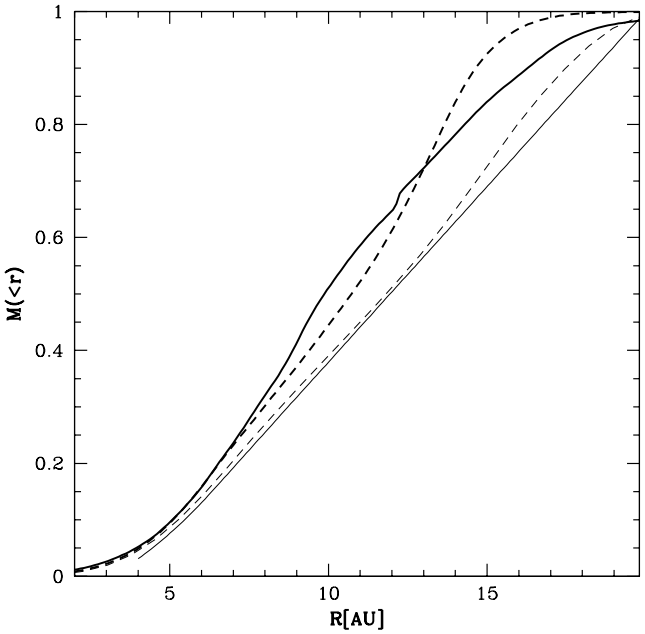

FIG. 4.—Evolution of the cumulative surface mass profile of model DISH1, with profiles at $T = 0$ (*thin solid line*), 200 yr (*dashed line*: at this time the spiral arms have the strongest amplitude and are about to fragment), and 350 yr (*thick solid line*: fragmentation has already occurred). The thin dashed line shows the profile at $T = 350$ yr for a disk model 10 times lighter than DISH1 (and with the same temperature), showing that the initial conditions evolve very little in the absence of significant self-gravity.

All disks develop trailing spiral instabilities after a few rotations, the strength and nature of which depend on the Q- and X_m-parameters and on the equation of state. The disk expands because of the spiral arms, which shed angular momentum outward and mass inward. As a result, its profile becomes more concentrated with time (Fig. 4). The expansion and readjustment of the disk profile is clearly an effect of self-gravity; as shown in Figure 4, a disk 10 times lighter than the standard models studied in this paper undergoes minimal changes in its mass profile over several orbital times. Although initially the minimum Q of the disks, $Q_{\rm min}$, is located at their outer boundary, their rapid expansion in response to the spiral instabilities causes a drop of the surface density in the outer part, and therefore $Q_{\rm min}$ shifts farther inward, near 13–14 AU (see Fig. 1). The larger pressure gradients developing in adiabatic runs tend to erase disk substructure created by gravitational instability, while locally isothermal runs provide the most favorable conditions for the development of the instabilities through damping of those same pressure gradients. The nature of the spiral pattern, namely, which modes are dominant, depends on the details of disk structure.

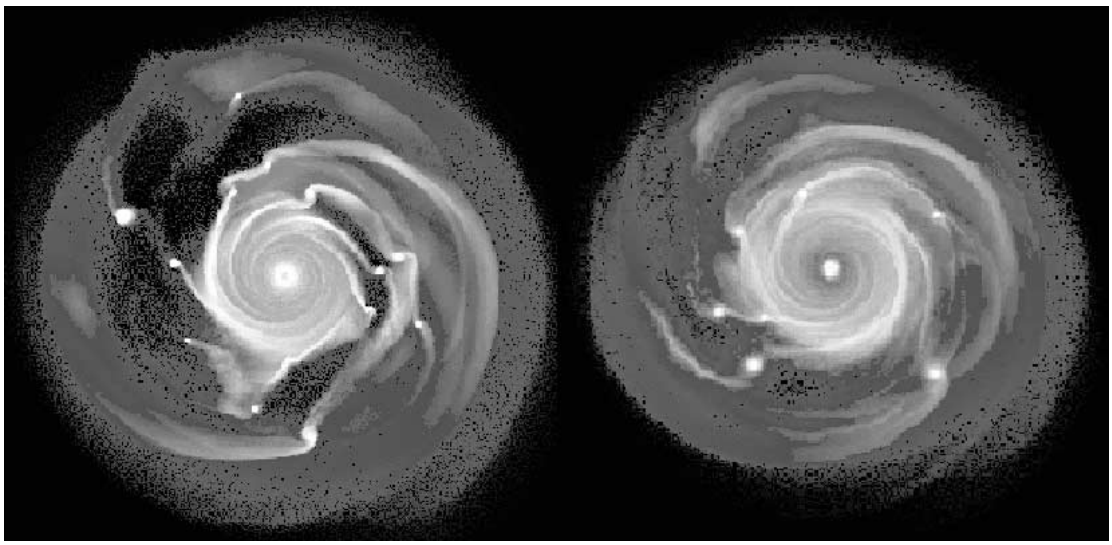

FIG. 5.—Gray-scale density maps of run DISH1 (see Table 1) after 350 yr, when the locally isothermal equation of state is used throughout the entire calculation (*left*) and when the simulation switches to an adiabatic equation of state (*right*) once the overdensities have grown past some threshold (see text, § 4.1) Brighter shades are for higher densities (densities between 10^{-14} and 10^{-6} g cm^{-3} are shown using a logarithmic scale; the same applies to all density maps shown in this paper), and the disks are shown out to 20 AU. [*See the electronic edition of the Journal for a color version of this figure.*]

In general we observe that, for a given Q, disks with lower mass and temperature tend to produce higher order spiral patterns. This is likely due to a combination of swing amplification being stronger for higher order modes in lighter disks [$X_m \sim (\Sigma m)^{-1}$] and a lower pressure allowing higher order modes to grow. Such trends in the spiral pattern as a function of disk parameters have been previously noted in other works on self-gravitating disks (e.g., Nelson et al. 1998, 2000a; Rice et al. 2003).

4.1. *Locally Isothermal Runs*

Disks evolved isothermally undergo fragmentation for $Q_{\rm min} \leq 1.4$. In these disks Q drops below unity between 12 and 15 AU after ~200 yr, and shortly after, several clumps appear that become gravitationally bound over a fraction of the orbital period, reaching densities of the order of 10^{-7} g cm^{-3} in their centers, up to 6 orders of magnitude higher than the initial local density (their further collapse being limited only by numerical resolution; see § 4.3). As a criterion to identify bound clumps, we check that $2E_{\rm th} + U < 0$, where $E_{\rm th}$ is the thermal energy and U is the gravitational binding energy of the clump. We adopt the algorithm described in Kazantzidis et al. (2004), which is similar to that used in the publicly available group finder SKID[4] but takes advantage of the fast gravity calculation performed by the same tree algorithm used in GASOLINE to compute the gravitational potentials of particles. Clump formation still proceeds when we switch to adiabatic conditions once the critical density threshold is reached. Clumps are fewer in the latter case, but several gravitationally bound ones are still present (Fig. 5). Newer calculations with explicit treatment of heating and cooling and in which the gas evolves adiabatically only in the overdense regions show very similar results in terms of numbers and densities of the clumps (L. Mayer et al. 2004, in preparation). We note that our adiabatic conditions include irreversible heating from artificial viscosity; the temperature near the spiral overdensities indeed rises above the isothermal value, reaching 80 K, but this is not enough to suppress clump formation at this stage because of the high density contrast already achieved (see Mayer et al. 2002). For $Q_{\rm min} \sim 1.65$ strong spiral arms are observed, but these saturate at some point (Laughlin et al. 1997), reaching a near stationary pattern after almost 20 orbital times. At even larger values of $Q_{\rm min}$, ~1.9, very weak spiral arms form and then saturate. The evolution of the Q-parameter strongly suggests that the threshold between fragmentation and self-regulation must lie near $Q_{\rm min} \sim 1.4$. Indeed, models whose initial $Q_{\rm min}$ is 1.65 reach $Q_{\rm min} \sim 1.15$ locally (between 12 and 15 AU), which is only slightly higher than the $Q < 1$ reached locally in the disks that fragment into clumps. A simulation with $Q_{\rm min} \sim 1.5$ also does not lead to fragmentation (see Fig. 6); the spiral arms reach a consistent amplitude after about 200 yr, but then weaken and saturate. Therefore, $Q_{\rm min} \sim 1.4$ really seems to mark the threshold for fragmentation in our calculations. Recent local grid-based calculations of razor-thin self-gravitating disks also find that the disk fragments into clumps for $Q < 1.4$ (Johnson & Gammie 2003).

As mentioned in § 2, the Q-profile is not enough to characterize the evolution of the different disk models; in addition to having different types of spiral patterns, disks with the same $Q_{\rm min}$ but different temperatures/masses yield clumps with varying mass. In particular, for a given $Q_{\rm min}$, lighter and colder disks produce less massive clumps, but the mass of the clumps does not scale linearly with the mass of the disks. For example, model DISH3, which has a mass of 0.085 $M_\odot$, produces several protoplanets with masses below a Saturn mass, which eventually grow up to a Jupiter mass or slightly above that (see below); therefore, in this case protoplanets have masses up to 3–4 times smaller than those arising in disks only 15% more massive (e.g., model DISL1).

The trend can be understood in terms of the dependence of the local Jeans mass on disk temperature and mass. Only overdensities whose scale is above the Jeans length and below the Toomre critical wavelength (see Binney & Tremaine 1987) will be able to grow and survive. In particular, any overdensity whose size is larger than the Toomre wavelength will be

[4] See http://www-hpcc.astro.washington.edu/tools/skid.html.

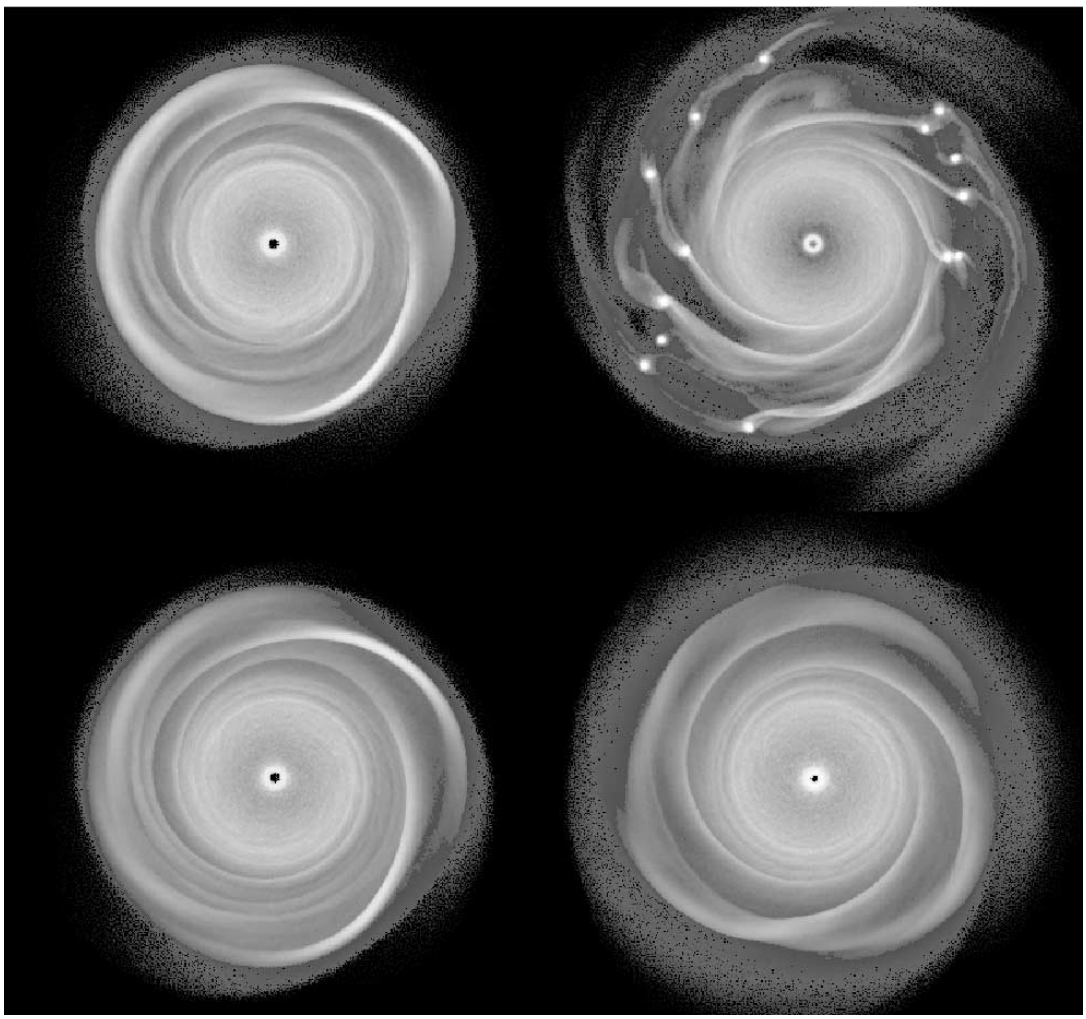

FIG. 6.—Gray-scale face-on density maps of run DISH3 (*top panels*; $Q_{\rm min} \sim 1.3$) and DISH3b (*bottom panels*; $Q_{\rm min} \sim 1.5$), at 200 (*left panels*) and 350 yr (*right panels*). See Table 1 for details on the models. The equation of state is switched to adiabatic close to fragmentation in run DISH3. Brighter shades are for higher densities (see Fig. 5), and the disks are shown out to 20 AU. [*See the electronic edition of the Journal for a color version of this figure.*]

sheared away by differential rotation irrespective of disk temperature, and it will not grow in the first place unless its mass and size are above the local Jeans mass or length, respectively. The Jeans length decreases with increasing density, while the Toomre wavelength increases with increasing density (see below). Initially, the Jeans length is bigger than the Toomre wavelength everywhere in the disk, but as the instability proceeds, its value drops below the latter, and fragmentation becomes possible. For disks with the same $Q_{\rm min}$, the Jeans mass is smaller for lighter (colder) disks. In fact, from the definition of the Jeans mass, we have $M_{\rm J} = (\pi/6)\rho(\pi v_s^2/G\rho)^{3/2}$ (ρ is the density, v_s the sound speed, and G the gravitational constant); hence, $M_{\rm J} \propto T^{3/2}$. However, the scaling between density and temperature at fixed Q is $\rho \sim T^{1/2}$, so ultimately $M_{\rm J} \sim T^{5/4}$. A difference of a factor of 2 in the temperature (required to compensate a difference of the order of $\sqrt{2}$ in the disk mass) introduces a difference of a factor of $\sim$2.4 in the Jeans mass and hence in the minimum mass expected for condensations. Figure 7 shows that, indeed, there is a factor of 2 or more variation in the minimum mass of the clumps at the onset of fragmentation when we compare disks with different masses and the same initial $Q_{\rm min}$.

Is the maximum size of the overdensities also consistent with simple theoretical expectations? According to the tight-winding (WKB) approximation, which is only valid for tightly wound (local) perturbations in a differentially rotating thin disk, only density perturbations whose scale is smaller than the Toomre wavelength, $\lambda_{\rm crit} = 4\pi^2 G\Sigma/\kappa^2$, can grow, where κ is the epicyclic frequency. The most unstable wavelength is $\lambda_{\rm mu} = 0.55\lambda_{\rm crit}$ for zero-thickness gaseous disks (Binney & Tremaine 1987). For most of our disks, for which $M_d = 0.1M_s$, we have $\lambda_{\rm mu} \sim 5$ AU at distances between 12 and 16 AU, the region where fragmentation occurs, and because $\lambda_{\rm mu}/R \sim 0.25$, i.e., $\lambda_{\rm mu}/R \ll 2\pi$ in the same range of radii, WKB results should still be valid for axisymmetric waves (Binney & Tremaine 1987). For nonaxisymmetric waves, the condition $X_m \gg 1$ (where X_m is the parameter defined in § 2) is a more appropriate one for the validity of WKB results, but this is also satisfied throughout most of the disk in the standard models, up to $m = 3$ (but see below on the swing amplification). We would expect clumps to occur at scales of the order of $\lambda_{\rm mu}$ and always below $\lambda_{\rm crit}$. In disks with nonnegligible vertical pressure, and hence finite thickness, both wavelengths will be somewhat smaller, as the disk will have an effective self-gravity lower than in the truly thin case. Numerical softening, in addition, can also be thought of as providing an artificial pressure on small scales. Romeo (1992, 1994) has computed correction factors for the stability properties of

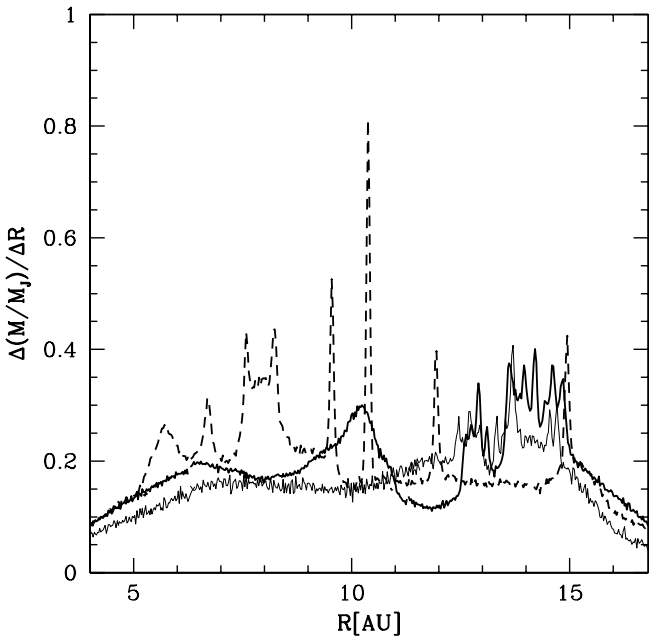

FIG. 7.—Radial mass profile at fragmentation. Results are shown for model DISH1 (*dashed line*), DISH3 (*thick solid line*), and DISgr (*thin solid line*). The mass is measured in units of $1M_{\rm Jup}$ using cylindrical bins equally spaced in radius. The peaks correspond to bound clumps. Although the differences in the masses of the disks are small (e.g., DISH1 vs. DISH3), the differences in the clumping mass scale are large because of the way the Jeans length scales with mass and temperature (see § 4.1).

self-gravitating stellar and gaseous disks that account for finite thickness and numerical softening. These factors are formulated as reduction factors for the surface density of the disks. In particular, if the wavenumber corresponding to the most unstable wavelength is $k_{\rm mu} = 2\pi/\lambda$, and the typical disk scale height resulting from pressure is h_d, then the correction factor for finite disk thickness reads $(1 + k_{\rm mu}h_d)^{-1}$; an identical expression can be obtained for the softening ϵ, where ϵ replaces h_d. In most of our simulations $\epsilon \ll h_d$, and hence we only consider the first reduction factor. Indeed, we find that $\lambda_{\rm mu}$ is reduced by more than 60% once we include the latter factor, becoming $\sim$2 AU, which turns out to be the typical size of the banana-shaped overdensities that appear along the spiral arms just before fragmentation. These overdensities rapidly fragment into multiple clumps, but during this phase the system has become so strongly nonlinear that any extrapolation of linear theory is meaningless. However, it is remarkable that while the system is still mildly nonlinear, the results of WKB theory are in good agreement with the simulations.

One ingredient that is not captured by the WKB approximation is swing amplification of nonaxisymmetric modes. Indeed, when strong swing amplification is expected, it also means that linear theory breaks down (Binney & Tremaine 1987). As we mentioned in § 3, most of our disk models (e.g., DISL1, DISH1, DISH2) are barely in the regime in which swing amplification is expected to be important at $R > 10$ AU. In particular, only for $m = 3$ (or even higher order) modes is the condition $X < 3$, necessary for strong swing amplification (see Binney & Tremaine 1987), satisfied at some point, and only quite late in the evolution (see Fig. 3). However, model DISL4, which has a lower mass for the central star, clearly departs from this picture, as $X < 3$ for both $m = 2$ and $m = 3$ modes early during the evolution (this suggests that WKB results are less applicable here). Indeed, the latter model exhibits a mixture of three- and two-armed modes visibly stronger than in the standard models (three-armed spirals are evident, but two of the three arms grow more in amplitude, and they begin to fragment, similarly to other simulations), as shown by comparing Figure 8 with Figure 6. Clumps appear almost 2 orbital times earlier in this model (the orbital time at 10 AU is 40 yr), after only 160 yr, and the stronger nonaxisymmetry produces orbits with eccentricities as large as $e = 0.35$, bigger than in the other models (see Mayer et al.

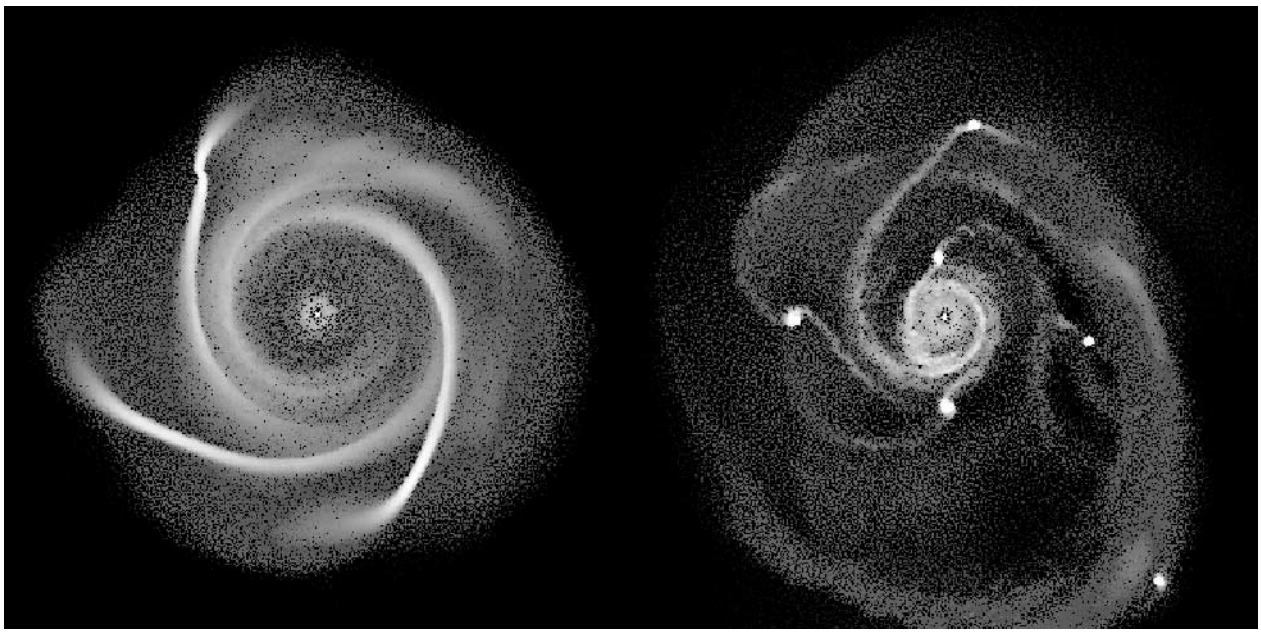

FIG. 8.—Gray-scale face-on density maps of run DISL4 after 120 (*left*) and 240 yr (*right*), out to a radius of 25 AU. Note the stronger spiral arms compared to the runs shown in Fig. 5 and the very eccentric orbital trail of the outermost clump in the right panel. [*See the electronic edition of the Journal for a color version of this figure*.]

2002 and § 4.5 of this paper), which then favor more frequent and violent interactions between the clumps.

4.1.1. *The Problem of the Initial Conditions*

How does fragmentation depend on the initial conditions of the disk simulations? An often heard argument against fragmentation models states that simulations with disks having a low Q-parameter by construction are unrealistic because the disk would respond to a developing gravitational instability by raising its temperature and adjusting its density profile, thus self-regulating to fairly high Q-values (Laughlin & Bodenheimer 1994; Laughlin & Różyczka 1996; Laughlin et al. 1997). The argument was originally developed against models starting with Q near unity somewhere in the disk, but one can easily imagine that mass redistribution due to milder nonaxisymmetric instabilities might always drive the disk to values higher than even the critical threshold that we claim here, $Q \sim 1.4$. Moreover, in the simulations presented so far, the disk has a sharp edge at $t = 0$. Although the disk very quickly (in 1 orbital time) expands as a result of its own evolution, the initial reflection of waves at the edge might artificially activate a feedback loop and sustain swing amplification, leading artificially to a faster and stronger growth of the instability. Motivated by these arguments, we performed a simulation (run DISgr in Table 1) in which the mass of the disk is increased gradually (over about 600 yr, 20 orbital times at $R = 10$ AU) by a factor of 10, going from $M_d = 0.0085$ (well below even the minimum-mass solar nebula) to $M_d = 0.085$ (the final mass is the same as in model DISH3). The mass is increased uniformly in the disk at every step (in practice, we grow the mass of each particle by a fixed constant). The disk is evolved isothermally in the first phase, and then adiabatically once the usual overdensity threshold is reached; the initial temperature profile is of the usual form (see § 2), and in particular, the outer minimum temperature is 30 K.

The disk stays considerably smooth for a long time (the initial $Q_{\min}$ is higher than 10), but as Q approaches 2, spiral instabilities start to appear and grow in amplitude, while the mass continues to increase (disk snapshots are shown in Fig. 9), until Q drops to near unity (Fig. 10) and clump formation occurs. Note that the disk surface density has not been redistributed significantly, since fragmentation occurs at roughly the same temperature and the same mass as those initially assigned to one of the fragmenting standard models initialized with a low $Q_{\min}$, model DISH3 (see Table 1). This experiment shows that mass redistribution when $Q_{\min}$ is higher than 1.4 or so is not important, since nonaxisymmetric torques are still too weak in this regime. Therefore, such a disk could still undergo clump formation provided that it cools down a bit. In addition, in this test the sharp outer edge disappears several orbital times before the first weak nonaxisymmetric pattern becomes apparent in the disk, yet fragmentation takes place more or less as in the other runs. We only note that in this simulation the dominant spiral patterns are of higher order ($m = 5$ or higher) than in any of the other runs; this is in part due to the fact that the disk enters the regime of nonaxisymmetric instability with a mass lower than in any of the other models, so that swing amplification is effective for high-order modes only during the first part of the evolution (see Fig. 10), but we cannot exclude the factor that which of the spiral patterns are dominant also depends on whether or not initial edge effects are present. The outcome of this run is in agreement with another result that points to negligible mass redistribution when $Q > 1.5$: the fact that disks with $Q_{\min} \sim 1.65$ maintain a value of $Q_{\min}$ similar to the initial one after the strongest phase of the instability is over (see Fig. 1). Of course, in reality the process of mass accretion from the molecular cloud is much more complicated than depicted here; the disk mass will not increase monotonically but will probably reach a maximum and eventually decrease as accretion onto the central star or photoevaporation take over accretion from the molecular cloud (Matsuyama et al. 2003). In addition, and most importantly, the temperature will not remain constant but will change as a result of heating and cooling.

4.2. *Adiabatic Runs*

Disks evolved adiabatically since $t = 0$ reach fragmentation only when starting from very low values of $Q_{\min}$, as small as 0.8. Such low values of $Q_{\min}$ are obtained with a combination of low temperatures and high masses ($M_{\rm disk} = 0.125\ M_\odot$; see the models marked "ad" in Table 1). With these very low values of the Toomre parameter, the disk is locally unstable to axisymmetric perturbations. Indeed, we see a ring forming in the outer part of the disks, but this is soon dissipated as Q rises because of heating by compressions and shocks, and spiral arms form after 2–3 orbital times. The precise value of $Q_{\min}$ needed for fragmentation decreases with increasing values of γ in the equation of state. The results are listed in Table 1 and shown in Figures 11 and 12. We note that even disks with $Q_{\min} \sim 0.8$ undergo fragmentation only if $\gamma = 1.2$, which would happen only if some cooling were present. In addition, for $\gamma = 1.3$ fragmentation does occur for $Q_{\min} \sim 0.8$, but the clumps are quickly washed out by the strong developing pressure gradients, while for $\gamma = 1.4$ overdensities in the spiral arms are washed out before being able to collapse. The temperature rises to more than 100 K along the spiral arms because of compressional and shock heating. Therefore, even at these very low values of Q, we do find that fragmentation depends on the equation of state, while an early work by Boss (1998) was finding fragmentation even for $\gamma = 1.4$ with $Q_{\min} \sim 1$. Our tighter limits are likely due to the inclusion of irreversible shock heating, which was absent in that study. Especially when these violently unstable modes are present, the inclusion of shock heating is very important, as was already emphasized in previous works by Pickett et al. (1998, 2000a, 2000b, 2003). Shock heating is evident along the trailing and leading edges of the spiral arms (see the temperature maps in Fig. 12); indeed, these are the locations where the velocities of the gas in the spiral arms and the mean sound speed differ the most.

As we mentioned above, these very low Q states are probably unrealistic, and this is shown by the rapid evolution that the disk undergoes, with $Q_{\min}$ rising by a factor of 2 in less than 5 orbital times. The clumps that form in run DISad4 are particularly big (see § 4.1 on how disk mass and temperature influence clump formation), growing to more than $10 M_{\rm Jup}$ after a few mergings. Finally, we note that in adiabatic simulations that start from higher values of $Q_{\min}$ (DISLad1, DISLad5, and DISLad6 in Table 1), fragmentation does not take place. The temperature along the spiral arms, on the other hand, grows from 56 to only about 70–80 K because the nonaxisymmetric modes are much weaker than in the runs with lower $Q_{\min}$. This temperature is only ~50% higher than that required to maintain Q below 1.4 and follow the path toward strong instability (the change in temperature has a negligible dependence on the magnitude of the artificial viscosity; see Table 1 for the different cases considered). Therefore, the magnitude of shock heating is not dramatic, and

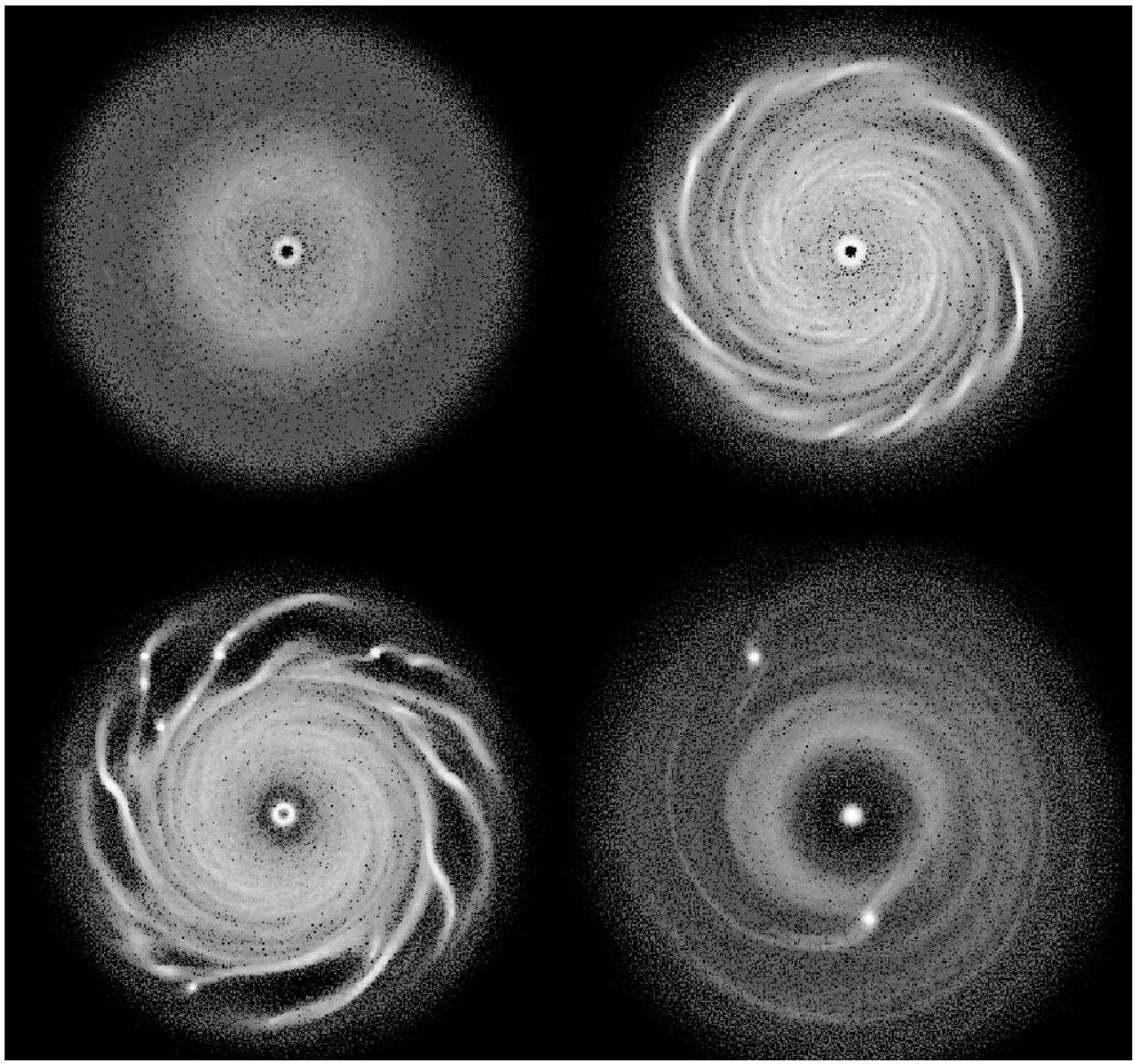

FIG. 9.—Gray-scale density maps of the growing disk simulation, run DISgr. Brighter shades are for higher densities (see Fig. 5), and the disk is shown out to 20 AU. From top to bottom and clockwise, snapshots are taken after 300, 480, 560, and 800 yr. [*See the electronic edition of the Journal for a color version of this figure*.]

in principle a modest amount of cooling should be enough to bring the disk toward the strongly unstable regime. However, the rise in temperature associated with shock heating occurs very fast, over less than an orbital time, while it remains to be seen whether cooling processes can counterbalance it by acting at comparable speed (but see Boss 2002a and Johnson & Gammie 2003).

4.3. *The Effect of Mass and Force Resolution on Clump Formation*

The physical interpretation of fragmentation proposed in § 4.1 clearly implies that the simulations ought to resolve the Jeans and Toomre wavelengths. The ability to resolve these characteristic physical scale lengths depends on both mass and force resolution in an SPH simulation; in particular, the resolution must be high enough to resolve the smallest among the two scales, usually the Jeans length. As extensively discussed by Bate & Burkert (1997), the gravitational softening and the SPH smoothing length should be comparable, and both should be smaller than the local Jeans length for the calculations to be trustworthy (see also Nelson 2003). When the gravitational softening exceeds the SPH smoothing length, any eventual collapse will be slowed down or halted, while artificial fragmentation might occur in the opposite situation. Bate & Burkert also showed that the local Jeans mass should always be resolved by no less than $2N_{\rm neigh}$, where $N_{\rm neigh}$ is the number of neighbors used in the SPH smoothing kernel; $N_{\rm neigh} = 32$ in our runs. Thanks to the extremely high resolution adopted in this work, the local Jeans mass is always resolved by several hundred to several thousand particles; hence, we are orders of magnitude above the minimum requirements. On the other hand, shortly after clump formation the smoothing length becomes significantly shorter than the gravitational softening inside the clumps, and hence the collapse of the clumps is ultimately slowed down once they have shrunk down to a scale comparable to the latter (at such a point the force softening acts like an artificial pressure force). This means that in the isothermal calculations the clumps become effectively

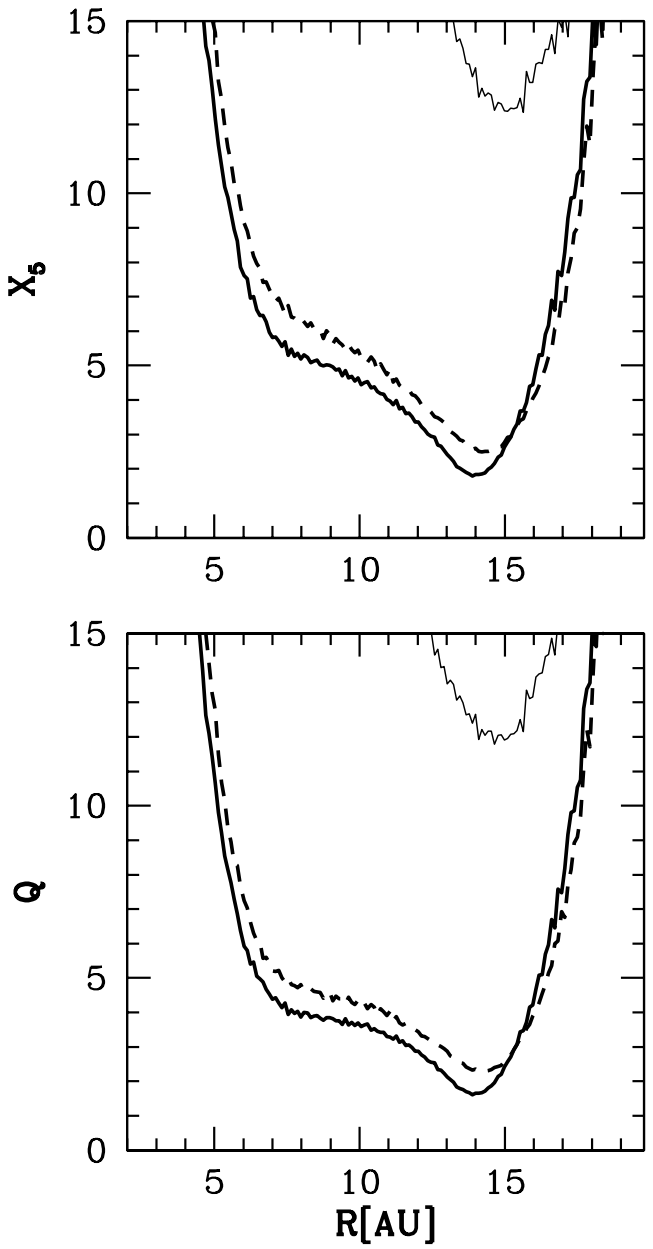

FIG. 10.—Evolution of the Toomre Q-parameter (*bottom*) and of the X_5-parameter (*top*) for the growing disk in run DISgr (see Table 1, § 4.1.1.). The profiles are shown at 300 (*thin solid line*), 450 (*dashed line*), and 500 yr (*thick solid line*: just before fragmentation starts). Note that the X_5-parameter is just below the threshold for instability (<3) even at late times, which might explain why high-order modes dominate in this run.

superadiabatic near the center, where the resolution limit is reached. Hence, the difference with the calculations in which the equation of state is switched to adiabatic exists only in the very first stage of clump formation, when the softening is still comparable to the other relevant scale lengths. The bottom line is that any detailed analysis of the internal structure of the clumps must be postponed to future simulations with even higher resolution. For the moment, only mean properties of the clumps, calculated as averages over the entire systems (which are comfortably larger than the softening), can be and are discussed (see § 4.5).

A very important result emerging from our set of simulations is that clump formation is enhanced with increasing mass resolution for the same initial conditions (with the softening scaling as $m_g^{1/3}$). A larger number of clumps is seen at higher resolution, although the mass scale of fragmentation is basically unaffected (Fig. 13). Indeed, the minimum clump mass depends on the local Jeans mass, which in turn is determined by the disk density and temperature, not by resolution (§ 4.1). Force resolution alone also has an effect. A softening larger than or comparable to the maximum unstable Toomre wavelength suppresses clump formation even in unstable disks. In fact, a large softening acts as an additional pressure (in practice, it suppresses gravity) at the crucial scale at which the perturbation has the highest amplitude, or, alternatively, one can view it as producing an effective increase in the local Toomre parameter. Following Romeo (1994), the maximum allowable softening for our disks should be around 0.37 AU, which, of course, is close to the effective $\lambda_{\rm mu}$ calculated in § 4.1. Actually, we find that a softening $\sim$3 times smaller than the latter is necessary to go beyond the stage of the mildly nonlinear regime and enter that of clump formation (see Table 1). Indeed, the calculation done by Romeo applies to the study of spiral structure in marginally unstable (galactic) disks and not to the strongly unstable regimes that we are investigating here. When Q locally drops below 1 and the strongly nonlinear regime is reached, a higher force resolution, close to the rapidly dropping Jeans length, should be required in order to keep following the dynamical evolution properly. Therefore, in a simulation both mass and force resolution must be high enough to follow the fragmentation process in the disk. The maximum and minimum allowable softenings can be accurately determined only through convergence tests, the only a priori prescription being to balance the softening and smoothing length (Bate & Burkert 1997). We note, however, that the line dividing stable and unstable disks appears to be independent of softening in our simulations; disks that are stable with our "standard" choice of parameters remain stable even with a softening 10 times smaller (for example, in Table 1 compare run DISH4 with DISH4b and DISH2 with DISH2c). This suggests that the threshold for stability, $Q_{\rm min} \sim 1.4$, is a robust physical result, at least under the thermodynamic conditions adopted.

4.4. *Dependence on Artificial Viscosity*

We find that the values of both α and β have some impact on disk evolution and fragmentation. Whereas disks that do not fragment are found to behave so irrespective of the artificial viscosity, disks that fragment do so more or less severely and on slightly different timescales depending on the value of these parameters. Here we discuss the results of locally isothermal runs with varied viscosity parameters (all listed in Table 1); hence, we neglect the artificial heating eventually induced by viscosity. The latter is present in adiabatic runs, but we already showed that in such runs disks never form bound clumps unless we reduce the value of γ significantly below the canonical 1.4 (see § 4.2), which makes them rather unsuitable for analyzing the effects of viscous heating on clump formation. Therefore, we decide to postpone the analysis of artificial viscous heating to a forthcoming paper in which we also implement radiative heating and cooling in the disks (L. Mayer et al. 2004, in preparation).

In general, a smaller value of β or α enhances fragmentation, and the opposite happens for larger values. Only $\beta \geq 3$ can completely suppress fragmentation; however, such high values of β are not a good choice for flows with moderate Mach numbers ($\sim$1–1.5), such as those occurring in our simulations (see, e.g., Hernquist & Katz 1989; Thacker et al.

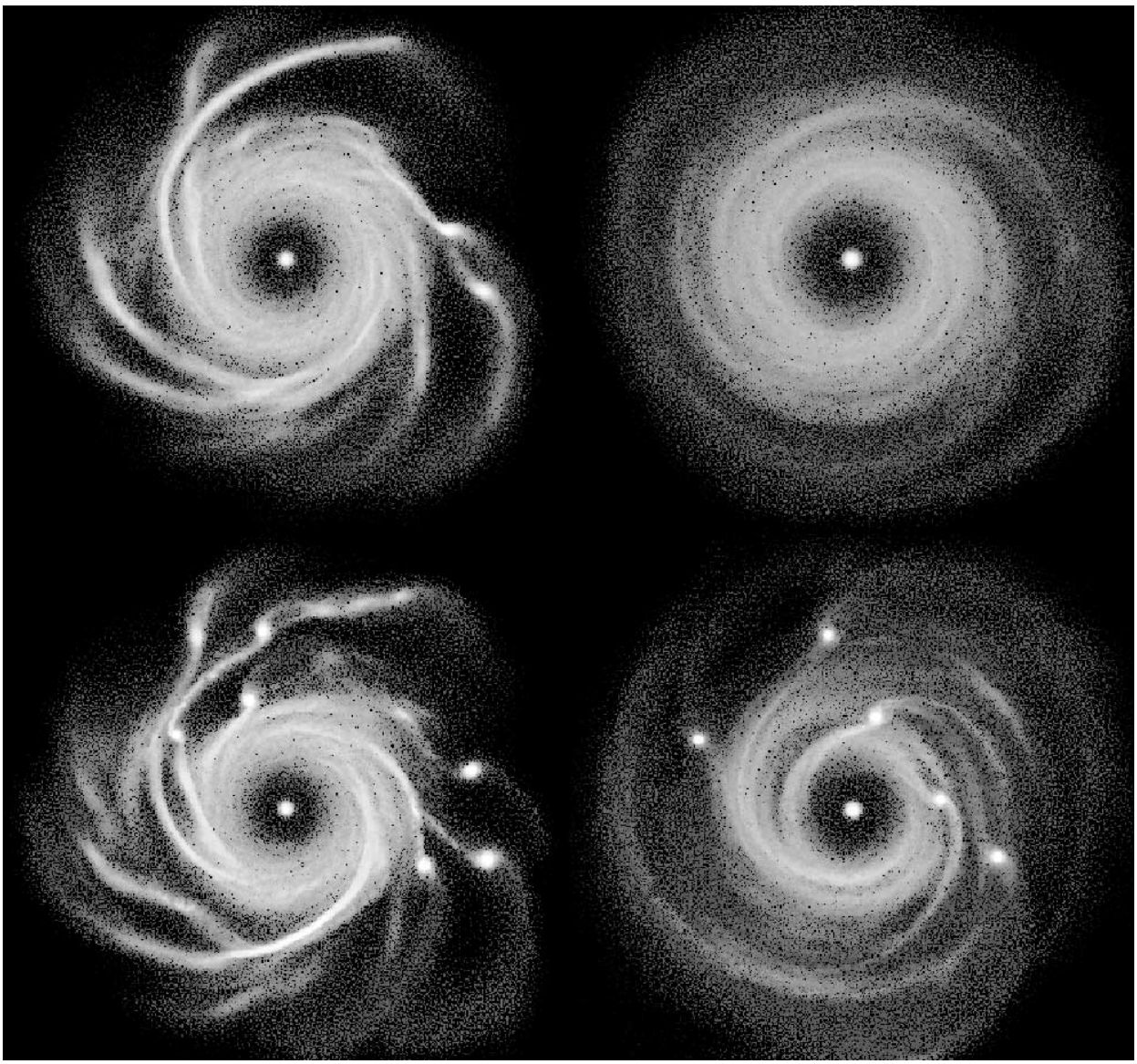

FIG. 11.—Gray-scale face-on density maps of adiabatic runs out to 20 AU. Brighter shades are used for higher densities (see Fig. 5). The two top panels show model DISLad3 at 200 (*top left*) and 300 yr (*top right*), while the two bottom panels show model DISLad4 at 200 (*bottom left*) and 300 yr (*bottom right*). [*See the electronic edition of the Journal for a color version of this figure.*]

2000), but could mimic the behavior of disks with a high degree of turbulence (Nelson et al. 1998). Why is viscosity affecting clumping? In general, artificial viscosity makes the velocity and density fields smoother, which helps to reduce postshock oscillations and noise but could in principle suppress small-scale physical features in the velocity field of the fluid. When the collapse of an overdensity begins, particles locally acquire radial motions, and hence the radial velocity dispersion will rise. A high viscosity can damp these radial motions and hence the collapse. The fact that the disk velocity dispersion profile is both lower and smoother with higher artificial viscosity is an indication of such an effect (Fig. 14). In particular, disk models that produce several bound clumps with the standard values of artificial viscosity (DISL1) clearly have their localized peaks in the radial dispersion profile completely smeared out for higher viscosity (DISL1e and DISL1f). The question arises of whether the dependence on artificial viscosity is strong enough to change the threshold Q for fragmentation. The answer is negative based on our simulations, since disks with Q above the threshold remain stable irrespective of the values of α and β. This is highlighted by the comparison between runs DISH2 and DISH2b, in which a disk with initial $Q_{\min} = 1.65$ is evolved first with the standard parameters and then with $\alpha = 0$ and $\beta = 0.5$. While the velocity dispersion profiles look somewhat different, the dynamical evolution is substantially identical, and in both runs the disk does not fragment.

4.5. *Long-Term Evolution of Disks and Protoplanets*

Once the disks enter the fragmentation phase, time steps become extremely small locally, and the computation becomes extremely demanding. One million particle runs require almost 400,000 steps for up to 350 yr of evolution, 80% of which cover only the last 2–3 orbital times ($\sim$100 yr), after clumps begin to form. We resort to lower resolution runs (200,000 particles) to probe the disk evolution over a more extended time (see also Mayer et al. 2002). In these runs the thermodynamics switch to adiabatic as soon as the local

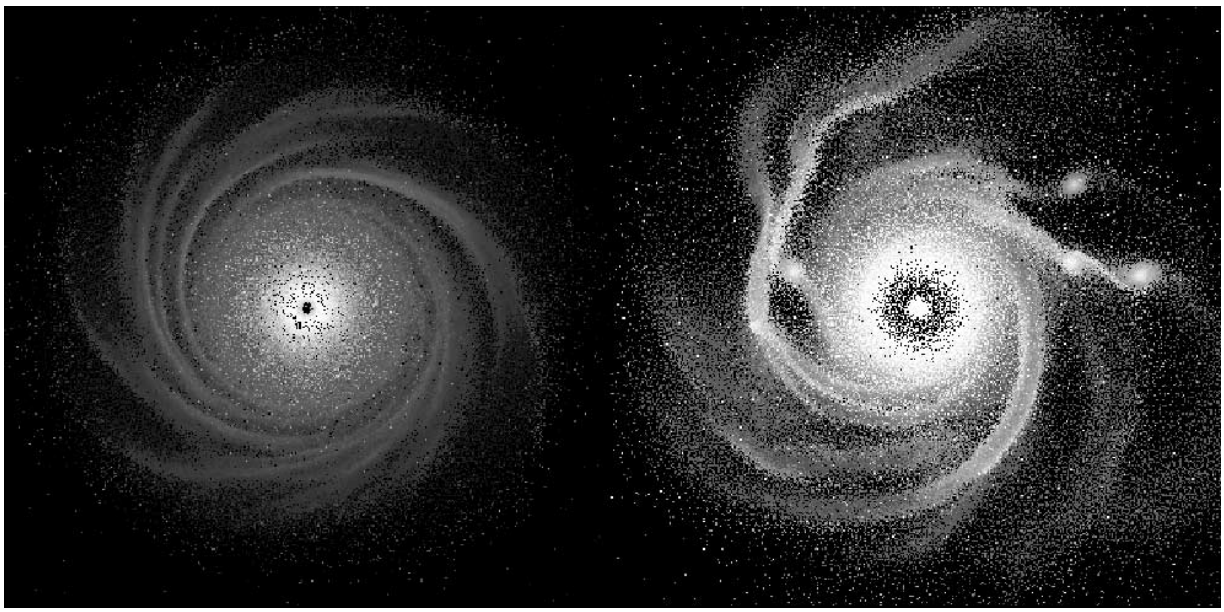

FIG. 12.—Gray-scale face-on temperature maps of the adiabatic run DISLad4. Brighter shades are used for higher temperature (the scale goes from 20 to 2000 K). Two snapshots are shown at 120 (*left*) and 220 yr (*right*; at this point clumps have just started forming), out to 20 AU from the central star. Note the strong increase of temperature along the edges of the spiral arms and at the location of the clumps. [*See the electronic edition of the Journal for a color version of this figure.*]

density becomes 10 times higher than the initial value (see § 4.1). The central temperature of the clumps grows up to 300–400 K rapidly after formation takes place owing to strong compression, while their collapse proceeds on their internal dynamical time (of the order of days); however, the clumps would certainly become much hotter in the center if their collapse were not halted at a scale comparable to the gravitational softening. Two low-resolution simulations ($N =$ 200,000 particles), one for a disk with mass $M = 0.1\ M_{\odot}$ (same as in run DISL1) and one for the growing disk (run DISgr, final mass $M = 0.085\ M_{\odot}$), were run for, respectively, 30 and 20 more orbital times (the reference orbital time being measured at 10 AU) after the onset of the fragmentation. Because of their extremely high densities, protoplanets are never destroyed by the tidal field of the central star (their tidal radius is more than 10 times larger than their typical size) but are tidally perturbed by and undergo a series of mergers with neighboring protoplanets (this phase lasts about 10 orbital times) until only three and two protoplanets remain in, respectively, model DISL1 (see also Mayer et al. 2002) and model DISgr (see the last snapshot of Fig. 9 for run DISgr).

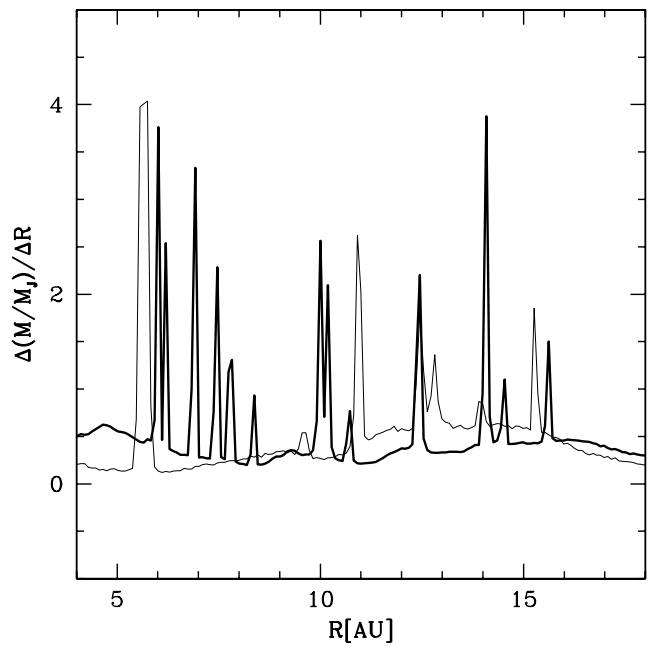

FIG. 13.—Differential mass profile after 350 yr in two simulations with the same disk model and different resolutions, run DISH1 (*thick solid line*) and DISL1 (*thin solid line*). The mass is measured in cylindrical radial bins, and the unit is $1M_{\rm Jup}$. See Table 1 for details on the simulations. Peaks correspond to gravitationally bound clumps. Clearly, several more clumps are present in the higher resolution run.

4.5.1. *Masses of Protoplanets and Disk Dispersal*

The gravitationally bound masses of the planets remaining at the end of the simulations (see § 4.1) range between $2.4M_{\rm Jup}$ and $6.6M_{\rm Jup}$ in the higher disk mass case, and between $0.07M_{\rm Jup}$ and $1.7M_{\rm Jup}$ in the lower disk mass case. Hence, both super-Jupiters and planets with masses as small as that of Saturn seem to be a possible outcome of the instability mechanism. This natural flexibility in accounting for the entire range of masses of gas giants known so far is an important feature of the disk instability model.

The final masses of the planets are not just the result of merging but also of the accretion of ambient gas, and the accretion rate in turn depends on the ambient pressure and thus on the equation of state of the gas (D'Angelo et al. 2003). In the extended simulations the equation of state is normally adiabatic, which means high pressure support of the surrounding gas (no cooling), and thus should yield a lower limit on the accretion rate. The mean accretion rate measured over the 500 yr following the onset of fragmentation in the run employing model DISL1 is quite low, in the range of 10^{-7} to $10^{-6}\ M_{\odot}\ {\rm yr}^{-1}$ (different accretion rates are found for different planets; in particular, the lowest are found for planets closest to the star, as expected from the higher ambient pressure). The accretion rate is declining toward the end of the simulations

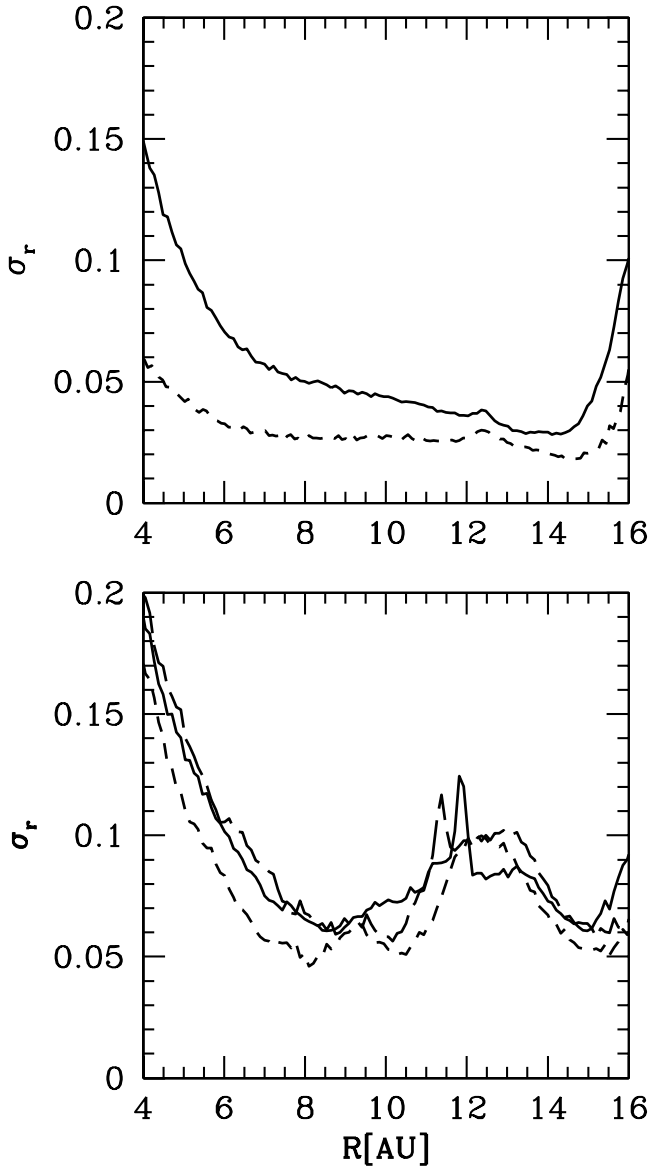

FIG. 14.—Effect of varying artificial viscosity on the velocity field of the disk in locally isothermal simulations. The radial velocity dispersion as a function of radius is shown. Regions where the radial dispersion is higher correspond to zones of higher nonaxisymmetric motion or even collapse (the peaks). *Top*: Two weakly unstable models, DISH2b (*solid line*) and DISH2 (*dashed line*), after 200 yr of evolution. *Bottom*: Three strongly unstable models, DISL1 (*solid line*), DISL1e (*long-dashed line*), and DISL1f (*short-dashed line*), at the time of maximum growth of the spiral overdensities (just before fragmentation in the case of run DISL1, after about 160 yr).

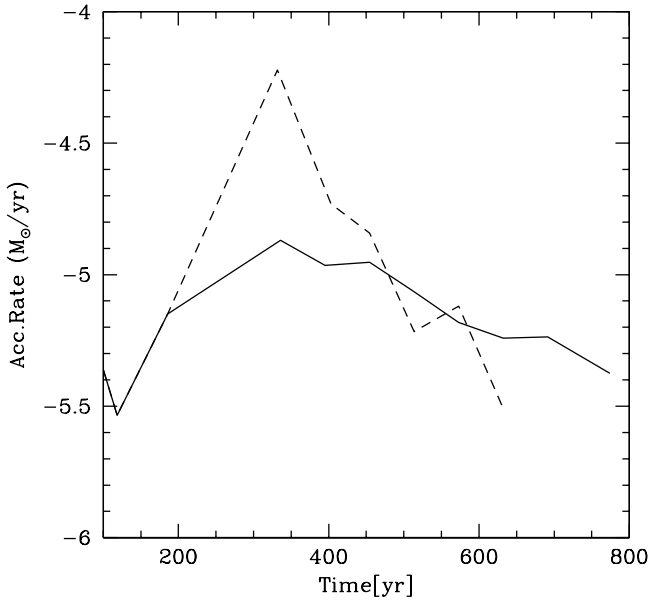

FIG. 15.—Accretion rate of gas onto the central star. The flow of mass inside a region of size equivalent to the gravitational softening length of the star (2 AU) is calculated. The simulations employ model DISL1 (see Table 1), the solid line is used for the case in which the equation of state is switched to adiabatic above the assigned density threshold, and the dashed line refers to the case in which the equation of state is locally isothermal for the entire duration of the calculation (see text, § 4.5).

and during the last couple of orbital times is practically zero. Therefore, at least in these adiabatic runs, the measured protoplanetary masses after ~1000 yr are certainly a good estimate of the final ones. However, the mass accretion rate onto the central star is still high, even at the end of the simulations, being more than $10^{-6}\ M_\odot\ \mathrm{yr}^{-1}$ (see Fig. 15); using the value of the accretion rate in the last stage of the simulation would yield a disk dissipation timescale as small as 20,000 yr (this is likely a lower limit given that the accretion rate is still declining at the end).

In passing we note that the short disk dispersal timescales predicted here as a result of gravitational instability would solve the puzzle of short disk lifetimes (shorter than a million years in at least 30% of the stars in Taurus; see Armitage et al. 2003). Other solutions, such as photoevaporative flows (Clarke et al. 2001; Armitage et al. 2003), require an input from external irradiation sources, since the heat generated by internal star-disk accretion shocks is probably insufficient (Matsuyama et al. 2003), but this would not work in, for example, Taurus because there are no massive stars capable of generating such a strong photoevaporating flux. One might worry that the accretion rate of gas toward the center and, in general, any motion of the particles in the disk might be partly caused by the artificial viscosity, which is well known to produce spurious losses of angular momentum (Thacker et al. 2000). We tested this latter possibility by stopping mass accretion in model DISgr when $M_d = 0.01\ M_\odot$ (the mass is an order of magnitude smaller than our typical disk masses) and running it forward in time for 20 orbital times (see Fig. 4). The disk remains very smooth in this case ($Q_{\min} > 8$), and the accretion rate toward the end is nearly 2 orders of magnitude lower than in the other runs, being stationary at around $3 \times 10^{-7}\ M_\odot\ \mathrm{yr}^{-1}$. We interpret the last number as the residual accretion rate due to artificial viscosity (note that this is a conservative choice because the disk is never perfectly axisymmetric because of the inevitable Poisson noise in the initial conditions). The corresponding accretion timescale (defined as the time required by a massive disk of about $0.1\ M_\odot$ to accrete onto the star by the latter mechanism only) is close to half a million years, much longer than any of the timescales considered here. Hence, the fact that a high accretion rate is a necessary consequence of gravitational instability and provides a way to clear out the disk very rapidly seems a well-grounded inference. Indeed, Pickett et al. (2003) have recently obtained similar accretion rates (although they probed the disk evolution on a shorter timescale) using high-resolution grid-based simulations that have a much lower numerical viscosity. Notwithstanding the high mass of the protoplanets, most of the mass at the end of the simulations, about 70%, is still in

the disks. Of this mass almost 30% is accumulated in the inner 2–3 AU as a result of accretion triggered by the non-axisymmetric torques and, partly, by artificial viscosity, while the rest is still at $R > 5$ AU, and hence is still important in determining the orbital evolution (and eventual migration) of the planets (see Lufkin et al. 2004). Accretion rates peak at values higher than $10^{-5}\ M_{\odot}\ \mathrm{yr}^{-1}$, significantly in excess of the rate measured in T Tauri stars (Gullbring et al. 1998); the strong bursts of infrared luminosity found in disks caught during the early stages of their evolution, such as FU Orionis, are suggestive of such high accretion rates and might indeed be explained with such strong inflows due to gravitational instability (Lodato & Bertin 2003).

We tested how accretion rates depend on the equation of state by resimulating for a few orbital times model DISL1 with a locally isothermal equation of state during and past fragmentation. The protoplanetary masses are 4 times higher than in the adiabatic simulation after 600 yr (at this point we stop the simulation), the time-averaged accretion rate being about 50 times higher. By this time, however, planets have also carved gaps that are not present in the adiabatic runs (compare Fig. 16 with the last snapshot of Fig. 9); although material can still flow to the planet, the accretion rate is strongly reduced after gap formation (Bryden et al. 1999). In particular, it is about 100 times smaller than the rate at which gas accretes onto the central object; therefore, the disk will be dissipated well before the planets can grow significantly further (Fig. 15). Once planets have carved a gap, further accretion of disk material should occur on the viscous timescale; the only physical source of viscosity in our disks is self-gravity (Lin & Pringle 1987; Laughlin et al. 1997) because, by design, we lack other possible sources such as, for example, magnetic fields. The effect of artificial viscosity (Bryden et al. 1999), on the other hand, is negligible at our resolution, as we discussed above. Therefore, at this point the accretion rate onto the protoplanets should in principle be proportional to that onto the central star. However, not everywhere in the disk is the dominant bulk motion directed toward the center. In fact, the directions of the torques arising from gravitational instability change with radius, and typically, while the material in the inner regions loses angular momentum, the material in the outer part gains it. The transition radius appears to be associated with the location of the dominant unstable spiral modes as determined by visual inspection; this typically corresponds to between 10 and 15 AU from the center in our simulations. Therefore, we expect that the protoplanets located at $R <$ 10 AU should accrete much more mass and much more rapidly than those located at $R > 10$ AU (the actual accretion rate is determined by the local flux of mass around the protoplanet, hence by the local details of the gravitational torques). The final masses of the protoplanets in this last run are between $10M_{\mathrm{Jup}}$ and $25M_{\mathrm{Jup}}$, therefore intermediate between those of brown dwarfs and those of extrasolar planets (Udry et al. 2002). However, we believe that these numbers are not to be taken seriously. Indeed, by marking particles belonging to a clump at the final time and tracing them back, we found that most of the material accreted by the protoplanets comes from the midplane and usually from a narrow annulus coplanar with the orbit of the planet in the adiabatic runs, while it occurs in a much more isotropic fashion in the isothermal runs, with a large fraction of accreted particles originally located at high distances from the plane. This big difference is certainly due to the fact that the vertical pressure gradients and those across the spiral shocks near the planet are artificially low in the isothermal simulation (shock heating is instantaneously damped). In fact, after 600 yr of evolution, the vertical structures of the isothermal and adiabatic run are dramatically different, the scale height being almost 10 times as big in the latter. This is reflected in the comparison between the vertical temperature profiles (Fig. 17).

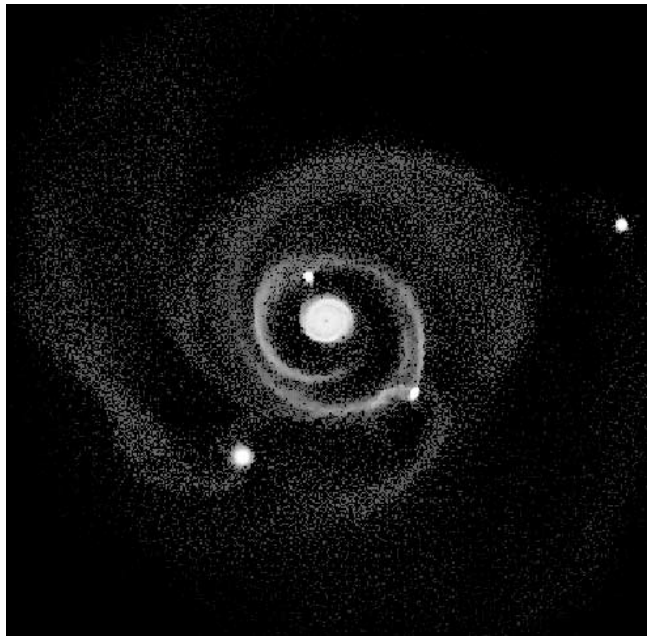

FIG. 16.—Gray-scale face-on density map of a run employing model DISL1 with a locally isothermal equation of state, even after the appearance of the overdensities (see text, § 4.5). The box is 30 AU on a side, and the snapshot is taken after 450 yr. Compared to the simulations employing adiabatic conditions once the overdensity threshold is reached (see Fig. 14 in this paper and Fig. 2 in Mayer et al. 2002), it is evident that protoplanets are carving much clearer gaps, considerably more mass is piling up at the center, and the disk is being dispersed much more quickly (the mean density is very low except in the central regions). [*See the electronic edition of the Journal for a color version of this figure*.]

In general, a conclusive answer on the final masses of the planets formed via gravitational instability has to await a more realistic treatment of the disk thermodynamics, with heating and cooling correctly modeled both inside and outside the overdense regions (L. Mayer et al. 2004, in preparation).

4.5.2. *Orbits and Migration of Protoplanets*

In all the simulations protoplanets have eccentric orbits and end up at mean distances between 3 and 12 AU. The eccentricity e, defined as $e = (R_{\mathrm{apo}} - R_{\mathrm{peri}})/(R_{\mathrm{apo}} + R_{\mathrm{peri}})$, with R_{apo} and R_{peri} being, respectively, the apocenter and pericenter distance, runs between 0.1 and 0.3. These eccentricities correspond to the mean values found for extrasolar planets (Marcy & Butler 1998; Marcy et al. 2000). Larger eccentricities (comparable to the highest measured for extrasolar planets, $\sim$0.7) are measured in a run in which the same disk as in run DISL1 is evolved with a locally isothermal equation of state throughout the evolution (see below). Very high eccentricities could also result from dynamical relaxation of these systems of massive protoplanets over timescales of several hundred thousand years (Terquem & Papaloizou 2002; Papaloizou & Terquem 2001).

There are hints that protoplanets undergo some orbital migration (Mayer et al. 2002), but this does not seem to have a preferred direction because of the chaotic nature of the torques

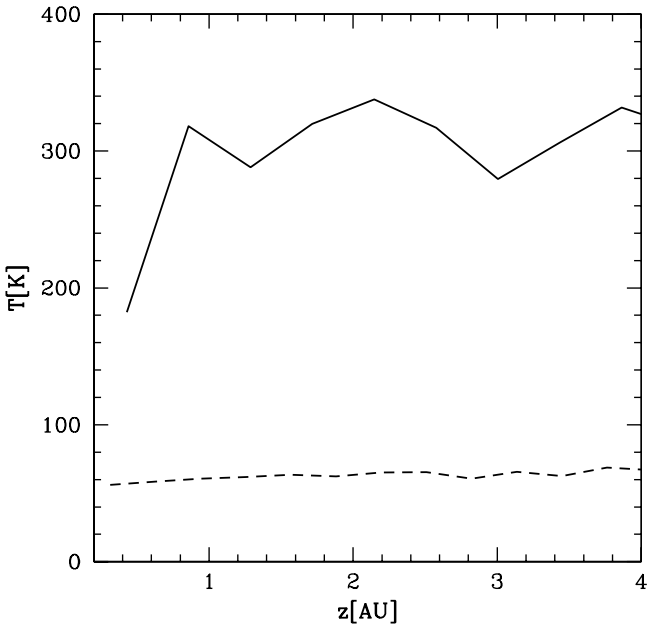

FIG. 17.—Temperature profile along the disk vertical axis for model DISL1 after 600 yr for a fixed locally isothermal equation of state (*dashed line*) and for the case in which the equation of state is switched to adiabatic prior to clump formation (*solid line*).

present in the system (as a result of both the nonaxisymmetric global potential and the gravitational interactions with the other planets). This supports the idea that current models of planet migration become unrealistic once nontrivial disk profiles (P. Artymowicz & W. Peplinski 2004, in preparation) and interactions between several bodies are simultaneously taken into account (Lufkin et al. 2004). Although the protoplanets are massive, their masses are still as small as a few times 10^{-2} the disk mass in the adiabatic run after nearly 1000 yr of evolution. As a consequence (Ward 1997a, 1997b; Tanaka et al. 2002), we do not expect migration to occur on the viscous timescale of the disk as in the classic type II regime, and indeed, no clear gap is visible. Instead, gravitational torques between the disk and the protoplanets should still determine the orbital evolution as in the type I regime, albeit with the additional complication of disk self-gravity and mutual interaction between the protoplanets. On the contrary, in the isothermal run the protoplanets acquire much higher masses because of the higher accretion rates, gaps are carved by them, and the orbital evolution resembles that of the type II regime (see above), with the difference that the disk is being dissipated very quickly. Although we are not able to probe the systems for timescales long enough to draw conclusions, an efficient inward orbital migration is difficult to imagine in the traditional framework, because the disks acquire a significantly steeper inner density gradient (inside the orbits of the planets) because of accretion of gas onto the central star; outward migration should thus be more likely (Masset & Papaloizou 2003).

So does this mean that hot Jupiters are difficult to explain within the present model? Not necessarily. There are indeed several possibilities of having efficient migration as soon as we move out of the standard framework. Indeed, one of the mechanisms recently proposed to explain the observed orbital distribution of extrasolar planets relies on dynamical relaxation of a population of massive planets formed rapidly through gravitational instabilities (Papaloizou & Terquem 2001). This model assumes that the disk is dissipated on a timescale much shorter than that required for relaxation (this being on the order of thousands of orbits, or several tens of thousands of years for the orbital timescales typical of our simulations), an assumption that seems to be marginally supported by our simulations. As mentioned above, when the disk is still present, the net gravitational torque caused by the instability will have different directions depending on the location in the disk. In the locally isothermal run with model DISL1, there is one protoplanet significantly inside 10 AU, and this might sink toward the center in only a few thousand years if it just follows the "bulk" accretion flow estimated for the disk (Fig. 15). The same would happen even in the adiabatic run, but of course the sinking timescale would be longer this time, of the order of a few tens of thousands of years. One is tempted to speculate that if protoplanets find themselves in a region where the torques due to gravitational instability become negative, they could drift toward the center rapidly with the rest of the disk. However, the situation is complicated by the fact that torques can fluctuate in sign and amplitude both spatially and temporally (R. H. Durisen 2003, private communication), and hence the resulting migration might be similar to a random walk (Rice & Armitage 2003).

Aside from the latter possibilities, certainly a more straightforward prediction of our model is that inward migration should not be very efficient; this would actually be a welcome feature of disk instability, since the problem of planets sinking too fast toward the central star is difficult to solve within the standard framework of migration in light disks without self-gravity (e.g., Bate et al. 2003). Ways that have been proposed to halt migration, such as the interaction between a planet on an eccentric orbit and the surrounding disk (Papaloizou & Larwood 2000; Papaloizou 2002), are naturally included in the gravitational instability model studied here, although the individual effects are difficult to disentangle given the complexity of the evolution. On the other hand, we have to keep in mind that our simulations still lack several ingredients, such as stellar winds, photoionization, and magnetic turbulence (Balbus & Hawley 1991; Nelson & Papaloizou 2003; Matsuyama et al. 2003), that can substantially affect the inner disk, eventually creating cavities or severely affecting the density profile, with consequent effects on the speed and direction of migration.

4.5.3. *Internal Structure of Protoplanets*

The surviving protoplanets are differentially rotating, nearly spherical bodies slightly flattened by rotation (the ratio between the major and minor axes is $\sim$0.9; see Fig. 18). We measure equatorial rotation speeds (this is measured at the outermost radius for which particles are gravitationally bound to the clump) and then calculate a tentative final rotation speed of the protoplanets by allowing contraction down to the mean density of Jupiter, assuming conservation of angular momentum. We find values in the range 3.5–40 km s^{-1}, which nicely encompasses the (equatorial) rotation speeds measured for Saturn and Jupiter, 8.7 and 14.6 km s^{-1}, respectively. We note that a long-standing problem has always been how Jupiter can still maintain a high rotation speed despite the fact that some kinetic energy must have been dissipated by atmospheric friction after its formation; the present model suggests a solution in that protoplanets could form with rotation speeds well in excess of the speed that Jupiter has today. Among the surviving protoplanets the obliquities go from a few to 160°. Large obliquities seem to be associated with

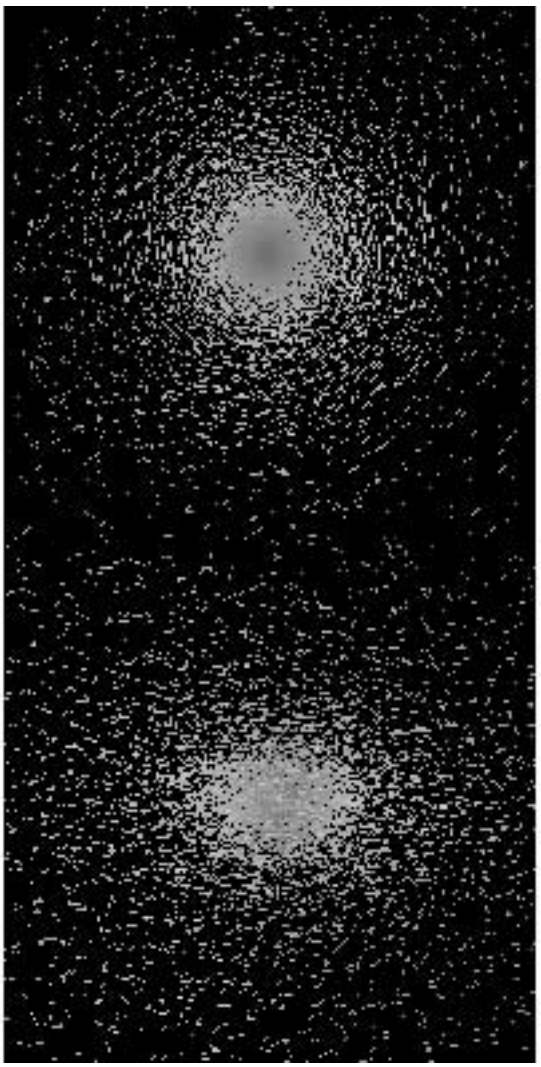

FIG. 18.—Close-up view of a gravitationally bound clump in run DISH1 (with the adiabatic switch) at 350 yr. We show the projection perpendicular to the angular momentum axis of the clump on top, and that along the angular momentum axis at the bottom. It is apparent that the clump is a flattened spheroid. A gray-scale plot of the velocity field is shown, in which the darker the color is, the higher the velocity. The clump is clearly in differential rotation. The box extends out to 0.5 AU and contains about 17,574 particles. [*See the electronic edition of the Journal for a color version of this figure*.]

mergers (and thus transfers of angular momentum); planets experiencing fewer mergers are in both runs those with smaller obliquities. Our model would then naturally explain the wide range of obliquities that we find in our solar system once we consider gas and ice giants together (a possible common origin of gas and ice giant planets via gravitational instability is discussed in Boss 2002b and Boss et al. 2002).

5. CONCLUSIONS

The main result shown in this paper is that fragmentation into long-lived, tidally stable, gravitationally bound protoplanets with masses and orbits comparable to those of observed extrasolar planets is possible in marginally unstable protoplanetary disks ($Q_{\rm min} < 1.4$). The requirement is that the disk cool efficiently (as implicit in the locally isothermal approximation) until the spiral arms approach fragmentation; once the local density grows by roughly an order of magnitude, gravity is strong enough for the collapse to proceed even with purely adiabatic conditions. We also showed that resolution, both in mass and in the gravitational force calculation, is a decisive factor in order to model disk evolution and fragmentation properly. In SPH simulations in particular, it is crucial that gravity and pressure be resolved at comparable levels for most of the extent of the simulations. In fact, no matter how many particles are used, if the gravitational softening is too large, the spiral arms do not reach the critical amplitude for fragmentation, as the dynamical response of the system is altered, and in particular, the growth of the amplitude of the modes is suppressed. Once fragmentation is approached, high resolution is also needed for the clumps to continue collapsing. Because the survival of clumps subject to strong tidal stresses depends on their binding energy, and thus on the density they are able to reach, the fact that in our simulations clumps survive for several tens of orbital times is also a consequence of the high resolution employed. Their densities are several orders of magnitudes higher than the mean density, resulting in a tidal radius 10 times larger than their typical size. Therefore, they will eventually survive for timescales much longer than those probed here and can thus be associated with protoplanetary objects. Their destruction can only happen through mergers with other protoplanets or from a strongly increased tidal field in the case in which their orbit migrates inward substantially. Pressure gradients near the clumps might drive dust and planetesimals and enrich the gaseous protoplanets up to metallicities beyond the solar value (Haghighipour & Boss 2003).

Previous works on gravitational instability fell short of the resolution needed to follow the very nonlinear stage of disk evolution. In addition, fixed boundaries were certainly a problem; in run DISL1, for example, both the outermost and the innermost clump go, respectively, farther out and farther in than the initial outer and inner radius of the disk, because of their eccentric orbits; therefore, even with enough resolution, typical fixed grids would not have been able to follow two out of three clumps (see also Pickett et al. 2000a; Boss 2000). Recently, Pickett et al. (2003) identified several banana-shaped overdensities in their grid simulations; these structures have densities and shapes strikingly similar to what we find in our disks just before clump formation starts. As they discuss and test, their simulations seem to lack enough azimuthal resolution to be conclusive about the evolution of the overdensities. As a comparison, their grid cell size is $\sim$5 times bigger than the gravitational softening in our 10^6 particle runs in the outer, more unstable regions; indeed, in runs in which the gravitational softening is increased by a factor of 3 or more with respect to the nominal value (DISL1b, DISL1c in Table 1), we also witness a suppression of fragmentation (§ 4.3).

Although the global nonaxisymmetric instabilities seen in the simulations and the resulting fragmentation cannot be captured by the WKB approximation, the maximum scales of the overdensities in the mildly nonlinear regime seem to be understandable in terms of the maximum Toomre wavelength. The minimum masses of the forming clumps are instead controlled by the local Jeans mass; because of the scaling with disk mass and temperature, disks with similar Q-profiles can produce smaller or bigger clumps depending on their mass and temperature. The smallest clumps have masses lower than that of Saturn at formation. Further mass growth due to merging and accretion shifts the typical masses to objects comparable to or bigger than Jupiter, but still it is clear that gravitational instability does not produce only super-Jupiters. Indeed, the

mean mass among the surviving protoplanets in the extended runs is $\sim 2.5 M_{\rm Jup}$.

We investigated whether starting a disk with a low $Q_{\rm min}$ at the beginning of a simulation, as we always do, might artificially enhance fragmentation. We showed that a disk grown from a very light state over tens of orbital times still produces several gravitationally bound protoplanets once it reaches values of both temperature and mass comparable to those of one of the models that undergo clump formation (§ 4.5). This result suggests that the weak nonaxisymmetric instabilities occurring when $Q_{\rm min} > 1.4$ are unable to affect appreciably the disk surface density by transporting mass inward. In fact, had the surface density been reduced significantly, a temperature significantly lower than that of the reference run would have been required in order for the disk to reach $Q_{\rm min} < 1.4$ and fragment. Such a negligible effect of spiral instabilities occurring at high-Q states is confirmed by the analysis of the runs that start with Q in the range 1.5–2; for example, Figure 1 shows that the isothermal run DISH2 has a final Q-profile (and also a final surface density profile) quite similar to the initial one. In reality, the growth of the disk mass with time is determined by the balance between the accretion rate onto the central star, as determined by both gravitational instability and other processes, for example, viscosity produced by magnetic fields and the accretion rate of material falling onto the disk from the molecular cloud envelope. Therefore, the mass will not grow at a constant rate as assumed here; instead, the process will be strongly time-dependent. However, hydrodynamic simulations of disk formation (Yorke & Bodenheimer 1999) that include radiative transfer but neglect magnetic fields do find that disks reach Q-parameters in the range 1.3–1.5 early in their evolution. Whether fragmentation will actually occur will then depend on how well a disk can radiate away the thermal energy produced by compression and shocks along the edges of the growing spiral arms. This is the most important, still open question concerning the final outcome of gravitational instabilities with the inclusion of realistic thermodynamics (Pickett et al. 1998, 2000a, 2000b, 2003; Mejia et al. 2003). In fact, even in our growing disk simulation we were keeping the local temperature constant before the appearance of the overdensities. Recent, lower resolution SPH simulations that solve for heating and cooling find that fragmentation can proceed when the cooling time is comparable to the disk orbital time (Rice et al. 2003), basically confirming previous simpler numerical and analytic calculations by Gammie (2001). Similar conclusions are reached in the recent three-dimensional calculations with volumetric cooling by Pickett et al. (2003). Such short cooling timescales are actually achieved in the simulations of Boss (2001, 2002a, 2002b), which include radiative transfer in the diffusion approximation with realistic disk opacities, because of efficient vertical energy transport by convection. We will address these issues in a forthcoming paper using very high resolution simulations that incorporate different forms of radiative cooling. These simulations will also allow us to model more realistically the accretion rate onto the protoplanets and thus produce a better prediction for their masses.

Finally, within the gravitational instability model the appearance of the protoplanets and a rapid disk dispersal seem to be linked. Although it is premature to estimate a robust disk dispersal timescale (even this can vary depending on the way the disk thermodynamics is treated), our calculations suggest that most of the disk material originally at tens of AU from the central star will be accreted in less than 10^5 yr. Material originally located outside the strongly unstable region will instead gain the angular momentum shed by the strong spiral arms and avoid rapid accretion. The disk will thus become more extended and comparatively less dense in the outer part, as suggested by the evolution of the mass distribution of our disk models. Therefore, a prediction of the gravitational instability model is that there must be a population of fairly young protoplanetary disks (considerably less than a million years old) in which a gap in the mass density of gas exists between an inner and an outer zone. Where and how sharp the transition is between these two regions is yet to be determined and will require simulations starting from disks more extended than our models and capable of probing their evolution for a much longer timescale. The *Spitzer Space Telescope* and other upcoming missions (Evans et al. 2003), by looking in the mid-infrared wavelengths, will allow us, for the first time, the tracing of the structure and evolution of the gaseous component in protoplanetary disks using the rotational emission lines of their main constituent, molecular hydrogen, and will provide a direct estimate of disk dispersal timescales.

L. M. thanks all the participants of the workshop "Circumstellar Disks and Protoplanets" organized by Tristan Guillot at the Nice Observatory in 2003 February and, in particular, Hal Levinson, Doug Lin, Pawel Artymowicz, Tristan Guillot, Paolo Tanga, Patrick Michel, Alessandro Morbidelli, and Ricardo Hueso for insightful and stimulating discussions. We also thank the referee, Richard Durisen, for several useful remarks and comments that improved considerably the paper and for stimulating discussions on the disk instability model. The numerical simulations were performed on LeMieux at the Pittsburgh Supercomputing Center, on the Z-Box at the University of Zurich, and on the Intel cluster at the Cineca Supercomputing Center in Bologna (Italy).

REFERENCES

Armitage, P. J., Clarke, C. J., & Palla, F. 2003, MNRAS, 342, 1139
Artymowicz, P., Lubow, S. H., & Kley, W. 1998, in Planetary Systems: The Long View, ed. L. M. Celnikier & J. T. V. Tran (Editions Frontières), 381
Balbus, S., & Hawley, J. F. 1991, ApJ, 376, 214
Balsara, D. S. 1995, J. Comput. Phys., 121, 357
Barnes, J., & Hut, P. 1986, Nature, 324, 446
Bate, M. R., & Burkert, A. 1997, MNRAS, 288, 1060
Bate, M. R., Lubow, S. H., Ogilvie, G. I., & Miller, K. A. 2003, MNRAS, 341, 213
Beckwith, S. V. W., Sargent, A. I., Chini, R., & Gusten, R. 1990, AJ, 99, 924
Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
Boss, A. P. 1993, ApJ, 417, 351
———. 1996, ApJ, 469, 906
———. 1997, Science, 276, 1836
———. 1998, ApJ, 503, 923
———. 2000, ApJ, 545, L61
———. 2001, ApJ, 563, 367
———. 2002a, ApJ, 576, 462
———. 2002b, Earth Planet. Sci. Lett., 202, 513
Boss, A. P., Wetherill, G. W., & Haghighipour, N. 2002, Icarus, 156, 291
Briceño, C., et al. 2001, Science, 291, 93
Bryden, G., Chen, X., Lin, D. C. N., Nelson, R. P., & Papaloizou, J. C. B. 1999, ApJ, 514, 344

Cameron, A. G. W. 1978, Moon Planets, 18, 5
Charbonneau, D. T., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45
Clarke, C. J., Gendrin, A., & Sotomayor, M. 2001, MNRAS, 328, 485
D'Alessio, P., Calvet, N., & Hartmann, L. 2001, ApJ, 553, 321
D'Angelo, G., Kley, W., & Henning, T. 2003, ApJ, 586, 540
Durisen, R. H., Pickett, B. K., Bate, M. R., Imamura, J. M., Brandl, A., & Sterzik, M. F. 2000, in IAU Symp. 200, Birth and Evolution of Binary Stars, ed. B. Reipurth & H. Zinnecker (Potsdam: Astrophys. Inst. Potsdam), 187
Dutrey, A., Guilloteau, S., Duvert, G., Prato, L., Simon, M., Schuster, K., & Menard, F. 1996, A&A, 309, 493
Evans, N. J., II, et al. 2003, PASP, 115, 965
Gammie, C. F. 2001, ApJ, 553, 174
Guillot, T. 1999a, Planet. Space Sci., 47, 1183
———. 1999b, Science, 286, 72
Guillot, T., & Showman, A. P. 2002, A&A, 385, 156
Gullbring, E., Hartmann, L., Briceño, C., & Calvet, N. 1998, ApJ, 492, 323
Haghighipour, N., & Boss, A. P. 2003, ApJ, 586, 1442
Haisch, K. R., Jr., Lada, E. A., & Lada, C. J. 2001, AJ, 121, 2065
Hernquist, L. 1993, ApJ, 409, 548
Hernquist, L., & Katz, N. 1989, ApJS, 70, 419
Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2002, Eos Trans. AGU, 83 (47), Fall Meeting Suppl., abstract P22C-0415
Johnson, N. M., & Gammie, C. F. 2003, ApJ, 597, 131
Kawakita, H., et al. 2001, Science, 294, 1089
Kazantzidis, S., Mayer, L., Mastropietro, C., Diemand, J., Stadel, J., & Moore, B. 2004, ApJ, 608, 663
Kuiper, G. 1951, Proc. Natl. Acad. Sci., 37, 1
Laughlin, G., & Bodenheimer, P. 1994, ApJ, 436, 335
Laughlin, G., Korchagin, V., & Adams, G. 1997, ApJ, 477, 410
Laughlin, G., & Różyczka, M. 1996, ApJ, 456, 279
Lin, D. N. C., & Pringle, J. E. 1987, MNRAS, 225, 607
Lissauer, J. J. 1993, ARA&A, 31, 129
Lodato, G., & Bertin, G. 2003, A&A, 408, 1015
Luf'kin, G., Quinn, T., Wadsley, J., Stadel J., & Governato, F. 2004, MNRAS, 347, 421
Marcy, G. W., & Butler, R. P. 1998, ARA&A, 36, 57
Marcy, G. W., Vogt, S., Fischer, D., & Butler, R. P. 2000, BAAS, 196, 35.14
Masset, F., & Papaloizou, J. C. B. 2003, ApJ, 588, 494
Matsuyama, I., Johnstone, D., & Murray, N. 2003, ApJ, 585, L143
Mayer, L., Quinn, T., Wadsley, J., & Stadel, J. J. 2002, Science, 298, 1756
Mayor, M., Udry, S., & Queloz, D. 1999, BAAS, 31, 05.01
Mejia, A. C., Durisen, R. H., & Pickett, B. K. 2003, in ASP Conf. Ser. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. D. Deming & S. Seager (San Francisco: ASP), 287
Monaghan, J. J., & Gingold, R. A. 1983, J. Comput. Phys., 52, 374
Navarro, J., & Benz, W. 1991, ApJ, 380, 320
Nelson, A. F. 2003, in ASP Conf. Ser. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. D. Deming & S. Seager (San Francisco: ASP), 291
Nelson, A. F., Benz, W., Adams, F. C., & Arnett, D. 1998, ApJ, 502, 342
Nelson, A. F., Benz, W., & Ruzmaikina, T. V. 2000a, ApJ, 529, 357
Nelson, R. P., & Papaloizou, J. C. B. 2003, MNRAS, 339, 993
Nelson, R. P., Papaloizou, J. C. B., Masset, F., & Kley, W. 2000b, MNRAS, 318, 18
Papaloizou, J. C. B. 2002, A&A, 388, 615
Papaloizou, J. C. B., & Larwood, J. D. 2000, MNRAS, 315, 823
Papaloizou, J. C. B., & Terquem, C. 2001, MNRAS, 325, 221
Pickett, B. K., Cassen, P., Durisen, R. H., & Link, R. 1998, ApJ, 504, 468
———. 2000a, ApJ, 529, 1034
Pickett, B. K., Durisen, R. H., Cassen, P., & Mejia, A. C. 2000b, ApJ, 540, L95
Pickett, B. K., Mejia, A. C., Durisen, R. H., Cassen, P. M., Berry, D. K., & Link, R. P. 2003, ApJ, 590, 1060
Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., & Greenzweig, Y. 1996, Icarus, 124, 62
Quinn, P. J., Hernquist, L., & Fullagar, D. P. 1993, ApJ, 403, 74
Rice, W. K. M., & Armitage, P. J. 2003, ApJ, 598, L55
Rice, W. K. M., Armitage, P. J., Bate, M. R., & Bonnell, I. A. 2003, MNRAS, 339, 1025
Romeo, A. B. 1992, MNRAS, 256, 307
———. 1994, A&A, 286, 799
Stadel, J. G. 2001, Ph.D. thesis, Univ. Washington
Tanaka, H., Takeuchi, T., & Ward, W. 2002, ApJ, 565, 1257
Terquem, C., & Papaloizou, J. C. B. 2002, MNRAS, 332, 39
Thacker, R. J., Tittley, E. R., Pearce, F. R., Couchman, H. M. P., & Thomas, P. A. 2000, MNRAS, 319, 619
Throop, H. B., Bally, J., Esposito, J. L. W., & McCaughrean, M. J. 2001, Science, 292, 1686
Toomre, A. 1964, ApJ, 139, 1217
Udry, S., Mayor, M., Naef, D., Pepe, F., Queloz, D., Santos, N. C., & Burnet, M. 2002, A&A, 390, 267
Wadsley, J., Stadel, J., & Quinn, T. R. 2004, NewA, 9, 137
Ward, W. R. 1997a, ApJ, 482, L211
———. 1997b, Icarus, 126, 261
Weidenschilling, S. J. 1977, Ap&SS, 51, 153
Wetherill, G. W. 1990, Annu. Rev. Earth Planet. Sci., 18, 205
Yorke, H. W., & Bodenheimer, P. 1999, ApJ, 525, 330