

PLANETARY SYSTEMS

Pawel Artymowicz

*Physical Sciences
University of Toronto at Scarborough
1265 Military Trail, Ontario, Canada M2A 1N7*

Abstract.

We provide a non-technical introduction to some current topics in planetary system origins, relevant to the understanding of a decade's worth of discovery and characterization of systems other than the solar system. We consider both the radial velocity exoplanets (planets outside the solar system) and the dust disks known as transitional and debris disks, which are natural and conspicuous parts of young planetary systems. Our emphasis is on new dynamical theories, including new concepts in disk-planet interaction and the resultant three types of migration of protoplanets in disks. Migration is considered the key issue for planet survival during the T Tauri disk phase. We also discuss a variety of observable features arising from the dust-gas coupling in the presence of stellar radiation pressure. Nonaxisymmetric features have been both observed and predicted theoretically, although detailed comparison has not yet been performed.

Keywords: Keywords

PACS: PAC numbers

1. Exoplanets and theories of their formation

Theories and observations are tightly interwoven in any science, but the science of the origin of planetary systems provides one remarkable example of their mutual complementarity and motivation each provides for the other. The history of this subject is as old as physical science itself. Extrasolar planetary systems have been predicted by pure thought, starting from the prescient ideas of atoms and motion-allowing vacuum, by Leucippus and Democritus (cf. Dick 1982) in ancient Greece. The oldest and the newest theories regarding "other worlds" (planetary systems) are in general remarkably similar, each describing both their evolution, including formation from rotating rarified gas which forms a turbulent nebula ("single whirl") where refractory elements fractionate out and form planets, as well as their enormous possible diversity. But on that last point, modern theory is still catching up with the ancient predecessors, spurred into action by the remarkable success of observers who, as of this writing, within the past decade succeeded in finding 188 extrasolar planets (cf. Marcy et al. 2002, Schneider 2006).

In contrast, the theory of the 1980s that we summarize in this brief introductory section (for review see Lissauer 1993), called standard formation scenario, was tailored to the explanation of planets in one system only under the tacit assumption that our system is typical and representative of other. In this model, planetary systems form from the protoplanetary disks (also known as primitive solar nebulae, protostellar accretion disks, or T Tauri disks). Planetesimals, comet-sized primitive rock+ice bodies, formed from dust, accumulate in orbit via binary collisions, and in less than 1 Myr build protoplanets (bodies larger than 1000 km).

Terrestrial planets are finally assembled in the inner solar system within 30 Myr. A crucial role in their growth is played by gravitational focusing, which works best in kinematically cold planetesimal disks. Cooling (circularization of orbits) is provided by the nebular gas drag and by planet's launching of spiral density and bending waves into the disk.

Outside the ice condensation boundary, at a distance of several AU from the sun, protoplanets grew quicker and larger because of the availability of water ice, and grew up to a mass of several Earths, gathering around them a massive hydrogen+helium envelopes. In the "core-accretion" or "core-instability" scenario, the primitive atmosphere becomes unstable and accretes onto the core without mixing, when the core mass exceeds 8-15 Earth masses, a value in line with core masses in our system (see reviews in Mannings et al. 2000 and Reipurth 2006). Jupiter was thought to have been born: (i) at or near its present location, because of ice boundary location, (ii) on a circular orbit ($e \approx 0$), due to circular motion of the protoplanet and the disk, and (iii) with mass determined by the process of tidal gap opening in a viscous disk with parameter $\alpha \sim 10^{-2}$.

1.1 Diversity of solar systems

Although most theoretical concepts from the standard planet formation theory of the 80s are perfectly usable even today, in the hindsight we see how limited some of the predictions regarding extrasolar systems turned out to be. Ironically, the first planets ever found outside the solar system (by Wolszczan and Frail 1992, in an unlikely place: around a millisecond pulsar PSR1257+12) are of terrestrial type and although closer to their sun than our terrestrial planets are to the sun, exhibit neatly near-circular orbits with two major bodies very close to a mean-motion resonance (3:2), not unlike Jupiter and Saturn (5:2). However, a theoretical expectation (Boss 1995) that the solar and extrasolar systems share the basic blueprint (giants outside a few-AU radius, terrestrial planets inside) was soon afterwards disproved by the first 'hot jupiters' such as 51 Pegasi discovered via Doppler shift or, in other words, radial velocity variation, by Mayor and Queloz (1995) to have orbital period of only 4 days. While more data (among others, about long-period systems) are needed before we can be certain that the solar systems is atypical, if not unique, in the universe, the accumulating statistics certainly makes that conclusion very plausible (cf. Marcy and Butler 1998, Marcy et al. 2002, review books by Mannings et al. 2000, and Reipurth et al. 2006):

- Large planetary companions (with $a < 3$ AU and at least Saturn's mass) exist around at least $\sim 10\%$ of normal stars. They can be found at any distance from the star currently observable, with a possible depletion at radii less than 0.5 AU.
- "hot Jupiters" (minimum masses $m > 0.1m_J$, semi-major axes $a < 0.1$ AU) exist around $\sim 1\%$ of sun-like stars. There typical period is 3-4 days. There is a noticeable excess or pile-up of hot jupiters at these periods compared with a few times longer periods, although they are a small minority of all exoplanets.
- Very massive planetary companions are found with the frequency $dN \sim m_i^{-0.9} dm_i$, i.e., the logarithm of mass has a very flat distribution up to the minimum mass $m_i =$

$m \sin i > 10m_J$ (10 Jupiter masses).¹ This decreasing tail of planet-like bodies overlaps with a low-mass tail of massive companions (stars or brown dwarfs with mass 13 to 80 m_J), but the crossover mass is as yet unknown.

- Eccentric planets (i.e., planets on elliptic orbits) are common. There is a clear e - P correlation resembling closely that of the pre-main sequence and main-sequence close binary stars. Orbits with $a < 0.1$ AU tend to have $e \approx 0$, which might be due to tidal circularization by the star. There is a weak direct correlation of e with $m \sin i$. Also, $m \sin i$ correlates positively with distance or orbital period: there is a deficit of massive super-jovian planets closer than about 0.3 AU from the stars and a deficit of sub-jovian planets at more than ~ 0.5 AU (e.g., Papaloizou and Terquem 2005). It is hard to tell which of these correlations are primary (a physical connection acquired at birth or later such as planet migration sensitive to their mass, or eccentricity-inducing stellar flybys frequent far from the host star) and which merely a statistical consequence of the primary ones without any direct connection.

- Most transiting exoplanets have jovian-type radii, first detected in the case of HD 209458b, which has a record-low mean density 0.38 g/cm^3 (Charbonneau et al. 2001, 2006). They must be gaseous planets like our Jupiter rather than giant rocks.

Preliminary models of interior structure based on the transits were constructed. The difficulty in modeling is to know how much of the stellar radiation is absorbed (and where) by the planet. While HD 209458b must have a very small core (or no core at all), this seems to be an exception. Other radial velocity hot jupiters have mean densities in the range 0.7 - 1.5 g/cm^3 (Charbonneau et al. 2006). The preliminary ideas about the internal structure of three exoplanets are contrasted with the data about the solar system giant planets in Figure 1. Notice that in the case of exoplanets, by "core" we mean an estimate of the total heavy element contents, which can either reside in a well-defined core or be mixed with the gas.

HD 149026b has a very substantial core of heavy elements (70 Earth masses), and resembles a super-Neptune rather than Saturn, despite a Saturn's mass. This illustrates just how diverse the exoplanets are internally, not only orbitally. The large core of HD 149026b and the likely 'normal' core of HD 189733b exclude the origin of these planets in a disk fragmentation event (cf. Boss 2000). We feel that this should come as no surprise. Disk fragmentation model (also called Giant Gaseous Protoplanets or GGP scenario) was always appealing because of short formation time, but has had serious problems justifying its assumed unstable initial conditions.²

¹ The unknown inclination of the orbit i does not allow a unique mass determination, giving only the minimum mass $m \sin i$. Assuming randomly oriented orbits, this mass is on the average $\pi/2$ times smaller than the true mass.

² Disks defend themselves from approaching danger of a global breakup by growing spiral, galaxy-like wave modes, which in the non-linear regime dissipate energy of rotation and heat up the disk, thus curbing the instability (e.g., Laughlin and Bodenheimer 1994). This stabilizes galaxies at Safronov-Toomre number $Q > 1$. A very optically thick protoplanetary disk cannot cool fast enough for breakup (on orbital time scale, cf. Rafikov 2005). In fact, only isothermal disk models can be persuaded to fragment. Other assorted difficulties include: difficulty of reproducing typical sub-jovian and jovian planets (GGPs at the outset have jovian masses and are bound to grow by accretion even further during a long disk lifetime, ending up as super-jupiters or brown dwarfs); and finally, difficulty of fractionating a solid core of an exoplanet from a well mixed initial GGP.

**Comparison of gas and rock/ice masses (in Earth masses, M_E)
in giant planets and exoplanets (2006)**

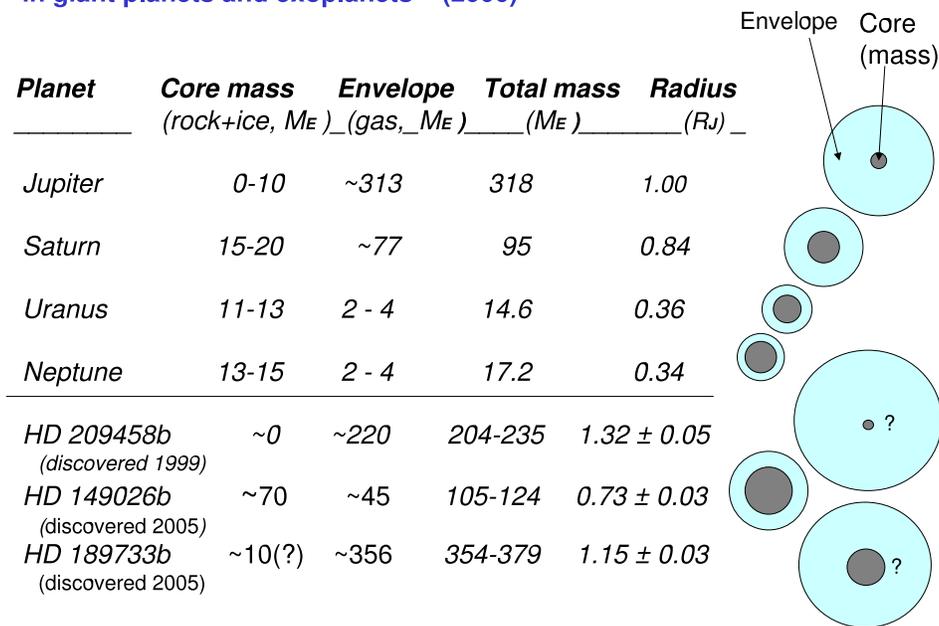


FIGURE 1.

In summary, ironically, pulsar PSR 1257+12 remains the only truly solar-like system, regarding the orbits and terrestrial masses of the planets. We now know some 50 times more(!) giant exoplanets than the giants in our system. The diversity of exoplanets far exceeded our naive expectations before 1995, and makes our system look atypical. Their orbits are too close to their stars, too elongated, or both. Chemical composition of host stars is clearly correlated with the frequency of occurrence of close giant exoplanets. (One example is HD149026b in Fig. 1, which has a star with more than twice the solar metallicity.) It is unclear why this is so, probably the metal-rich protoplanetary disks both produce more planets and enhance their delivery to the inner part of the planetary system (see Reipurth et al. 2006 for a collection of excellent reviews hereafter referred to as PPV or "Protostars and Planets V").

1.2 Dusty disks as young planetary systems

For consolation after the perplexing differences and surprises that radial velocity systems provided we might turn to another area of astrophysical research. For two decades now, and hence significantly longer than for instance the detection of radial velocity variations of stars, astronomers have studied another aspect of planetary systems: dusty circumstellar disks around main-sequence stars. Dusty disks better match the standard expectations about the mineralogy, spatial extent etc. of a planetary system like our own. These disks descend from the primordial solar nebulae that accompany all forming in-

intermediate and low mass stars. Late evolutionary stages of disk systems, following the loss of the hydrogen+helium gas from disks, are observed as a class of infrared-emitting systems known as Vega-type, Vega-excess or β Pic-type systems. Detected by IRAS and ISO satellites thanks to the presence of a large area (up to $\sim 10^{30}$ cm²) of solid grains much different from the ISM dust, complemented by a varying but generally much smaller amount of gas, the Vega-type systems comprise as many as 15% of nearby main-sequence field stars of type A to K (for review see Lagrange et al. 2000).

Such systems were quickly realized to represent a stage in planetary system evolution (e.g., Aumann et al. 1984, Smith & Terrile 1984), although it was not immediately clear which one (some were suggesting a stage of "snowballing" necessary to build large solid particles). An opposite view, that solids in the disks are intensely eroded and reduced to dust, rather than accumulating, now prevails (Artymowicz 1997, Vidal-Madjar et al. 1998, Lagrange et al. 2000). Planetesimals, i.e. comets and asteroids, replenish the dust via ice sublimation and mutual collisions. The disks ages range typically from a few to a few hundreds Myr.

High-resolution imaging has revealed structure in some of the disks implying a possible influence of planetary masses (more about this later). An important finding by Dominik et al. (1998) of a Vega-type infrared excess from 55 ρ^1 Cancri, a star orbited by a hot Jupiter-type planet, provided the first link between radial velocity and Vega-type systems. (Both represent planetary systems, seen from a very different perspective.) The disk has been, however, difficult to spot in the visible/near-IR scattered light; it has not been confirmed in HST imaging despite the initial claim based on Earth-bound observations (Trilling & Brown 1998; Schneider et al. 2000). Imaging of dust in the known exoplanetary systems will be an interesting future area of studies. We may notice that the prototype β Pic was also an isolated example of a scattering disk for about a decade, before improved sensitivity allowed imaging of any other objects, so current problems are temporary. We return to the dusty planetary systems after discussing the dynamics of disk-planet interaction.

2. Disk-planet interaction

The discovery of 'hot jupiters' highlighted the conceptual difficulties associated with forming such planets in situ, either in the critical core mass accumulation followed by gas accretion scenario, or the gravitational instability scenario for giant planet formation. Simply speaking, the disk-planet interaction and the resultant large-scale migration of planets became the only hope for understanding the observations. In this section, we discuss the disk-planet coupling along the lines of the most recent review by Papaloizou et al. (2006).

The idea of migration is not new, as the basic mechanisms were proposed in the early 1980s and 1990s. (see Lin et al. 2000 and references therein). In the standard theory of migration from 1990s, two modes of migration were known: type I and type II, the former applying to small-mass embedded protoplanets and the latter to gap-forming massive protoplanets. Both predicted disturbingly short radial infall times that in the type I case threatened the survival of embryo cores in the $1 - 15M_{\oplus}$ regime before they could accrete gas to become giant planets. At the end of 1990s, the main question

became how to resolve the type I migration issue and how to account for the observed radial distribution of exoplanets.

2.1 Migration type I

When the mass of the protoplanet is small the response it induces in the disk can be calculated using linear theory. Density waves propagate both outwards and inwards away from the protoplanet. These waves carry positive and negative angular momentum respectively and accordingly a compensating tidal torque is applied to the orbit resulting in type I migration. Like other types of migration discussed later, type I migration is a differential effect, a residual of two opposing torques from the protoplanetary disk gas located inside and outside the orbit of a planet.

The tidal torque

When hydrodynamic equations are linearized about a basic state consisting of an unperturbed axisymmetric accretion disk, the response to individual Fourier component (harmonics with azimuthal periodicity or number of arms given by integer m) can be calculated and at the end summed up. The gravitational potential ψ of a proplanet in circular orbit is

$$\psi(r, \varphi, t) = \sum_{m=0}^{\infty} \psi_m(r) \cos\{m[\varphi - \Omega_p t]\}, \quad (1)$$

where φ is the azimuthal angle and $2\pi/\Omega_p$ is the orbital period of the planet of mass M_p at orbital semi-major axis a . The total torque acting on the disk is given by $\Gamma = -\int_{Disk} (\Sigma \vec{r} \times \nabla \psi) d^2r$ where Σ is the surface density of the disk.

An external forcing potential $\psi_m(r, \varphi)$ with azimuthal mode number m that rotates with a pattern frequency Ω_p in a disk with angular velocity $\Omega(r)$ triggers a response that exchanges angular momentum with the orbit whenever, neglecting effects due to pressure, $m(\Omega - \Omega_p)$ is equal either 0 or $\pm\kappa$, with $\kappa \equiv \Omega$ being the epicyclic frequency in a Keplerian disk. The first situation occurs when $\Omega = \Omega_p$ and thus corresponds to a corotation resonance (CR). The second possibility corresponds to Lindblad resonances (LR): an inner Lindblad resonance (ILR) for $\Omega = \Omega_p + \kappa/m$ and an outer Lindblad resonance (OLR) outside the planet's orbit for $\Omega = \Omega_p - \kappa/m$.

Lindblad resonances

Torques at LRs (Goldreich and Tremaine 1979, 1980) are of crucial importance for the standard theory of disk-planet interaction, and provide a dominant effect for: disk gap opening, shepherding of planetary rings by satellites, migration, as well as planetary eccentricity damping in the type I situation (Ward 1986, Artymowicz 1993ab). The reason for neglecting CRs is that in the linear theory with a planet on a fixed orbit they have been shown to be subdominant to LRs, e.g., by Tanaka et al. (2002).

The torque arising from the component of the potential with azimuthal mode number m is found, for a Keplerian disk, in standard form reads (Goldreich and Tremaine 1979)

$$\Gamma_m^{\text{LR}} = \frac{\text{sign}(\Omega_p - \Omega)\pi^2\Sigma}{3\Omega\Omega_p}\Psi^2, \quad (2)$$

with

$$\Psi = r\frac{d\psi_m}{dr} + \frac{2m^2(\Omega - \Omega_p)}{\Omega}\psi_m. \quad (3)$$

where the expression has to be evaluated at the location of the resonance. Torque exerted on the planet from an outer Lindblad resonance is negative and causes its inward migration, while the torque due to an inner Lindblad resonance is positive corresponding to an acceleration and outward migration.

Let us denote by c_s the sound speed, by $r\Omega$ Keplerian speed, and by $h = c_s/(r\Omega)$ the inverse Mach number in disk, also equal to the geometrical aspect ratio of the disk, or the ratio of disk's half-thickness H to radius r . The total torque is obtained by summing contributions over m from $m = 1$ to $m = \infty$. Contributions fall dramatically above a limiting value of m equal $m \approx h^{-1}$. We can express this pressure torque cutoff condition as $\xi = mh = mc_s/(r\Omega) > 1$. For a finite ξ the true (as opposed to nominal) positions of the Lindblad resonances can be found from a WKB dispersion relation of density waves

$$m^2(\Omega - \Omega_p)^2 = \Omega^2(1 + \xi^2). \quad (4)$$

The effective positions of the resonances are shifted with respect to the standard theory. If a denotes the orbital radius of a planet, an m -armed potential $\psi_m(r)$ creates nominal positions of ILR/OLR at $r_{LR,nom} = a \mp 2a/3m = a \mp (2H/3)\xi^{-1}$, where $H = c/\Omega$ the disk's semi-thickness or scale height obeying $H \ll r$. In Goldreich-Tremaine theory LRs converge to $r_{LR,nom} = a$ for high-order harmonics of the potential. (This caused some spurious infinities in initial calculations of type I migration.) The effective resonance positions, where the potential actually excites the waves and around which the waves have their first wavelength, are given by (Artymowicz 1993a)

$$r_{LR} = a \mp (2H/3)\xi^{-1} \sqrt{1 + \xi^2}. \quad (5)$$

In particular, when $m \rightarrow \infty$, Lindblad resonances pile up at $r = a \pm (2H/3)$ instead of $r = a$.

The exciting potential $\psi_m(r)$ becomes so localized for $\xi \rightarrow \infty$ and the pressure-shifted LR positions, where the wave-potential coupling takes place within the first wavelength, so far from the peak of exciting potential (at $r = a$) that the resulting torque on planet decreases exponentially with m . Torque cutoff has good analytical approximations useful for disk plus point mass problem (Artymowicz 1993b).

Using those analytical approximations, Ward (1997) compared the torques due to individual ILRs with those due to OLRs. We show in Figure 2 this comparison for a very small planet that does not perturb the surface density significantly (Papaloizou et al. 2006). The main reason for different shapes of the ILR and OLR torque curves is that these resonances belonging to the same m are not located symmetrically with respect

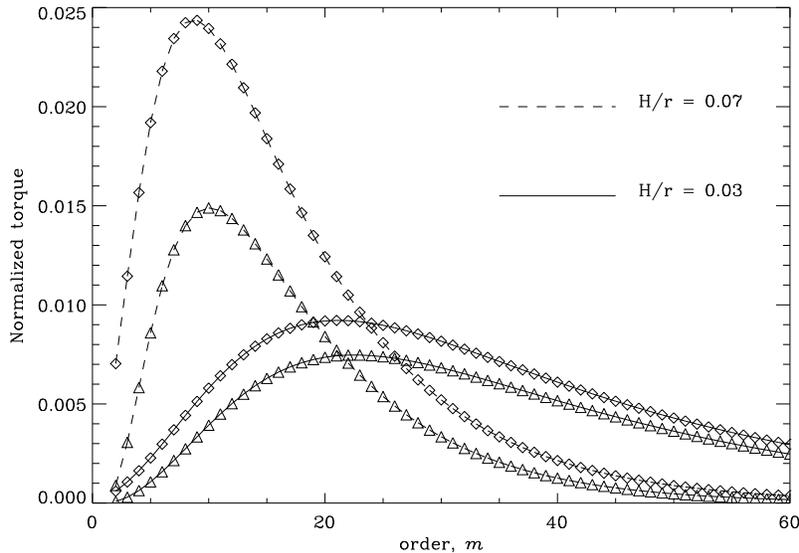


FIGURE 2. Individual inner and outer torques (absolute value) in a $h = 0.07$ and $h = 0.03$ disks, as a function of m . For each disk thickness, the upper curve (diamonds) shows the outer torque and the lower one (triangles) the inner torque. These torques are normalized to $\Gamma_0 = \pi\mu^2\Sigma a^4\Omega_p^2h^{-3}$, where $\mu = M_p/M_*$.

to the planet's orbital radius. We can see a clear dominance of OLRs over ILRs, that is the dominance of outer disk over the inner disk. Virtually independently of the assumed unperturbed density profile $\Sigma(r)$, the differential torque results in an *inward migration* of a protoplanet (Artymowicz 1993b, Ward 1997, Tanaka 2002).

The most recent linear calculations by Tanaka et al. (2002) that take into account 3D effects, and are based upon the value of the total tidal torque, including the corotation torque (fully unsaturated since it is a linear estimate), give a migration timescale

$$\tau \equiv a/\dot{a} = (2.7 + 1.1\alpha)^{-1} \frac{M_*^2}{M_p \Sigma a^2} h^2 \Omega_p^{-1}, \quad (6)$$

for a surface density profile $\Sigma \sim r^{-\alpha}$. Notice that the all tidal torques scale with the square of the perturbing mass (disk response to planet scales as one power of M_p , force on planet from the disk scales as another M_p). Inertia of the planet scales as M_p , hence the acceleration and migration rate scale linearly with M_p , while the timescale as $\tau \sim M_p^{-1}$ (cf. the above equation). For an Earth-mass planet around a solar mass star at $r = 1$ AU, in a disk with $\Sigma = 1700 \text{ g cm}^{-2}$ and $h = 0.05$, we obtain $\tau = 0.16$ Myr.

A number of sophisticated numerical calculations with planets on fixed circular orbits have shown a good agreement with the linear theory (for review see Papaloizou et al. 2006). Only at Neptune's mass, protoplanets in numerical calculations sometimes migrate significantly slower than the linear prediction, and of course after gap opening, which occurs a higher mass, the migration slows down. However, type I is a grave danger to Earth-type planets which nominally all should travel toward the star in much less than the disk lifetime, which equals several Myr. The apparent robustness of type I migration

in our theories creates potential difficulties for the accumulation scenario for the critical jovian cores.

Possible solutions to the seemingly unstoppable migration conundrum include:

(i) unusual LR location and strength if gas is strongly magnetized (Terquem 2003, but no guarantee this always helps and not worsens the situation);

(ii) the disk self-gravity modifying the spiral density patterns and thus changing the pull of the disk on the planet (has been modeled by Nelson and Benz 2003, Pierens and Huré 2003, but does not make a big difference and actually speeds up migration);

(iii) hypothetical high eccentricity of a planet might slow or reverse migration (Papaloizou and Larwood 2000, but there is little support for such an eccentricity);

(iv) detailed disk thermodynamics and radiation transfer including shadows cast by the disk around the planet on itself (a moderate slow-down of migration was found by Menou and Goodman 2004, Jang-Condell and Sasselov 2005);

(v) MHD turbulence was proposed to disrupt type I migration and replace it with a random walk (Nelson and Papaloizou 2004; however, would it not be an asymmetric random walk with precisely the same consequences?);

(vi) One tantalizing possibility is that migration type III (see below) supersedes type I. (3-D calculations of this problem are needed.)

2.2. Migration type II

When the planet grows in mass the disk response cannot be treated any longer as a linear perturbation, nor the unperturbed surface density taken as smooth power law. Nonlinear numerical hydrodynamics such as the calculation shown in Figure 3 show that a deep gap around the planet opens for standard disk parameters and planet's mass ratio similar to Saturn, and generally between Neptune's and Jupiter's ($\mu = M_p/M_* = 10^{-4} - 10^{-3}$). Several, not always compatible, criteria for this event have been proposed, so let us enumerate a few.

Lin and Papaloizou (1993) proposed two criteria, a thermal and a viscous criterion. The thermal condition requires that the planet's gravity be strong enough to overwhelm pressure in its Roche lobe radius (Hill sphere radius), $r_L = (\mu/3)^{1/3}r$, or in other words that $r_L > H$ (where we have $h = H/r$). After raising both sides of equation to the 3rd power, it reads

$$\mu > 3h^3. \quad (7)$$

This condition used to be questionable derived from the Rayleigh instability concept, with which it has no connection; afterwards the justification involving nonlinearity of the disk response was substituted.³ This second justification is also problematic (there are numerical examples violating the thermal criterion, where the gap exists without the criterion being satisfied. The concept of nonlinearity may either be irrelevant (viscosity

³ Papaloizou et al. 2006 write that the flow perturbation becomes non-linear and the planetary wake turns into a shock in its vicinity. Dissipation by these shocks as well as the action of viscosity leads to the deposition of angular momentum, pushes material away from the planet and a gap opens.

alone may, in principle, provide for damping of *linear* density waves, or at least not very useful since weakly nonlinear shocks do accompany even an Earth-mass object, which clearly does not open a gap.) We see the status of the thermal criterion as uncertain.

The condition that angular momentum transport by viscous stresses be interrupted by the planetary tide is approximately

$$\mu > 40 \frac{v}{r^2 \Omega} \quad (8)$$

if one follows the derivation by Lin and Papaloizou (1993), where use is made of an uncertain assumption that the distance at which Lindblad torques cease to act is equal to one Roche lobe. That is not accurate, as the CR zone is in reality about 2.5 times wider; moreover if $(2H/3) > 2.5r_L$ then $2H/3$ takes over the role of the limiting minimum distance of interaction and leads to the following viscous criterion

$$\mu^2 h^{-3} > 100 \frac{v}{r^2 \Omega}, \quad (9)$$

which happens to be virtually equal to the side-by-side product of the standard thermal and viscous criteria, eqs. (7) and (8). Essentially, all those criteria appear to be useful and approximately correct in practice, because they are mutually dependent, and in addition even the independent criteria like eqs. (7) and (8) accidentally coincide for the parameters corresponding to the Jupiter in a Minimum Mass Solar Nebula. This may not be a general result, so attention to the correct derivation and form of the gap opening criterion is needed. In that respect eq. (9) might be preferable, as it allows for the role of both gravity, gas pressure, and viscosity. Indeed, Crida et al. (2006) analyze this simultaneous dependence on all the physical processes deeper than we can do it here.

Once the gap is open, the planet is traveling centered in it, as if shepherded by the two disk edges, following their motion with the disk material accreting slowly onto the star. The time scale of migration type II is, roughly, the viscous timescale $\tau_{II} \sim r^2/v \sim 0.1$ Myr from the distance of Jupiter, $r = 5.2$ AU. While still shorter than the lifetime of primordial disks, this timescale is formally longer than τ of embedded planets.

There are two important exceptions to this description of type II migration. One could be called migration type IIb. It happens when the gas in the disk immediately surrounding and interacting with the planet becomes depleted to the point that the planet is more massive than that gas. Migration slows down below the type II speed, approximately in proportion to the gas:planet mass ratio. The second exception is when the LR torques responsible for centering the planet in the gap become weaker than the CR torques from the CR zone: the centering and symmetry of type II has been observed to spontaneously break down: the situation described below as migration type III.

Numerical modeling and how to test it

Many important results on the behavior of bodies embedded in disks have been obtained through computations. The corresponding migration regime is called type II migration (e.g., Lin and Papaloizou, 1986; Ward, 1997). In fact, what comes out

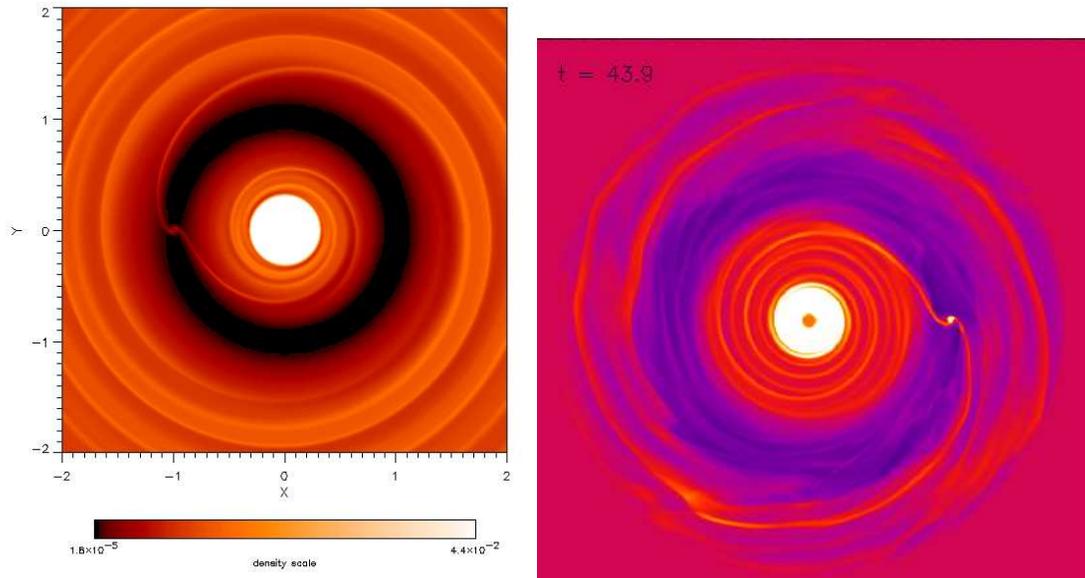


FIGURE 3. Surface density profiles for an initially axisymmetric disk model about one hundred orbital periods after the introduction of a planet, which remains on a prescribed, circular orbit. The mass ratio is a Jupiter's $\mu = 10^{-3}$. The program in the left image is ZEUS (2nd order scheme with artificial viscosity), on the right PPM (higher order scheme, no explicit viscosity). Notice different amount of vorticity and wavelets in the disk.

of the numerical calculations is often immediately interpreted in terms of standard type II migration. That, after a closer look, is unlikely. The simulated drift is likely a combination of type II and type III.

Medium-resolution hydrodynamical calculations (several hundred by several hundred cells) of planet-disk interaction in the type II regime were performed by Bryden et al. (1999) and Lubow et al. (1999). Since protoplanetary accretion disks are assumed to be vertically thin, these first simulations used a two-dimensional ($r - \phi$) model of the accretion disk. The vertical thickness H of the disk is incorporated by assuming a given radial temperature profile $T(r) \propto r^{-1}$ which makes the ratio H/r constant. Typically, simulations assume $H/r = 0.05$ so that at each radius, r , the Keplerian speed is 20 times faster than the local sound speed. Initial density profiles typically have power laws for the surface density $\Sigma \propto r^{-s}$ with s between 0 and 1.5. More recently, fully 3D models have been calculated using the isothermal equation of state (D'Angelo et al. 2003a). Spiral arms are weaker and accretion occurs primarily from regions above and below the midplane of the disk.

The viscosity is dealt with by solving the Navier Stokes equations with the kinematic viscosity ν taken as constant or given by an α -prescription $\nu = \alpha c_s H$, where α is a constant. From observations of protostellar disks, values lying between 10^{-4} and 10^{-2} are inferred for the α -parameter but there is great uncertainty. Some codes require also additional artificial viscosity for proper post-shock oscillation damping; a popular code ZEUS belongs to this category. All of them also have a small spurious viscosity in the form of resolution-dependent diffusivity (sharp features cannot propagate on grids

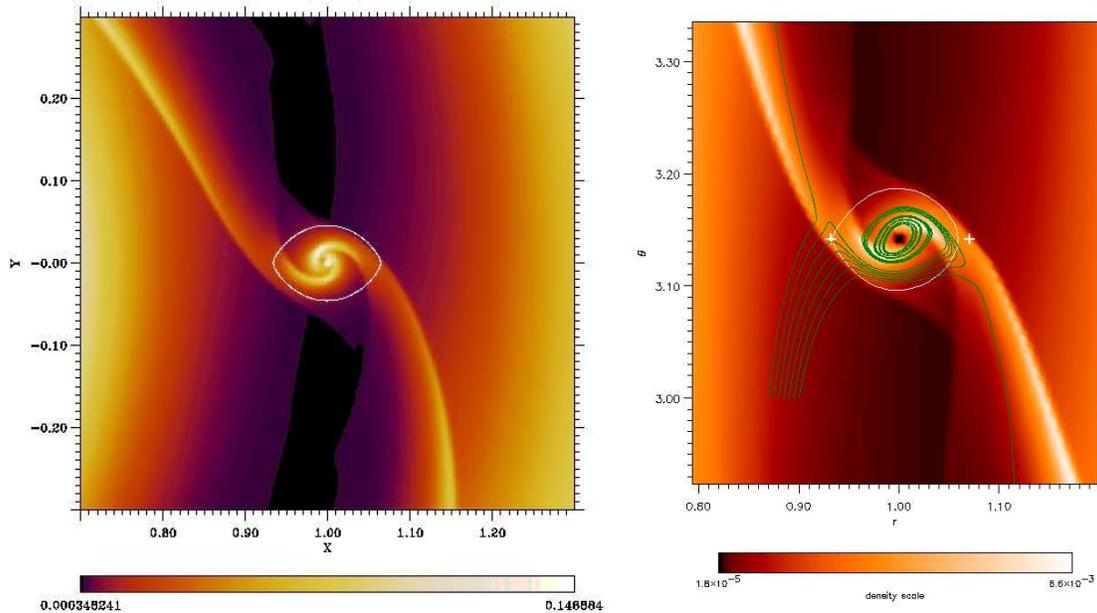


FIGURE 4. Left panel: Jupiter’s vicinity in a global simulation on a variable-resolution grid using Piecewise Parabolic Method (PPM) by Artymowicz (1999, unpubl.). Surface density of disk gas is color-coded. Notice planetary wakes in the disk gap and the complex set of shocks near the well-resolved Roche lobe (white oval). Right panel: Streamlines of gas superimposed on shaded density plot in Lubow et al.’s (1999) calculations. The flow splits into a part that returns to the disk, and a part that enters the Roche lobe, to eventually be accreted by a planet occupying the central sink cell.

freely without being washed out with time); measured amounts of that effect are small. Full MHD-calculations have shown that the viscous stress-tensor ansatz may give (for sufficiently long time averages) a reasonable approximation to the *mean* flow in a turbulent disk. The embedded planets are assumed to be point masses (using a softened potential). The disk influences their orbits through gravitational torques which cause orbital evolution. The planets may also accrete mass from the surrounding disk.

A typical result of such a viscous computation obtained with a hydrocode based on ZEUS second-order finite differences algorithm (Lubow et al. 1999) is displayed in Fig. 3 (left panel). Here, the planet with mass $M_p = 1M_{Jup}$ and semi-major axis $a = 1$ is *not* allowed to move and remains on a fixed circular orbit, an approximation which is made in many simulations. Clearly seen are the major effects an embedded planet has on the structure of the protoplanetary accretion disk: spiral wave patterns with tightness of the spiral inversely proportional to the temperature (i.e. $h = H/r$). It is interesting and a bit disconcerting to realize how much difference we can see between second-order, viscous calculation and a higher-order, non-viscous flow simulated with PPM. J

Figure 4 presents gas flow near the planet in two different numerical methods. The PPM model on the left does not allow accretion, while the right-hand panel presents a model allowing a fast accretion of the gas. Despite this diverging assumptions about the gas accretion, flow and gas features are similar.

Assessing the accuracy and robustness of various numerical hydrodynamics algorithms and their implementations is a difficult task. We do not know analytical solutions

to realistic multidimensional problems we face in disk-planet interaction. We cannot simulate a high-Reynolds number, supersonically rotating fluids in laboratory to make a direct comparison or test the codes against experimental data, which has always been important as a method of validation of hydrocodes. We hope to achieve a numerical convergence to a stable answer when we increase the number of resolution elements (cells) to our calculation, but the problems and algorithms are complicated enough for there to be no mathematical proof of convergence to a unique, let alone the correct solution. In other words, even if each code converges, they may still converge to different, wrong flows. Often there is no unique way to implement the method, or to set up a given calculation. There is usually a choice of coordinate system, making it rotating or nonrotating, in the disk-planet problem. There is always a choice of the length of timesteps. We would like to know which choice is 'best'. We also need to be able to tell numerical instabilities or problems (sometimes called features) with the code from a genuine physical processes. One example of the latter could be whether or not a planet embedded in a disk is a giant vortex generator.

To address some of the above concerns and to make a first wide ranging comparison of codes, a comparison project was organized by (but not restricted to) the European Research and Training Network "Planet formation". As many as 17 groups participated and the results have been published by de Val Borro et al. (2006). The test problem was actually a small number of different 2-D calculations with prescribed initial conditions (smooth-density axisymmetric gas nebula between 0.4 and 2.5 times a , the planet-star distance serving as unit distance), a softened-gravity planet on a fixed circular orbit (softening radius ϵ in the formula for planet's potential $\psi(r) = -GM_p/\sqrt{r^2 + \epsilon^2}$ equal to 0.6 times the disk scale height H), boundary treatment (non-standard, to avoid wave reflections, of escape of gas from the disk) and other details such as a prescription for a gradual introduction of a planet into the disk to avoid persistent transients. Two mass ratios were studied ($\mu = 10^{-3}$ and $\mu = 10^{-4}$) to approximate the masses of Jupiter and Neptune relative to the sun. Left completely open were the choices of a numerical method, the geometry of the grid (if any) and some minor computational details as required by the codes. In each case, the grid resolution around the planet was to be effective the same: square grids with about 64 cells per unit of radius.

Figure 5 presents the similarities and differences between 15 different numerical methods, the names and specifics of which are described in de Val Borro et al. (2006) and the URL address cited in the list of references, but can be ignored here). Each panel is a polar coordinate plot with radius on the horizontal axis and the azimuthal angle in units of π radians on the vertical. A tiny black dot at position (1,0) represent the planet of Jupiter mass ratio, at the time of 100 P (orbital periods) from its gradual introduction into the calculation over the first 3 periods. The color scale is identical in all figures, and covers the $\log \Sigma$ between -1.65 and +1 (in arbitrary units, equal to the initial density of the disk). Some important conclusions can be glimpsed from the figure. all the numerical codes do not provide the same answers. While the gap width is fairly uniform, some simulations produce a deeper gap than others. With the exception of a particle method SPH in the lower-right corner all the codes find rarified gas blobs surrounding Lagrange triangular points L4 and L5 at positions $(1, \pm 1/3 \cdot \pi)$, and the characteristic spiral wakes extending into both the inner and outer disks from the vicinity of the planet. Some codes,

especially of the second-order finite volume kind, produced from 1 to 3 vortices at the outer edge of the gap (seen as red blobs at radius $r = 1.25 - 1.5$). Some codes produce so strong vortices that the vortices are accompanied by non-linear, sheared spiral wakes, and visibly distort the edge. On the other hand, survival of large vortices is likely an artifact of a 2-D modeling, which does not extend to 3-D models.

Figure 6 compares 15 axisymmetrized profiles of density, $\Sigma(r)$, at time $100 P$. Most profiles are impressively similar, apparently more so than the detailed and time-dependent views of the disk alone. Torques have also been evaluated and with some exceptions are within 50% of each other. One of the reasons for this relatively good agreement is, however, a rather large smoothing constant ϵ used. This type of a comparative research will be of much practical value for those who try to derive general truths from a limited number of numerical models.

2.3 Migration type III

The terminology 'type III migration' refers to migration for which an important driver is material flowing through the coorbital region. If type I denoted embedded migration, type II the open-gap migration, then type III could be thought of as partial-gap migration. However, neither its nature nor rate are in any sense an interpolation of type I and II migration. Rather, migration type III can be orders of magnitude faster than either of its cousins. For brevity, we shall often just call it fast migration.

Corotational flows

In a frame which corotates with the perturbation pattern such as a planet, an inviscid gas on orbits closer to a planet than approximately $2.5r_L$ (2.5 Hill sphere radii) describes horseshoe and tadpole trajectories in motion trapped and librating with respect to the planet. In a local approximation similar to the Hill's problem of celestial mechanics, the streamlines have an analytical approximation shown in Figure 7 (left-hand panel). These streamlines approximate the more exact streamlines shown in the right panel of Fig. 4. To the lowest order, however, the flow and its density distribution remain point-symmetric about the planet and the time-average torque on the planet is zero even if the material is distributed asymmetrically at first. This phenomenon is known in literature as CR saturation. One important result that the approximate theory of flow gives us is the estimate of the half-width of the CR region, or the separatrix distance x_s . It is very weakly dependent on migration and in practice always very close to $2.5 r_L$.

Symmetry breaking and the induced CR torque

This symmetry is, however, broken as soon as the planet starts to migrate. Initial motion or gradient of density in a disk remove saturation and give rise to a new mode of migration, in which the radial drift causes the gas flow to exert on the planet a finite

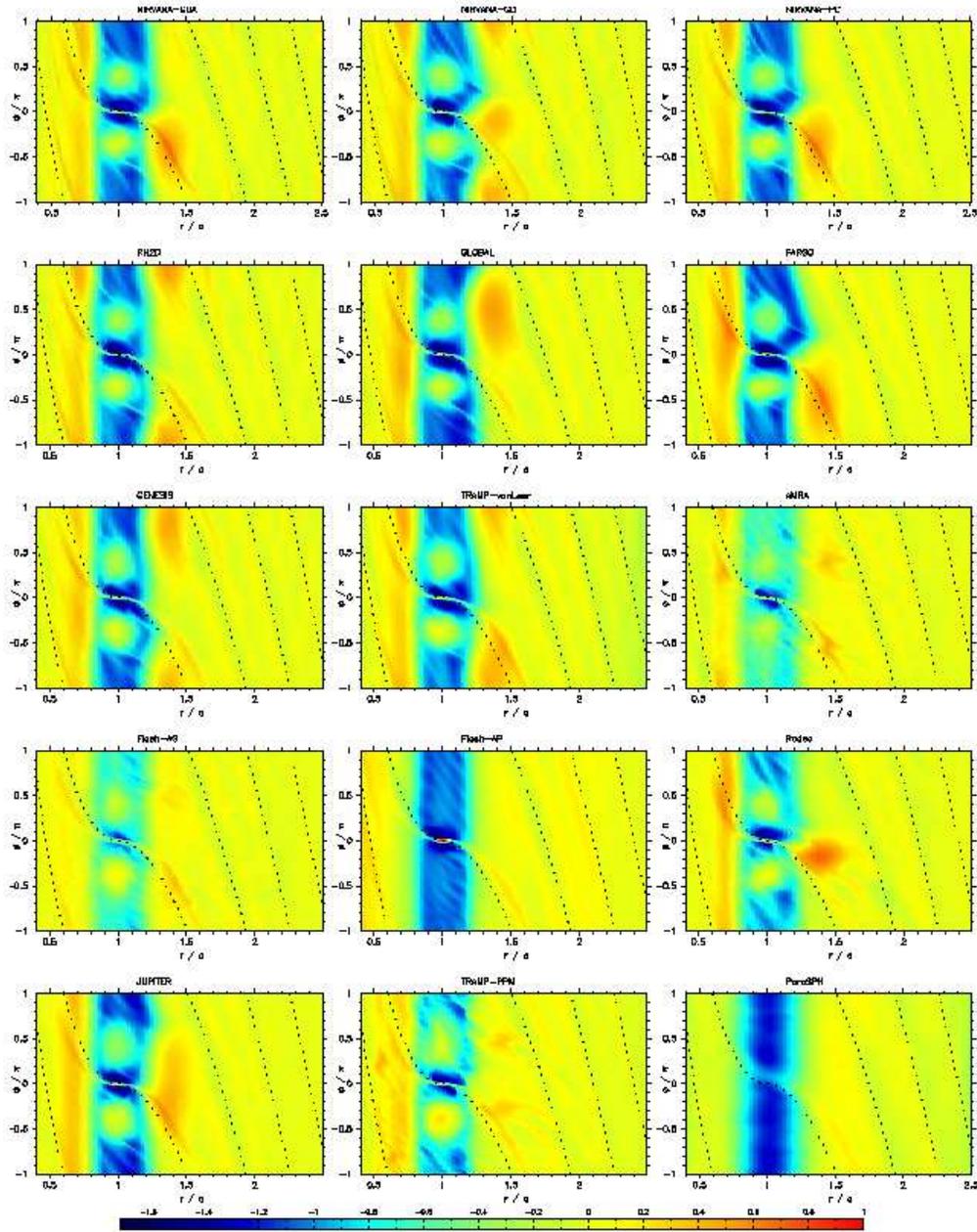


FIGURE 5. Comparison of surface density profiles in 15 numerical hydrodynamics codes (de Val Borro et al. 2006). See text for explanation.

induced torque which *supports* the migration. Whether or not this leads to a large enough positive feedback to support fast migration type III, depends on the density difference between the solid-line disk region and the dashed-line corotational regions in Fig. 7, as well as the speed of the asymmetry-causing migration. The right-hand panel of this figure shows a case of outward migration of the planet on a

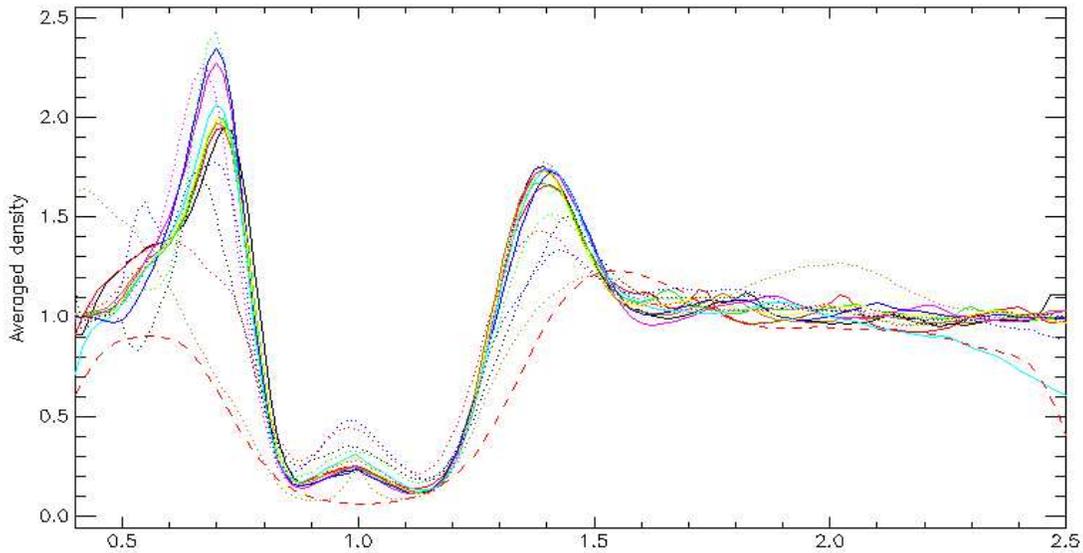


FIGURE 6. Comparison of axisymmetrized surface density profiles in 15 numerical hydrodynamics codes (from de Val Borro et al. 2006). See text for explanation.

circular orbit by definition does not have any motion in this corotating frame, but its motion in the inertial frame has an outward velocity component \dot{a} , so that the distant disk material in the comoving frame is endowed with the equal and opposite radial speed $-\dot{a}$. This causes the distant trajectories to bend into the shape of parabolae. Originally (at $\dot{a} = 0$) given by vertical lines following the shearing-sheet-like velocity distribution (streamlines pointing downward when $x > 0$ and upward at negative x), for finite \dot{a} all of them have an equal amount of negative speed ($-\dot{a}$) added along the x axis. Identical symmetry breaking applies if the disk rather than the planet has radial motion due to viscosity, because the relative motion is still the same. What matters is how much mass crosses the CR region, exchanging angular momentum with the planet, from the outer to inner disk, and vice versa.

The sign of the effect is easy to deduce from the right panel of Fig. 7. Since the dashed-line librating region, slaved to the planet, is of lower density than the ambient disk, either due to initial conditions or the creation of a gap by a planet, or both, there is more material in front of the planet, i.e. in the direction of its motion ($y > 0$) than behind the planet. The excess material in front exerts a positive gravitational pull (torque) on the planet, thus providing positive feedback to the outward migration assumed at the outset.

On a historical and pedagogical note, it was rather difficult to visualize the possibility of type III migration purely theoretically, because all the theory used to be done in the inertial or uniformly rotating but not expanding frame of reference. It is also much easier to fix a planet on a circular or elliptic orbit than to let it evolve freely. The expectation that migration rates will be four orders of magnitude below the orbital speed certainly argued for fixing the orbit completely to avoid any drift due to numerical artifacts. In fact, when W. Kley and the present author collaborated in the mid-1990s on allowing the simulated stellar companions including planets to freely evolve in m , a , and

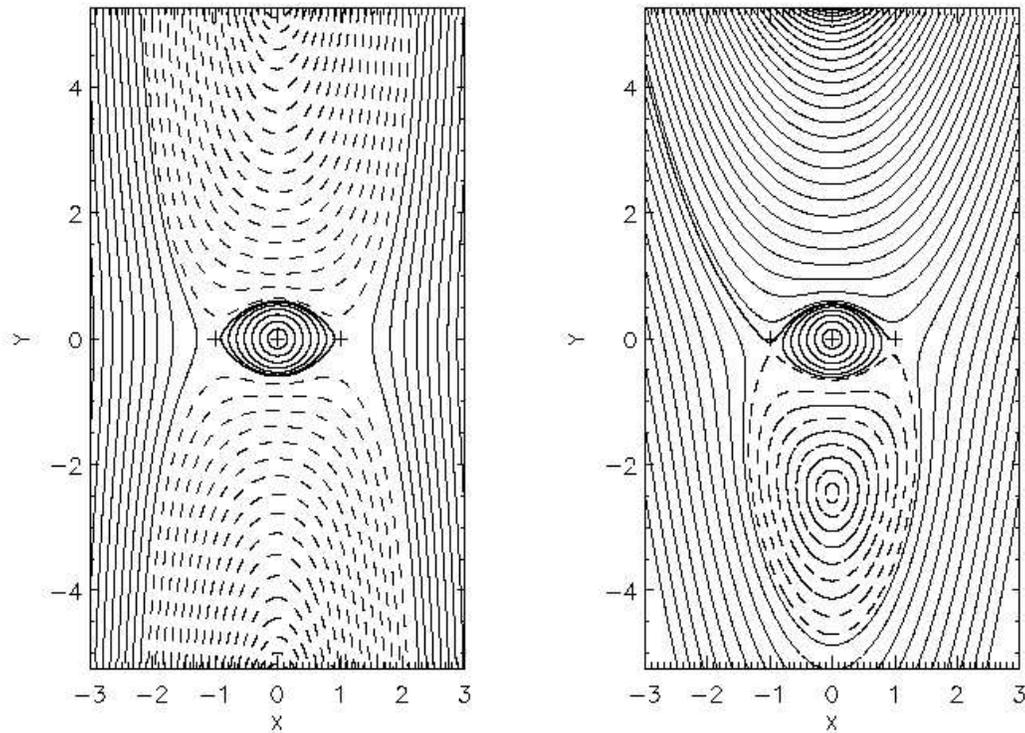


FIGURE 7. Fluid streamlines around a planet in local coordinates (x – radial coordinate, y – azimuthal; distances in Roche lobe (Hill) radii; the planet is located at $x = 0$ and $y = 0$.) Horseshoe region of closed streamlines defining corotational zone is shown by dashed lines. The left panel shows the symmetric situation in the case $\dot{a} = 0$, the right panel asymmetric trapped zone in libration w.r.t. planet for $\dot{a} = +\dot{a}_f$.

e (mass, semi-major axis and eccentricity), they stumbled upon an unexplained and quite abnormal inward migration rate of giant planets in their model, the characteristic time of which was about $50 P$ with very little dependance on planet's mass. The code was scrutinized briefly for programming bugs, nothing was found, and the project was abandoned. In the hindsight, it was most probably an unrecognized migration type III. Dismissing unexpected phenomena must be done judiciously. Other researchers also recognized migration type III only a posteriori. publications by other researchers, after being originally explained away as type II migration. It pays to remember that coincidence of migration rates or other results with prior expectations based on preconceived notions of what is the dominant physical process, might be just that - a coincidence.

Masset and Papaloizou (2003) studied the limit of slow migration and concluded that the induced torque scales linearly with \dot{a} . They concluded that this situation leads to a possible runaway, an exponentially growing \dot{a} in sufficiently massive disks. They proposed the name "runaway migration", which turned out slightly misleading. Artymowicz (2004) independently found a fast but not exponentially unstable migration. He later provided (Artymowicz 2006) a more general formula valid for torque in slow and fast migration in which the torque saturates at \dot{a} . In each particular disk a planet may have

zero, one or two *stable migration speeds*, depending on the initial conditions of motion. For instance, inward migration might be as likely as outward migration. Here is the analysis in a nutshell.

The specific angular momentum that a fluid element near the separatrix takes from the planet when it switches from an orbit with radius $a - x_s$ to $a + x_s$ is $\Omega a x_s$ where x_s is the radial half width of the horseshoe region estimated to be $2.5 r_L$ (Artymowicz 2006, and Fig. 7).

Imagine the simplest possible situation, that the disk and the CR libration regions both have constant densities and that their difference is denoted by Σ_Δ . If $\dot{a} > 0$ but small, the CR region retains contact with the planet both in front and behind it. The orbital drift will, however, make the orbital radial jump (or U-turn) ahead of the planet (at larger y) smaller than the corresponding jump behind the planet, by a difference that we shall denote by Δ . Essentially, in our simple model the radial jump behind the planet, on the separatrix or the last librating orbit in the right panel of Fig. 7), will always be the same x_s , while the jump in front will diminish until it disappears altogether at a characteristic value of \dot{a} which we denote \dot{a}_f and call fast migration speed. At that speed Δ reaches a maximum equal x_s and cannot grow further. The torque between the planet and disk saturates at the fast migration speed, which can be written as (Artymowicz 2006)

$$\dot{a}_f = \frac{3x_s^2}{8\pi a} \Omega_p, \quad (10)$$

while the synodic jump difference assumes the form

$$\Delta = x_s \left(1 - \sqrt{1 - |\dot{a}|/\dot{a}_f} \right), \quad (11)$$

valid for all $|\dot{a}| < \dot{a}_f$ (otherwise $\Delta = x_s$). If the planet does not migrate ($\dot{a} \rightarrow 0$) then Δ vanishes as required. The knowledge of Δ is equivalent to the knowledge of the disk-planet torque, and an induced migration rate that would result from that torque. These quantities will be proportional to Δ and to the surface density deficit Σ_Δ .

We use the nondimensional mass deficit parameter M_Δ defined as

$$M_\Delta = \frac{4\pi a x_s \Sigma_\Delta}{\mu M}. \quad (12)$$

More specifically, M_Δ is the mass that a full CR annulus of width $2x_s$ and area $4\pi a x_s$ would have, if filled with density difference Σ_Δ , normalized to the planet's mass $\mu M = M_p$ (M being the stellar mass).

The final step of the calculation is to equate the assumed and the induced (resultant) migration speeds. As a result, we obtain an *equilibrium*, steady migration speed for every possible Σ_Δ . These results are plotted in our Figure 8 (left panel). Imagine increasing slowly the mass deficit. Up to the point when $M_\Delta = 1$, the planet has insufficient feedback from the disk to migrate rapidly, and since the CR torques were the only ones taken into account, the planet has zero migration speed.

We can improve our simple theory by adding torques due to differential LRs, the same ones responsible for migration type I or II. Then the inward/outward symmetry

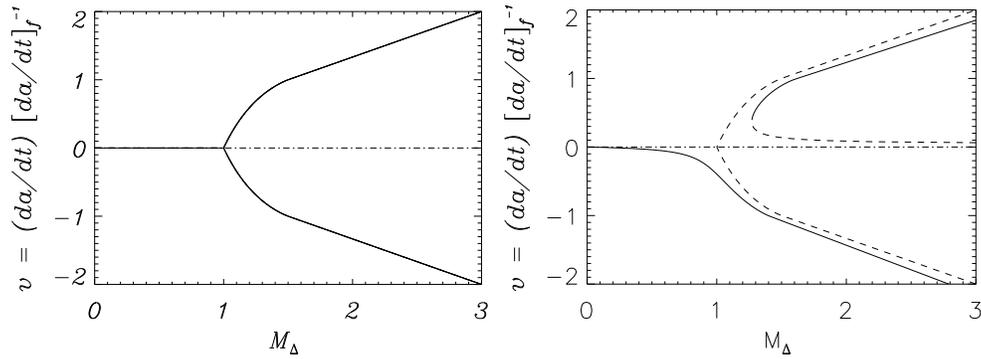


FIGURE 8. Bifurcation diagrams in a simple theory accounting for CR torques only (left panel), and a theory with the LR contributions added (right panel). The bifurcation to fast migration occurs at $M_\Delta = 1$ to 1.3.

is broken and the bifurcation diagram take the shape shown in the right panel of Fig. 8. We see that the abrupt loss of stability by the non-migrating equilibrium is replaced by a more gradual transition from type I/II migration for $M_\Delta < 1$ to the full speed of type III migration, plus the contribution from LR torques, provided that the migration starward. The outward migration, on the other hand, must undergo an abrupt jump from fast to slow migration when approaching the point $M_\Delta \approx 1.3$ from the right-hand-side. The fast migration speed \dot{a}_f is the characteristic speed of incipient migration type III when the disk conditions are marginally supportive of it. Higher speeds can be achieved for more massive disks or more empty corotational regions. What are these quantities in practice? A quick calculation tells us that in order to support fast migration of a Jupiter, the disk should be somewhat more massive than a Minimum Mass Solar Nebula. Saturn should be able to migrate easily in just twice the minimum nebula. The speed \dot{a}_f in the case of Jupiter is capable of changing the position of a planet by a factor of 2 in only 44 orbital periods! However unrealistic such a short time might appear, it is very well confirmed by direct numerical simulations at high resolution. Peplinski and Artymowicz (2006) substituted the time-dependent mass deficit parameter measured in the course of the simulation of a Jupiter in a massive solar nebula (4 times the MMSN). The evolution traced approximately the shape of a bifurcation diagram. The planet traveled outward at a speed higher than \dot{a}_f , reached the vicinity of the numerical grid boundary, lost its underdense libration region, and started migrating inward, again obeying the equilibrium speed at each radius, to within 25 percent accuracy. A snapshot from one of the PPM calculations of a large-scale migration is shown in Figure 9.

Consequences for planet formation

Fast migration, for the same disk profile and planet mass, can be directed either outwards or inwards, depending on the initial conditions. This type of planetary migration is found to depend on its migration history, the “memory” of this history being stored in the way the horseshoe region is populated, i.e. in the preparation of the coorbital mass deficit. Note that owing to the strong variation of the drift rate, the horseshoe streamlines

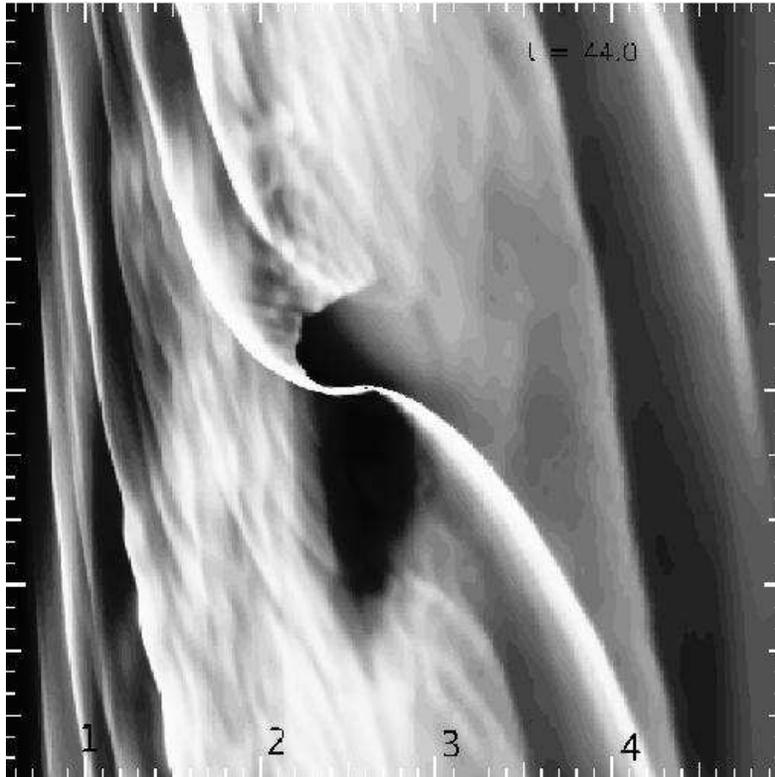


FIGURE 9. Density of gas in a PPM simulation using one variable-resolution mesh. Jupiter-mass planet rapidly migrates in a disk 2.5 times more massive than the Minimum Mass Solar Nebula (Artymowicz 2006). Vertical axis covers the full azimuthal angle, the horizontal is radius in units of initial star-planet distance. After being placed on a positive density gradient, the planet migrated outwards by a factor of 2.6 in only 44 orbits. A trapped coorbital region of low density (dark shade) is clearly visible below the planet (barely visible itself in this plot at radius 2.6). Migration speed corresponds to $M_{\Delta} \approx 3$.

are not exactly closed, so that the coorbital mass deficit can be lost and the runaway can stall. This has been observed in some numerical simulations, whereas others show sustained fast migration episodes for Saturn or Jovian mass planets that can vary the semi-major axis by large factors in less than 100 orbits (e.g., see Figure 9). To date, it is still unclear whether migration type III can in principle continue indefinitely (given that conditions in disk support it all the way in or out). Because of the need to take account of complex coorbital flows in a partially gap forming regime close to the planet, the problem of type III migration is very numerically challenging and therefore issues of adequate numerical algorithms, resolution and convergence remain outstanding. We cannot dwell here on the details of such calculations; they are discussed by Peplinski and Artymowicz (2006) and are the subject of active research by others.

The Minimum Mass Solar Nebula (MMSN) is not massive enough to allow superjovian planets to migrate fast. On the other hand, Saturn will migrate fast (in type III mode) in disks a few times the MMSN. This may be related to the unexplained fact that most of the extrasolar planets known as “hot Jupiters”, with a semi-major axis $a < 0.06$ AU, happen to have sub-Jovian masses. The detailed scenarios accounting for the actual masses

seen in radial velocity surveys need to be constructed, the challenge there is to study in detail the mass growth rate of a giant planet, based on internal restructuring and the supply of ample gas.

Migration type III has one big advantage over the standard two of its predecessors: Type I and type II motions are difficult to stop, which we discussed above at some length to show why this issue is considered a central difficulty of the current dynamical theory of planet formation. In contrast, type III motion in a partially open gap (in front or behind the planet) is extremely rapid but very fragile at the same time: any significant gradient of disk density will arrest it, be it due to varying efficiency of accretion (Shakura-Sunayev α), dead zones preventing midplane accretion, photoevaporation of disk parts by external ultraviolet flux, flybys of stars in the Orion-type dense stellar cluster, or other causes. Saving migrating planetary cores may not be difficult after all!

3. Transitional and Beta Pictoris-type disks: The origin of structure.

So many important processes in dusty disks around newly formed stars depend on the radiation pressure exerted by the star on dust and gas that it is worth for us to totally focus here on two main subjects: What drives the evolution of the disks, and why do the new observations by HST and ground-based telescopes show rich non-trivial structure, from blobs and lopsided appearance to pieces of spiral arms not unlike those we study numerically around protoplanets. Figure 10 is a roadmap to physical processes and their basic outcomes in circumstellar disks. Since axisymmetric ring formation has been discussed many times in the past, we shall almost exclusively concentrate on the possibility of gas-dust-radiation coupling to create a non-axisymmetric structure, possibly confusing our efforts to discover planets by their shepherding of dust.

3.1 Interactions of planetary disks with interstellar dust

At the upper-right corner of our diagram there is a spot succinctly stating that radiation pressure repels ISM dust grains, and hence the evolution of the disks is their internal matter, not a result of outside influences on Vega-type objects. Lissauer & Griffith (1989) proposed that sandblasting by ISM dust during rapid passages through atomic clouds depletes circumstellar disks; β Pic which happens to have a small velocity with respect to the clouds would be spared the sandblasting. Whitmire et al. (1992) proposed that all the nearby Vega-type stars have recently passed through molecular cloud and were sandblasted, which generated the finer dust from meteoroids. Artymowicz and Clampin (1997) revisited the issue and found that the ISM neither creates nor destroys the disks around A-type stars, mainly because the ISM grains are repelled by radiation pressure; the 'internal sandblasting' dominates over ISM sandblasting out to radii of order 400 AU, at which ISM may cause some one-sided disk asymmetries. something that we need to keep in mind right now that the imaging is revealing new objects and features.

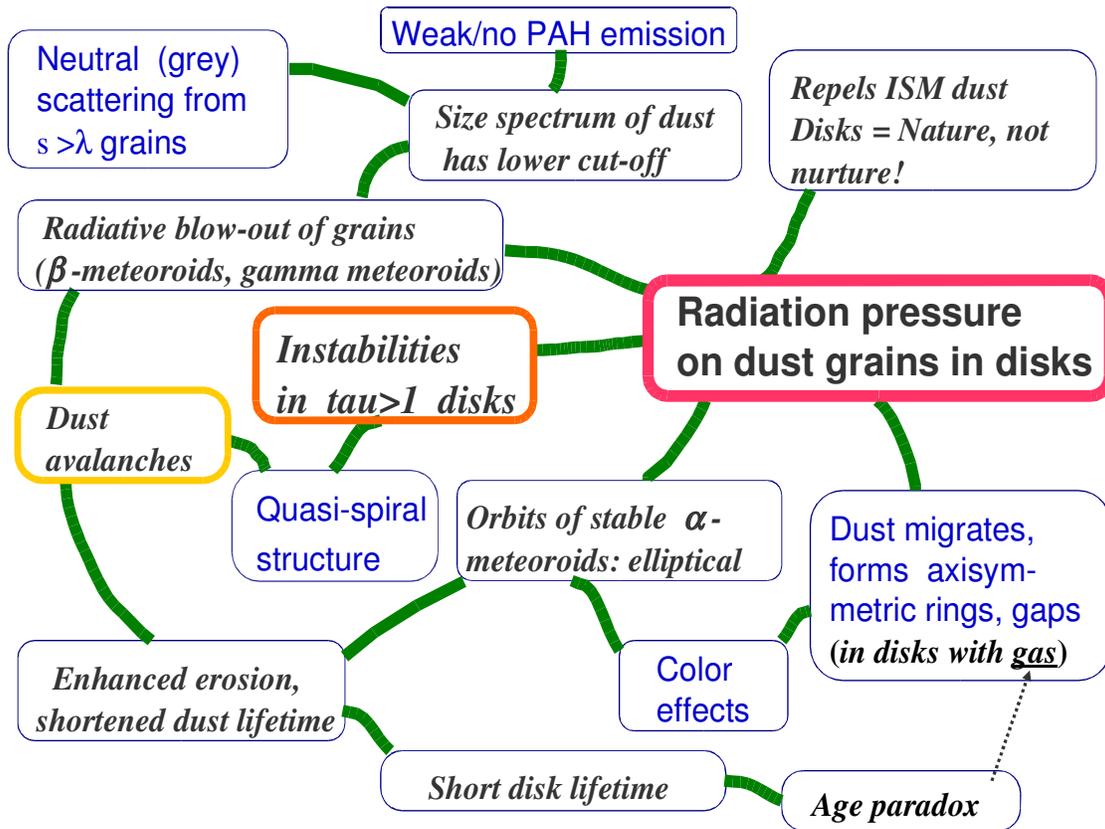


FIGURE 10. The many roles of radiation pressure in young planetary systems. Processes are written in a straight font, directly observable featured are in italics.

3.2 Size spectrum has a lower cut-off or knee

Radiation pressure equaling one-half the stellar gravity ($\beta = 0.5$) suffices for ejecting grains from a parent circular orbit to infinity, unless there is sufficient gas to counteract a rapid escape of small grains. The so-called β -meteoroids, dust below the appropriate radius s_β , for which $\beta = 0.5$, escapes. (Particles on bound orbits are sometimes called α -meteoroids.) Peak values of β much larger than unity are actually found around the known Vega stars, reaching several dozen in the case of Vega itself. In case of β Pic, common materials such as magnesium-rich silicate forsterite ($\text{Mg}_{0.95}\text{Fe}_{0.05}\text{SiO}_3$) and also water ice particles reach a peak of $\beta \sim 2 - 6$ for particle radius $p = 0.1 - 0.2 \mu\text{m}$.

As predicted, in β Pic there is a good observational evidence that particles below $2 \mu\text{m}$ radius are indeed rare (Artymowicz 1997). This is also true of very small Polycyclic Hydrocarbons (PAH particles). The effects of radiation are quite similar in different systems (e.g., Artymowicz and Clampin 1997). Luminosity differences simply modify the blow-out grain radius but the knee is always there. However, the cut-off must be where it is in β Pic or at larger sizes in order for the scattering not to introduce noticeable reddening or blueing with respect to the starlight (particles larger than the wavelength of the visible photons are required). Interestingly, new observations of β Pic by Golimowski

et al. (2006) do show for the first time a somewhat red scattered radiation at large distances from the star (> 300 AU) while a sister disk of AU Microscopii does show an opposite effect of strongly blue scattering (it is 60% brighter at B than I, according to Krist et al. (2005)). Perhaps the difference lies in the lower radiative blowout size of grains around a weaker source of AU Mic. HR 4796A is a narrow disk, and does not have color effects (Clampin et al. 2003).

3.3 Collisional dynamics and avalanches of dust in disks

Estimates show the possibility of the so-called dust avalanches (Artymowicz 1996) in HR4796A and, in fact, any disk with similar dimensions and dustiness ($f_d = 5 \times 10^{-3}$ in HR 4796A). The concept of an avalanche or, in other words, a chain reaction of outflowing debris is illustrated in Fig. 3. Three kinds of particles are distinguished in the figure. In addition to α and β particles, we define the γ -meteoroids as such (very small) grains, whose dynamics is strongly affected by gas drag. Slowed down in their outflow by the drag force, γ -meteoroids do not contribute significantly to dust processing, and we ignore them for a moment (cf., however, the next subsection).

Mutual collisions of stable disk particles ($\alpha + \alpha$) and the bombardment of disk grains by outflowing β -meteoroids ($\alpha + \beta$) contribute to the creation of fresh β -meteoroids, subsequently accelerated by radiation pressure force. On the average, $N + M$ fresh β -meteoroids per collision are produced (N from cratering/erosion and M from catastrophic disruption). Since the probability for a given β particle to strike a disk grain while crossing the disk annulus with optical thickness $d\tau$ is roughly equal to $d\tau$, a toy model of the growth of an avalanche can be constructed by neglecting, among others, the need for a grain to accelerate over a finite radial distance before collisions can become strongly erosive/disruptive. The growth equation for the number n of particles in the avalanche has the form $dn = (N + M)n d\tau$. The solutions of the toy model are thus exponentially growing avalanches of the form $n(\tau) = n_0 \exp[\int_0^\tau (N + M) d\tau] = \exp[\tau(N + M)]$, with n_0 an arbitrary constant. The optical thickness of the disk along its midplane, τ , is approximately its dustiness factor f_d divided by flattening ratio $z/r \sim 0.1$. Therefore, a very large amplification factor of an avalanche of order $\exp 30 \approx 10^{13}$ may be obtained if $(N + M)f_d \sim 30(z/r) \sim 3$.

From the physics of grain collision and more detailed models of a disk at radii $r \sim 100$ AU it follows that $N + M \sim 10^2 - 10^3$ (or $10^{2.5 \pm 0.5}$). Therefore, a plausible *upper limit on the dustiness of gas-free disks* exists in this model: $f_d < 3/(M + N) \sim 10^{-2 \pm 0.5}$. This value is 4 times larger than that in β Pic but only 2 times larger than that in HR 4796A. Hypothetical gas-free disks above that limit would tend to self-destruct on exponentially short time scales unless and until they reduce their dustiness down to the limiting value. The details of the evolution are yet to be determined.

3.4 Inferring the presence of gas (Age paradox)

The dust grinding rate of dust which, in a steady-state collisional cascade, is also the mass loss rate from the largest parent bodies (planetesimals), can be evaluated in detailed modeling of a disk. Dividing the expected mass of the rocks in a typical protoplanetary disk (e.g., $\sim 120M_{\oplus}$ estimated by Artymowicz 1997) by that rate yields the estimated half-life timescale of the whole disk (as opposed to the lifetime of the currently observed dust, which may be a thousand times shorter). Clearly, the disk's half-life should be longer than its age. But this is precisely the trouble with HR 4796A. We obtained a preliminary estimate of 3 Myr disk self-destruction time. HR 4796A and its disk are 8 ± 3 Myr old.

This paradox can be resolved if we relax one or more of the model assumptions. The most natural resolution is that there is an unobserved gas component with total mass at least ~ 4 times exceeding the dust mass, i.e. a fraction of Earth mass. This mass would be much smaller than the upper limit of $7 M_{\oplus}$ of gas, obtained from observations in molecular emission lines. Notice that the easiest method of direct detection of gas, via its absorption lines, may not be feasible in HR 4796A because of the viewing geometry with line of sight inclined by at least $1/3$ of radian, an angle several times larger than the opening angle $z/r \sim 1/10$ of a disk. Theory can thus be a valuable tool for predicting the presence of a hard-to-see gas component.

3.5 Vega-type system classes

The theoretical limit on the area of dust in Vega-type disks should modify the statistics of the observed fractional IR luminosity f_d . Artymowicz (1996) considered the pre-ISO statistics of the Vega stars and obtained a bimodal histogram of f_d . While there were many more examples (per $\log f_d$ bin) of disks with $f_d < 10^{-3}$ and $f_d > 10^{-2}$, the only two systems in the $10^{-3} - 10^{-2}$ bin were β Pic and HR 4796A. This provides a strong support for the division of Vega-type systems into "gas-poor" and "gas-rich" based on a very simple and readily available diagnostic (f_d). It would be very interesting to revisit the observed statistics with a fuller sample of disks and proper account of selection effects, if any⁴. Decin et al. (2003) recently observed a large sample of Vega-type stars with the ISOPHOT detector onboard ISO satellite, and studied the age dependence of the Vega phenomenon. After correcting some poorly known ages of stars, they've found little support for an otherwise popular idea that there should be a law of decline of the amount of orbiting dust in a planetary system seen as a power-law of dustiness vs. age. A power-law lower envelope has been found, a large spread of systems above it, but also a constant demarkation line of dustiness of order that of β Pictoris, a line that normal systems respect and do not exceed corresponding to 10^3 ! According to our theory, this maybe simply a limit enforced by avalanches.

⁴ Notice that there is no obvious selection against discovering systems with, say, $f_d = 0.03$, as opposed to $f = 0.003$ in sky surveys done in the past.

3.6 Gas-related effects in disks

The gas can have many observable consequences. For example in HR 4796A it could slow down the otherwise rapid escape of very small grains (γ -meteoroids). By the virtue of their slowness, γ -meteoroids might contribute significantly to the total area and observability of such fine dust. More generally, most gas-rich systems should be able to retain observable fraction of very small and transiently heated grains (which influences their light reprocessing and spectra) and/or PAH (polycyclic aromatic hydrocarbons). A correlation of PAH/small grain features with f_d is, indeed, present but awaits full analysis and description. Gas will also have a moderating effect on the overall dust processing rate in disks. If gas dynamics is truly important there, the HR 4796A disk may be a very promising laboratory for the study of distinctive nonlinear spiral density waves and resonantly truncated disk edges, hopefully indicating the position and orbit of planet(s). In contrast to this situation, a gas-poor disk like β Pic will not be able to maintain a sharp edge with or without planets. The dust will in general be trapped only temporarily at the outer Lindblad resonances in the circumbinary disk (binary=star+planet). A gradual gap, washed out by large velocity dispersion in disk, might then appear. Sharpness of profiles and disk features is a hallmark of the presence of gas.

Dust can migrate in disks both toward and away from the star, sometimes quite rapidly. The usual direction is inward, for the following reasons. In the solar nebula, there is a partial cancellation of the stellar gravity by the radial gradient of gas pressure, giving rise to a small decrease of orbital speed with respect to the Keplerian value $v_K = \sqrt{GM/r}$; we can denote it by $\eta = \Delta v/v_K \approx 0.005$. This causes an inward migration of solids due to the headwind $w = v_K \eta$ experienced by particles, and the associated loss of angular momentum at a speed proportional to w . The proportionality constant for large particles increases with decreasing size, for they are then better coupled to the gas (have larger area to mass ratio) while still orbiting along weakly perturbed Keplerian ellipses or circles. On the other hand, very small dust grains are coupled so well via strong gas drag force that they are almost frozen into the gas and corotate with it (at a sub-Keplerian speed $v_K - w$). This causes a small, size-independent net radial acceleration (gravity being slightly larger than centrifugal force), balanced by the radial drag component. In this regime, migration speed increases linearly with the particle size because of the linearly decreasing drag force. Large-scale migration of solids in turbulent gas disks was recently modeled by Stepinski and Valageas (1997). The maximum migration speed at intermediate particle sizes, for which the stopping time (velocity divided by gas drag deceleration) is equal to the dynamical time Ω^{-1} . Migration can remove such dust from a solar nebula.

In optically thin disks dust is subject to the combined gas drag and radiation pressure forces, which modifies the migration speed to $\dot{r} \sim (\beta/2 - \eta)v_K$, where $\beta < 1/2$ is the radiation to gravity ratio, and the coefficient of proportionality is a function of grain stopping time. For example, a $25 \mu\text{m}$ radius grain in HR 4796A disk has $\beta \approx 0.2$ and has an average circulation velocity $v_\phi = \sqrt{GM(1 - \beta)/r} \approx (1 - \beta/2)v_K$. The orbital motion of that grain is slower than that of the gas, thus the particle feels the push of a backwind, gains angular momentum, and spirals outwards. In some cases outflow speed may be independent of particle size (over a certain size range), because the dust mobility of

small (well-coupled) grains increases, while the radiation pressure drops with the grain size (Takeuchi and Artymowicz 2001). Poynting-Robertson drag can be included in the analysis but it is never dominant except in very low-density disks, e.g., in our Solar System.

3.7. HD 141569: Planets or dust mimicry?

HD 141569, a system shown in the left-hand side of Fig.2 is a quiescent Herbig-type star (type B9.5Ve) surrounded by a light-scattering (and IR emitting) disk (inclination 40 degrees away from edge-on). HD 141569 has a double-peaked H α emission indicating a rotating gas disk close to the star, as well as gaseous CO detected at radius of order 90 AU. Details of its structure were imaged for the first time with NICMOS/HST at 1.1 μ m (Weinberger et al. 1999) and at 1.6 μ m (Augereau et al. 1999). The crossed dark radial stripes are artifacts of the observation method, but the division of the disk into an inner main part and an outer ring is intrinsic to the object, although over-exaggerated by grayscale map. Analysis of the 1.1 μ m observations yielded the radial profile of the vertical optical thickness $\tau(r)$ times the unknown albedo, which we reproduce in Fig. 4. The profile has a moderate dip at $r = 250$ AU, interpreted by Weinberger et al. as a sign of a planet residing in the disk at that radius. If this feature is indeed due to a planet, its eccentricity must be small (which requires formation *in situ*), and a moderate mass (on the order of Neptune's mass; this follows from gap opening criteria). An obvious difficulty, however, arises in the formation theory because of the very large distance from the star. In the Solar System, the Kuiper belt region is 5 times closer and yet had insufficient mass (surface density) to form planets. Time needed to assemble planets at 250 AU also strongly disfavors *in situ* formation. Outward migration to this large radius remains an unlikely possibility. In fact, the gap is not deep enough to stop a predicted rapid inward migration type I (Ward 1997).

In this and similar cases one should avoid postulating planets indiscriminately to account for every morphological feature of the disk, unless supported by independent evidence. In HD 141569 other plausible explanations of the disk morphology exist, which naturally explain the puzzling differences in the appearance of the disk at the two wavelengths observed (Augereau et al. 1999 do not detect at 1.6 μ m the disk part inside $r = 180$ AU, prominent at 1.1 μ m).

If the disk is nearly optically thick in radial direction, shadowing of outer disk parts by inner disk parts in conjunction with a variable geometrical disk thickness might play a role, especially in the scattered light images like the NICMOS ones. At present we have no direct information on either vertical geometry or radial optical depth in this disk. Secondly, a variety of general scenarios can be constructed in which dust either originates or accumulates in a certain place in a disk, depending on dust grain size. We consider the dust migration under the combined gas drag and radiation pressure forces (Takeuchi and Artymowicz 2001) a much simpler explanation overall. If the disk is nonaxisymmetric, as in the case of HD 141569A (cf. Ardila et al. 2005), a nonaxisymmetric initial distribution of dust, combined with a method of multiplying the visible area of dust, might be of help.

Grigorieva et al. (2006) produced for the first time a comprehensive model of dust avalanche traced from the beginning in a presumed planetesimal-planetesimal collision, through the time of escape and exponential growth, to the eventual decline. Collisional avalanches are expected to be triggered by a localized disruptive event, such as the collisional breakup of a large cometary or planetesimal-like object. A fraction of the dust then produced is driven out by radiation pressure on highly eccentric or even unbound orbits. These grains moving away from the star with significant radial velocities can break-up or microcrater other particles further out in the disk, creating in turn even more small particles propagating outwards and colliding with other grains. Should this collisional chain reaction be efficient enough, then a significant multiplication in the number of dust grains could be achieved. In this case, the consequences of a single shattering event in terms of induced dust production could strongly exceed that of the sole initially released dust population, possibly too small to be detected in the absence of avalanches. The outwards propagation of the dusty grains might induce asymmetric features in the disk.

The novelty in this calculation is the full treatment of size distribution and mutual collisions. A concept of superparticles is used, these are entities following the same trajectory in space and containing a large number of similarly sized particles. During collisions, superparticles proliferate beyond the capability of a program to trace all of them, therefore regular pruning or re-sorting of superparticles done so as to preserve integrals of motion, is done to keep the total number of superparticles to about 10^5 to 10^6 . Amplification factors of 10^3 are routinely achieved in the system just a few times more dusty than β Pic. This would guarantee visibility of an avalanche coming from a collision of two large planetesimals. As to β Pic itself, however, typical parameters regarding the materials strength of dust grains lead to only small amplifications by a factor of 10 to 100, and so for visibility against the backdrop of a large disk would require initial cloud of dust that could only be created in a protoplanet-protoplanet collision, which makes this mechanism very unlikely.

3.8. Instabilities in $\tau > 1$ disks

Finally, disks made of dust and gas can under the right conditions develop their own spiral wavelets and even spectacular spiral arms somewhat like the spiral galactic arms. Radiation pressure acting on an overdense region in a disk (which could be a spiral arm, for instance) has the ability to compress it, if the disk has noticeable optical thickness. Then the front of the region is being pushed the most, and the back of the region is in a partial shadow of its own making. Relative to the center of the region, radiation causes a compression, which acts in complete formal analogy to self-gravity. Therefore, a criterion can be developed along the lines of the derivation of the Safronov-Toomre gravitational instability of a disk. This model has an analytical linear solution (Artymowicz, 2006), however, it must now be studied numerically in a fully nonlinear regime. In conclusion, there are several good explanations for nonaxisymmetric features in dusty circumstellar disks in addition to the perhaps most tantalizing one, which involves hypothetical exoplanets. All of them will need to be scrutinized before it will

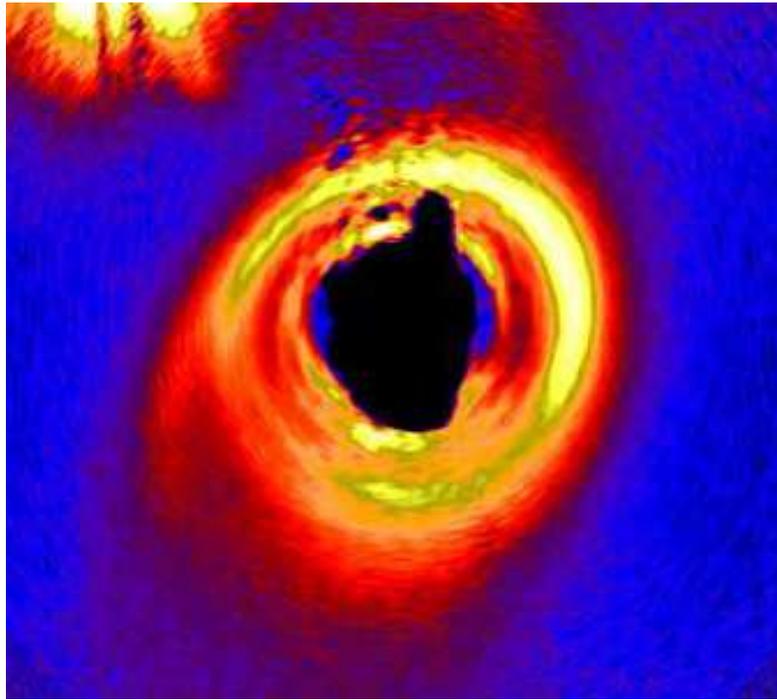


FIGURE 11. HD141569A seen in the near-IR light by HST (Clampin et al. 2003)

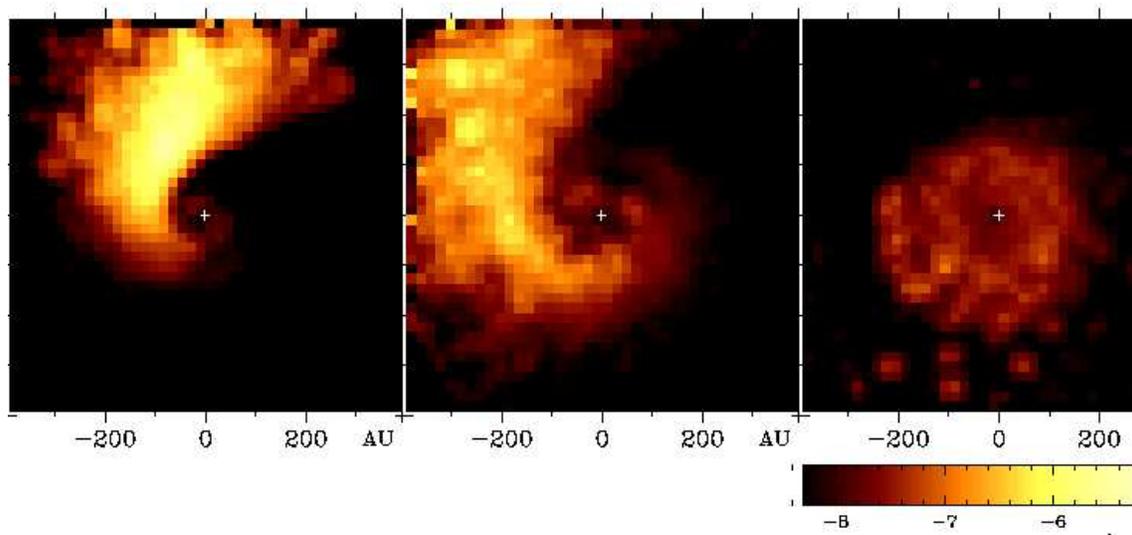


FIGURE 12. Three snapshots of a developing dust avalanche, growing in a β Pictoris-like disk. Disk particles are not shown. Brightness scale is logarithmic. (Grigorieva et al. 2006).

become clear which process is responsible for which particular morphology of a disk.

4. Conclusions

At the close of the previous millenium, we were fortunate to discover numerous planetary systems. Extrasolar planetary systems are fairly common, accompanying 10 to 20% of all stars. These numbers will further climb in the future, as observational capability to see both planets and their accompanying disks increases. Uncharted terrain of super-Earths and Neptunes lies ahead in the radial velocity surveys. Giant planets, of which we seem to know already 188, are not arranged about their suns as ours are around the sun - and we need to know why.

Currently the most serious challenge in observations is to discover and study Earth-like planets, to understand how prevalent potentially life-supporting planets are. In theory of formation, there are likewise broad challenges: to understand how and why protoplanets migrate but do not die en masse during their extreme youth. Migration type III is but the newest item in our arsenal to confront the riddles of nature. It is a striking mechanism, which only very recently gained the support of the majority of dynamicists. Despite being so quick and so brief that we won't probably actually see it occurring, rapid migration has an interesting future.

Finally, the study of the origin of nonaxisymmetric structures in disks will be time well spent, if we are going in the future to go hunting for planets in the shifting sands of dusty disks. If we know enough, we might even succeed!

Acknowledgments

Contributions from many co-authors of the cited papers are incorporated in this overview. This work was supported by the NSERC research grant 5810-2005-312219. The author acknowledges the angelic patience and help from the editors of this volume. Some of the results reported here were obtained with a PPM hydrocode SVH partly based on the VH-1 program, generously made public by J. Blondin.

REFERENCES

- . Ardila, D. R., Lubow, S. H., Golimowski, D. A., Krist J. E. et al. 2005, ApJ, 627, 986
- . Artymowicz P. 1993a, ApJ, 419, 155
- . Artymowicz, P. 1993b, ApJ, 419, 166
- . Artymowicz P. 2004, Astr. Soc. Pacific series, vol. 324, 39
- . Artymowicz P. 1996, in Role of Dust in the Formation of Stars. Eds. HU Käuffl & R Siebenmorgen (Berlin: Springer), 137
- . Artymowicz, P. 1997, Ann. Rev. Earth Pl. Sci., 25, 175
- . Artymowicz, P., Clampin, M. 1997. ApJ, 490, 863
- . Artymowicz P., and Peplinski, A. 2006, in preparation
- . Augereau, J. C., Lagrange, A. M., Mouillet, D., Ménard, F. 1999, A&A, 350, L51
- . Augereau, J. C.; Lagrange, A. M.; Mouillet, D.; Papaloizou, J. C. B.; Grorod, P. A. 1999, A&A, 348, 557
- . Aumann, H. H., Gillett, F. C., Beichman, C. A. et al. 1984, ApJ, 278, L23
- . Backman D. E., Paresce F., eds. 1993. In Levy et al. (1993), p. 1253
- . Boss A. 2000, ApJ, 536, L101
- . Bryden G., Chen X., Lin D. N. C., Nelson R. P., and Papaloizou J. C. B. 1999, ApJ, 514, 344

- . Charbonneau, D., Brown, T., Noyes R., Gilliland R. 2001, *ApJ*, 568, 377
- . Charbonneau, D., Brown, T., Burrows, A., Laughlin, G. 2006, in *Protostars and Planets V* (Reipurth et al. 2006)
- . Clampin, M., Krist, J. E., Golimowski, D. A., Ardila, D. R. et al. 2003, *AJ*, 126, 385
- . Crida A., Morbidelli A., and Masset F. S. 2006, *Icarus*, in press; astro-ph/0511082
- . D'Angelo G., Bate M., and Lubow S. 2005, *MNRAS*, 358, 316
- . D'Angelo G., Henning T., and Kley W. 2003a, *ApJ*, 599, 548
- . D'Angelo G., Kley W., and Henning T. 2003b, *ApJ*, 586, 540
- . Decin, G., Dominik, C., Waters, L.B.F.M., Waelkens, C. 2003, 598, 636
- . de Val Borro M., Edgar, R., Artymowicz, P., Ciecielag, P., Cresswell, P., et al. 2006, *MNRAS*, in print; cf. also <http://www.astro.su.se/groups/planets/comparison/index.htm>
- . Dominik, C., Laureijs, R. J.; Jourdain de Muizon, M.; Habing, H. J. 1998, *A&A*, 329, L53
- . Goldreich P., and Tremaine S. 1979, *ApJ*, 233, 857
- . Goldreich, P., and Tremaine, S. 1980, *ApJ*, 241, 425
- . Golimowski, D.,A., Ardila, D. R., Clampin, M., Krist, J. E., et al. 2006, *ApJ*, in print
- . Grigorieva, A., Artymowicz, P., and Thebault, P. 2006, *A&A*, in print
- . Heap, S. R., Linder, D. J., Lanz. T. M., Woodgate, B., Cornett, R., Hubeny, I., and Maran, S. P. 2000, *ApJ*, 539, 435
- . Jang-Condell H., Sasselov D. D. 2005, *ApJ*, 619, 1123
- . Krist, J. E., Ardila, D. R., Golimowski, D. A.; Clampin, M. et al. 2005, *AJ*, 129, 1008
- . Lagrange A M., Backman, D., Artymowicz, P. 2000, in Mannings et al. (2000), p. 639
- . Laughlin, G., Bodenheimer, P. 1994, *ApJ*, 436, 335
- . Lin D. N. C. and Papaloizou J. C. B. 1986, *ApJ*, 309, 846
- . Lin D. N. C. and Papaloizou J. C. B. 1993, in *Protostars and Planets III*, E. H. Levy and J. I. Lunine (Eds.), Univ. of Arizona, Tucson, p.749
- . Lin D. N. C., Papaloizou J. C. B., Terquem C., Bryden G., and Ida S. 2000, in Mannings et al. (2000), p. 1111
- . ubow S. H., Seibert M., and Artymowicz P. 1999, *ApJ*, 526, 1001
- . Lissauer J.J. 1993. *Ann. Rev. A&A*, 31, 129
- . Lissauer J.J., Griffith CA. 1989, *ApJ*, 340, 468
- . Lubow, S. H., Seibert, M., Artymowicz, P. 1999, *ApJ*, 526, 1001
- . Mannings, V., Boss A., Russell, S. (Eds.) 2000, *Protostars and Planets IV*, Tuscon: Univ. Arizona Press
- . Marcy G. W., Butler R. P. 1998, *Ann. Rev. A&A*, 36, 57
- . Mayor M., Queloz D. 1995. *Nature*, 378, 355
- . Menou K. and Goodman J. 2004, *ApJ*, 606, 520
- . Nelson A. F. and Benz W. 2003, *ApJ*, 589, 578
- . Nelson R., and Papaloizou J.C.B. 2004, *MNRAS*, 350, 849
- . Papaloizou J. C. B. and Larwood J. D. 2000, *MNRAS*, 315, 823
- . Papaloizou J.C.B., Nelson, R. P., Kley, W., Masset, F. S., and Artymowicz, P. 2006, in *Protostars and Planets V* (Reipurth et al. 2006); astro-ph/0603196
- . Papaloizou J.C.B., and Terquem, C. 2006, *Rept. Prog. Phys.* 69, 119
- . Peplinski, A., and Artymowicz, P. 2006, in prep.
- . Pierens A. and Huré J. M. 2005, *A&A*, 433, L37
- . Rafikov R. 2005, *ApJ*, 621, L69
- . Reipurth, Jewitt B. D., and Keil K. (Eds.) 2006, *Protostars and Planets V*, University of Arizona Press, Tucson; in print
- . Schneider, G., Becklin, E., Smith, B. A., Weinberger A., et al. 2000, *AJ*, 121, 525
- . Schneider, J., 2006, *Extrasolar Planet Encyclopaedia on-line database* at <http://exoplanet.eu>
- . Smith B. A., Terrile R. J. 1984, *Science*, 226, 1421
- . Stepinski, T., Valageas, P., 1997. *A&A*, 319, 1007
- . Tanaka H., Takeuchi T., and Ward W. R. 2002, *ApJ*, 565, 1257
- . Takeuchi T., and Artymowicz, P. 2001, *ApJ*, 557, 990
- . Terquem C. 2003, *MNRAS*, 341, 1157
- . Trilling, D., Brown, R. H. 1998, *Nature*, 395, 775
- . Vidar-Madjar, A., Lecavelier des Etangs, A. and Ferlet, R. 1998, *Pl. Sp. Sci.*, 46, 629
- . Ward, W. R. 1986, *Icarus*, 67, 164

- . Ward, W. R. 1997, ApJ, 482, 211
- . Weinberger A. J. et al. 1999, ApJ, 525, L53
- . Whitmire D.P., Matese J.J., Whitman P.G. 1992, ApJ, 388, 190
- . Wolszczan, A., Frail D. A. 1992, Nature, 355, 145