Growth and interaction of planets

Pawel Artymowicz

Stockholm Observatory, SE-13336 Saltsjöbaden, Sweden,
pawel@astro.su.se

Abstract. We discuss theories of origin and evolution of the newly discovered extrasolar planetary systems. As these systems failed to fulfill prior expectations concerning their orbital structure, we are challenged to extend and/or revise many preexisting theories. Important extensions include migration of bodies in disks and planetary eccentricity pumping by planet-planet interaction and primordial disk-planet interaction. Progress in observational techniques will allow us to find which of these two types of interaction is responsible for the observed variety of orbits and masses of planets. New insights into the formation of giant planets in our system can be obtained by asking why Jupiter and Saturn are not larger, closer to the sun and/or do not follow noticeably elliptic orbits.

1. Prior expectations

The extrasolar planetary systems have been anticipated long before they were discovered. Greek atomism of Leucippus and Democritus (cf. Dick 1982) lead to prediction of other "worlds" (planetary systems with their moons and suns) and, importantly, their evolution (formation in a rotating, turbulent nebula; fractionation of dust), and diversity (worlds with no moons, multiple suns etc.) The oldest and the newest concepts regarding "other worlds" are thus remarkably similar.

The "standard model" of solar system formation from 1980s served by default (or Copernican principle) for extrasolar systems as well. 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 (or only rock) containing bodies, form from dust, accumulate in orbit via binary collisions, and in less than 1 Myr form protoplanets. In <100 Myr terrestrial planets are assembled in the inner solar system. 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 formation scenario most supported by ground-based observations and spacecraft flybys, called "core-accretion" or "core-instability" scenario, the primitive atmosphere becomes unstable and accretes onto the core without mixing, when the core mass exceeds 7 and 10 Earth masses, a value in line with core mass estimates from Jupiter to Neptune (see reviews in Mannings et al. 2000).

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 $\alpha \sim 10^{-2}$.

2. The new worlds

2.1. Dusty disks

The standard model of planet formation agrees with numerous observations of circumstellar disks. Many dusty disks known as Vega-excess systems (because of infrared radiation detected by satellite IRAS around Vega, much exceeding the flux from the star itself), are good if not exact analogues of an early solar system. An A-type star β Pictoris was first imaged to reveal an extended, edge-on dust disk. This system, and many others thereafter, were recognized as truly planetary systems, grinding planetesimals and meteoroids to dust (Artymowicz 1997). The name 'replenished disks' describes their nature well. The observed amount of dust in β Pic, which could tightly cover the orbit of Uranus and weighs as much as several Moons, must have been resupplied thousands of times during the star's lifetime (20-100 Myr), partly owing to erosion-enhancing dynamical effects of radiation pressure on dust. There is evidence that protoplanet-sized or larger bodies hide in the disk. The strongest ones include: the warp in the inner 100 AU zone of the disk (Heap 2000), the need to perturb comets/asteroids from the disk to the immediate vicinity of the star, where they are seen spectroscopically as gas-and-dust shedding bodies. (For review see Lagrange et al. 2000).

Several transitional disks, of approximately 5-10 Myr of age, have recently been imaged (Koerner et al. 1998, Weinberger et al. 1999). Situated evolutionary between gas-rich solar nebulae and the gas-poor replenished dust disks, these disks (e.g., in HR 4796A and HD 141569) show evidence of features such as gaps and inner clearings, which may be due to planets or, alternatively, dust migration in optically thin disks with gas (Takeuchi & Artymowicz 2001).

2.2. Exoplanets

The first extrasolar planetary system was found in an unlikely place: around a millisecond pulsar PSR1257+12 (Wolszczan and Frail 1992). Belief in the ubiquity of planets was strengthened and searches might have been stimulated by its discovery but, ironically, the orbital structure was too similar to that of the inner solar system to prepre us for the other worlds around normal stars. Likewise, a strong expectation (Boss 1995) that the solar and extrasolar systems share the basic blueprint (giants outside a few-AU radius, terrestrial planets inside) provided no guidance to the discovery of the extrasolar giant planets.

Mayor and Queloz (1995) discovered the first 4-day period giant planet at a distance of only 0.05 AU from 51 Pegasi. Currently, more than 55 exoplanet candidates are known from radial velocity studies (cf. Butler et al. 2001; Schneider 2001, Marcy et al. 2001). Statistical conclusions are possible at this stage (although serious modeling of the observational biases is acutely needed for confirmation of finer points):

• Planetary companions (with $a < 3$ AU and at least Saturn's mass) exist around at least $\sim 7\%$ of normal stars.
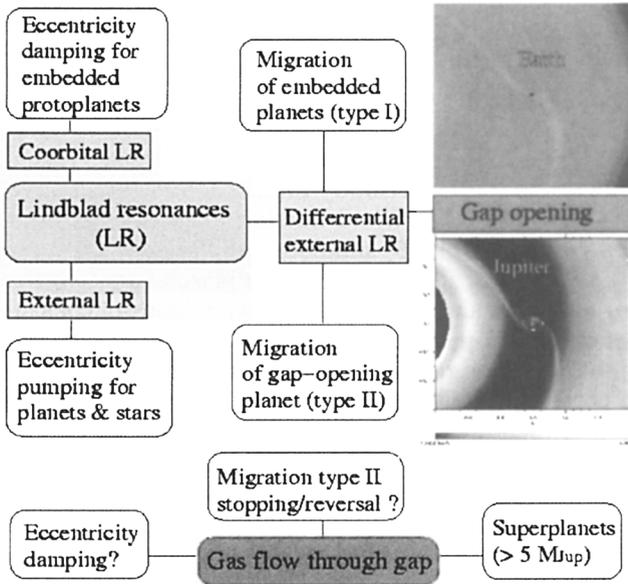
Figure 1. Theoretical concepts and effects of disk-planet interaction.

• "hot Jupiters" (minimum masses $m > 0.1 m_J$, semi-major axes $a < 0.1$ AU) exist around 0.7% of sun-like stars.

• Eccentric planets abound. There is a clear e–P correlation (Marcy et al. 2001) resembling closely that of the PMS and main-sequence close binary stars (cf. Mathieu 1994). Orbits with $a < 0.1$ AU tend to have $e \approx 0$. There is a weak direct correlation of e with $m \sin i$.

• 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 $m_i = m \sin i > 10 m_J$ (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 details of where this happens are unknown because of poor statistics.

3. Theories

3.1. Planet-planet interaction and other perturbations

Eccentricity-inducing star-planet interactions take place during flyby's frequently occurring in dense open clusters (Laughlin & Adams 1999), especially if the

---

[1] The unknown inclination of the orbit i does not allow unique mass determination, and yields only the minimum mass $m \sin i$ of a system. This mass, assuming randomly oriented orbits, is on the average $\langle 1/\sin i \rangle = \pi/2$ times smaller than the true mass.

planet is far away from its host star ($a > 10$ AU). On the other hand, such a perturbation cannot explain the eccentricities of short-period (hot) planets. Of course, planets in known binary systems, such as 16 Cyg B, can also be strongly affected (Holman & Wiegert 1999). Eccentricity and orbital inclination can undergo anticorrelated, large swings known as the Kozai effect (e.g., Lin et al. 2000). Strong mutual gravitational perturbations between the forming planets distort their orbits up to a high $e \approx 0.9$ or cause escapes (e.g., Weidenschilling & Marzari 1996). But the models of N-body interactions depend strongly on the initial configuration of bodies. To avoid making arbitrary assumptions about the initial state, Levison et al. (1998) simulated numerically the bottom-top accumulation of a swarm of protoplanets in the outer solar system, including the gas accretion from disk in a simple parametric way (non-selfconsistent). They presented statistics of the computed masses and eccentricities (migration in disks was not modeled), with a wide range of eccentricities including high ones, without a pronounced correlation with planet's mass. A slight anti-correlation may even exist, due to a tendency of any N-body systems toward energy equipartition (here, of the epicyclic motion associated with elliptic orbits). Thus, at least in a rough way, standard accumulation scenario may explain the statistics of eccentric exoplanets but, surprisingly, requires an extra mechanism for damping e (artificially included in a subset of Levison et al. calculations) in order to routinely produce low-eccentricity, solar-like systems. We feel that the best candidate for such a mechanism is the disk-planet interaction.

3.2.   Disk-planet interaction

There are two key processes underlying much of the orbital evolution in disks. The first one is Lindblad Resonance (LR), where a planet launches a spiral wave in disk. The second unifying concept is mass flow through gaps, which will be elaborated in sect. 3.6.

Lindblad resonances and their manifestations are shown in Fig. 1. Lower (upper) part of the diagram describes the behavior of protoplanets which do (do not) open gaps. The condition of LR requires that the (disk) fluid element moves periodically through the rigidly rotating pattern of the perturbing potential with a period equal to its natural radial oscillation frequency, called epicyclic frequency ($\approx$Keplerian angular frequency in a typical solar nebula). If the perturbing body's orbit has $e > 0$, a double Fourier decomposition of potential is carried out and the response of the disk is obtained by linear sum over harmonics (a theory valid for small protoplanets). Pioneered by Goldreich & Tremaine (1979, 1980), the LR theory was generalized by Artymowicz (1993) to handle point-like perturbers embedded in disks, and to derive the so-called torque cutoff at high azimuthal wave numbers of the potential.

3.3.   Gap opening

We mentioned in sect. 2.2 that in the standard theory gap opening was held responsible for limiting the mass of a giant planet. Indeed, the viscous gap opening criterion (Lin & Papaloizou 1979, one of two criteria of Lin & Papaloizou
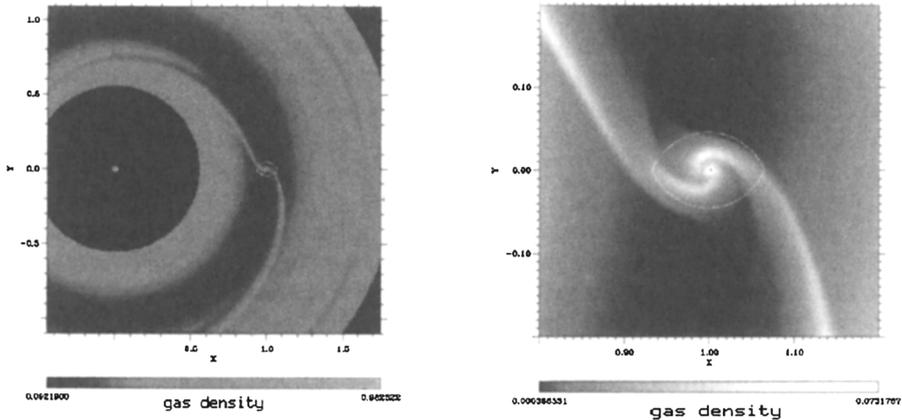
Figure 2. Left panel: a 25-Earth mass protoplanet opens a gap in the surrounding solar nebula (viscosity parameter $\alpha = 0.006$, thickness ratio $z/r = 0.05$). Gas density, in arbitrary units, is color-coded. Right panel: the same as left panel, but a 1 Jupiter-mass protoplanet (and a zoomed view). Notice the shock waves (wakes) between the disk and the Roche lobe of a planet. White ovals around planets mark the Roche lobe of Jupiter.

1993), $\mu \approx 40Re$, is satisfied by a 0.4–1 m_J planet residing in a standard solar nebula with Reynolds number[2] $Re \approx 10^5$.

One problem is that the standard gap criterion disagrees with some numerical simulations (Artymowicz 2000). Figure 2 (left panel) shows the offending Neptune-class planet that exerts gap-opening torques in the surrounding disk simulated by PPM (Piecewise Parabolic Method). The density ratio inside/outside the gap is of order 1:10. Unexpectedly, the standard viscous criterion predicts a 2.5 times larger mass is needed for this, and an additional "thermal" one a 5 times larger mass (i.e., standard criteria require ~100 Earth, i.e. Saturn mass, for gap opening in the simulated disk). More work on gap-opening criteria is clearly needed to find when gaps become observable and how far the planetary cores migrate within the disk, the question to which we now turn.

3.4. Distances: the puzzles of hot and warm Jupiters

Since it is difficult to imagine sufficient amounts of both the rocky material (needed for cores in core-accretion scenario) and the nebular gas for in-situ giant planet formation within a small fraction of AU from the star, it is generally thought that the hot Jupiters, or at the very least their solid cores formed much further out in the disk and migrated inwards (Lin et al. 1996). We distinguish

---

[2]Reynolds number can be estimated via the spectroscopically observed accretion rates of gas onto the PMS stars (this gives viscosity parameter $\alpha \sim 10^{-2}$, cf. review by Calvet et al. in Mannings et al. 2000; we have $Re^{-1} = \alpha(z/r)^2$, where $z/r \approx 0.05$ is the disk opening ratio.

between migration type I, in which the protoplanet is embedded in disk gas, and migration type II, where the planet is centered in the disk gap.

Armed with semi-analytical formulæ describing the LR torques (Artymowicz 1993), Ward (1997) presented a unified theory of migration type I and II. It predicts frighteningly large migration rates, corresponding to time scales for migration from 5 AU of only 10^4–10^5 yr. Observationally, for a while it seemed that planets migrate but preferentially stop close to the stars surviving as "hot Jupiters" (Lin et al. 1996). Since then, however, so many "warm Jupiters" were discovered, smoothly distributed over all observable distances (<3 AU) that the existence of any strong stopping effect near the star is unclear.

There are some empirical indications that solid material migrated long distance in the solar system, despite the fact that its final structure which does not suggest it.[3] An unexpectedly uniform abundance relative to solar of heavy elements including such volatiles as N_2, Ar, and Kr, was found by Galileo probe in Jupiter's atmosphere (Owen et al. 1999). However, if Jupiter formed from material that condensed at its present location then gases like Ar and N_2 should be depleted by 4 orders of magnitude, because they do not solidify or get trapped in clathrates at the temperature characteristic of Jupiter zone (Owen and Bar-Nun 1995). Planetesimals which were incorporated into Jupiter, or perhaps the whole core of Jupiter, appear to have been assembled outside the orbit of Neptune.

3.5. Eccentricities: the puzzle of elongated orbits

The υ And planets exhibit an unnatural (anti-equipartition) dependence of e on m, since the most massive superplanet ($\sim 10 \ m_J$) has by far the largest mean eccentricity. This could indicate the action of a disk during the formation period, or alternatively a contrived N-body interaction resulting in the loss of yet another, larger hypothetical superplanet.

Gravity of the disturbed disk(s) transfers energy and exerts feedback torques on the perturber (planet), thus causing a coupled orbital evolution of a and e in accordance with LR theory (Goldreich & Tremaine 1980, Artymowicz et al. 1991; Artymowicz 1993; see also Fig. 1). Therefore, significant migration necessarily implies significant eccentricity pumping or dumping. Artymowicz (1992) proposed that a sufficiently wide disk gap leads to eccentricity pumping by external Lindblad resonances in disk, and damping otherwise. The mass flowing from the disk onto a planet through a disk gap or from one side of the disk to the other provides an extra torque and energy transfer route, but these effects are more model-dependent (hence question marks following the outcomes presented in Fig. 1).

Eccentricity evolution is independent of the origin of affected bodies. While it is true that stars (brown dwarfs) and radial-velocity companions have similar eccentricity distributions, this does not at all mean that the former are small brown dwarfs, only that both types of objects may have their e pumped by the ubiquitous protostellar disks (for a diverging opinion see Black 1997). The crossover mass for eccentricity damping/excitation depends on disk parameters

---

[3]For instance, migration of a giant protoplanet through the inner solar system would have likely irreplaceably destroyed proto-terrestrial planets.

in a still poorly studied way. We estimate that the most important disk parameter is its Reynolds number, which varies by 2 orders of magnitude among the observed protostellar disks. This should result in exoplanetary systems exhibiting a transition from low to high e over a range of $m \sin i$ from 1 to 10 (or 20) Jupiters. Current searches do not yet yield reliable statistics of sub-Saturnian companions, whose eccentricities should be low if disk-planet interaction is of major and lasting importance in planet formation. It is also not yet known theoretically how large e might be generated (Papaloizou et al. 2001 obtain $e \sim 0.2 - 0.3$ in their models).

3.6.   Masses: the puzzle of superplanets

Superplanets (a loose term for companions roughly 5 to 20 m_J) might, in principle, be small brown dwarfs (forming directly from collapsing molecular cloud) rather than planets (understood to be objects grown in a two-stage accretion process in protoplanetary disks). Why not adopt a definition of a brown dwarf based on mass ($>$13 times Jupiter's perhaps) and just call massive superplanets brown dwarfs? One reason *not* to do it is that naming a body brown dwarf may falsely suggest we know how it formed. A growing number of systems (like v And, HD 168443, Gl 876) will surely be found, in which companions have very different masses and would have to be called stars/planets despite the apparently common origin.

   The reasoning behind an idea of a disk viscosity-based mass limit for Jupiter (Lin & Papaloizou 1993) included an assumption that the gap is empty and impermeable. We have seen above that gaps are cleared somewhat earlier than previously expected. That would lead to a final planetary mass *smaller* than computed from the standard viscous criterion, and thus to problems with the explanation of massive exoplanets. However, gap impermeability has been questioned by Artymowicz & Lubow (1996). Lubow and Artymowicz (2000) summarized the remarkable permeability of almost-empty gaps around binary stars to gas flows from circumbinary disks. Non-axisymmetric flows (and accompanying wakes around the planet growing in a disk, seen in Fig. 2) are a very robust phenomenon and are *not* restricted to eccentric binary systems or stellar mass ratio systems. In fact, a Jupiter in a standard minimum mass solar nebula (with viscosity $\alpha > 10^{-3}$) would double in mass in less than 1 Myr (Artymowicz et al. 1998; fig. 1 of Lin et al. 2000). Results of a number of different hydrodynamical codes, from SPH to ZEUS-type codes (Lubow et al. 1999, Bryden et al. 1999, Kley 1999) and PPM (fig. 2), are supporting a general conclusion: a protoplanetary disk must be extremely non-viscous ($\alpha < 10^{-4}$) to prevent Jupiter from growing further.

   That, however, clashes with empirical requirements based on accretion rates (footnote 2 above). A typical disk with $\alpha \sim 10^{-2}$ and lifetime $>$1 Myr can form a 10 m_J (super)giant planet. Thus, either we do not understand correctly the accretion rates in T Tau disks, or for some reason our calculation do not capture the physics or timing of exoplanet formation (most of which are not superplanets). This is a basic unsolved problem.

4.   Back to Jupiter

We sketched some old and new theories accounting for the amazing variety of planets in nature. Some mechanisms, such as migration and eccentric instability of orbits, constituted an old (pre-1995) *prediction*. While new ingenious theories help us interpret exoplanets, we must not forget about planets close to home. The ice boundary lost its predictive power for the location of Jupiter, because of (i) smaller than previously thought jump in surface density of solids across the boundary, evidenced by a rather modest 25% to 50% mass percentage of ice in comets (previously thought of as only slightly dirty ice), and (ii) effectiveness of migration across any predefined boundary. The concept of viscosity-dependent mass limit was replaced by a more uncertain and much higher estimate of achievable planet mass. Finally, disks were intuitively thought of being able to circularize planetary orbits because of their "dissipative nature", a wrong idea since planets have large mass/area ratio and are oblivious to gas drag. This idea died (or should have died) when it was realized that eccentricities can easily result from gravitational interactions of many sorts. We got a taste of knowledge (about the reasons for exoplenetary diversity) but lost a paradise (some paradigms about the solar system formation).

This loss is not a serious problem if we accept that planetary systems like our own, while abundant (up to 10^{20} may exist in the Universe, 10^6 new ones born every hour), are nonetheless not "typical". A natural explanation of properties of Jupiter, Saturn, Uranus and Neptune is provided by the late timing of their formation with respect to the dissipation of the primordial solar nebula. For instance, early photoevaporation of the solar nebula by nearby massive stars might provide the explanation of why these planets, if they formed in the sequence mentioned above, captured decreasing amounts of nebular gas.

5.   Core accretion: still the best bet?

A percieved problem of long formation timescales for giant planets in the standard core-accretion scenario has prompted a second look at an older idea of rapid giant gaseous planet formation by disk fragmentation (Boss 2001 and references therein). However, the long timescale problem may not exist (e.g., Lissauer 2001), and would be curable anyhow by adopting a sufficient mass of the disk (smaller than required for fragmentation). The disk fragmentation model, in turn, has its own problems. Notice, for instance, that existing models typically begin with disk configurations which are known in advance(!) to be *unstable* to axisymmetric perturbations, but lack convincing proof of why and how the disk would find itself in such precarious state. We know that the approach to an unstable configuration would have to occur rapidly to be successful, on orbital time scale. Otherwise, disks are known to defend themselves from a slowly approaching Safronov-Toomre gravitational instability, e.g., when their mass is being augmented slowly. Like the observed galactic disks and correctly modeled protostellar disks, by launching mass-transferring open spiral waves they tend to stay close to but on the safe side of the instability (Laughlin & Bodenheimer 1994, Nelson 2000).

6. Concluding remarks

Exciting discoveries have been made revealing the existence and diversity of planetary systems around stars, from dusty disks made of planet-building rocks crushed to sand and dust, to the giant and supergiant planets causing a detectable stellar wobble. It is still too early to conclude how exoplanets are normally forming, or even what kind of systems constitute the norm. Some of our prior theories of how Jupiter formed became questionable as an indirect effect of the discovery of exoplanets. This turmoil in the world of theory will be resolved in due time, and contribute to a better understanding of both solar and extrasolar systems. A surmised unified, future theory will likely rest upon the familiar fundament of the accumulation of solid planetesimals and planets in a solar-type nebula (accretion disk), a model which right now finds beautiful confirmation in the observations of the evolutionary descendants of such nebulae, the transitional and replenished dust disks around a significant fraction of normal stars.

References

Artymowicz, P., Clarke, C., Lubow, S.H., & Pringle, J.E. 1991, ApJ, 370, L35
Artymowicz, P. 1992, PASP, 104, 769
Artymowicz, P. 1993, ApJ, 419, 155
Artymowicz, P. 1997, Ann. Rev. Earth Planet Sci., 25, 175
Artymowicz, P. & Lubow, S.H. 1996, ApJ, 467, L77
Artymowicz, P., Lubow, S. H., and Kley, W. 1998, in Planetary systems - the long view, Eds. Celnikier, L. et al., Editiones Frontieres, p. 381
Artymowicz, P. 2000, in From Extrasolar Planets to Cosmology: The VLT Opening Symposium, Eds. J. Bergeron, A. Renzini, Springer Verlag, p. 391
Black, D. 1997, ApJ, 490, L171
Boss, A. 1995, Science, 287, 360
Boss, A. 2001, this volume
Bryden, G., Chen, X., Lin, D.N.C., et al. 1999, ApJ, 514, 344
Butler, R.P., Marcy, G. W., Vogt, S.S., Fischer, D. A. 2001, this volume
Dick, S.J. 1982, Plurality of Worlds, Cambridge U. Press
Duquennoy, A., & Mayor, M. 1991, A&A, 248, 485
Goldreich, P., & Tremaine, S. 1979, ApJ, 233, 857
Goldreich, P., & Tremaine, S. 1980, ApJ, 241, 425
Heap, S., Lindler, D., Lanz, T., et al. 2000, ApJ, 539, 390
Holman, M.J., & Wiegert, P.A. 1999, AJ. 117, 621
Kley, W. 1999, MNRAS, 303, 696
Koerner, D. W., Ressler, M. E., Werer, M. W., Backman, D. E. 1998, ApJ, 503, L83
Lagrange, A-M., Backman, D., & Artymowicz, P. 2000, in Protostars and Planets IV, eds. V. Mannings et al. (Tucson: Univ. Arizona Press), 703

Laughlin, G., & Adams, F. 1999, ApJ, 508, L171

Laughlin, G., & Bodenheimer, P. 1994, ApJ, 436, 335

Levison, H.F., Lissauer, J.J., & Duncan, M.J. 1998, AJ, 116, 1998

Lin, D.N.C., & Papaloizou, J.C.B. 1979, MNRAS, 186, 799

Lin, D.N.C., & Papaloizou, J.C.B. 1993, in Protostars and Planets III, eds. E.H. Levy and J.I. Lunine (Tucson: Univ. Arizona Press), 749

Lin, D.N.C., Bodenheimer, P., & Richardson, D.C. 1996, Nature, 380, 606

Lissauer, J.J. 2001, Nature, 409, 23

Lubow, S.H., & Artymowicz P. 2000, in Protostars and Planets IV, eds. V. Mannings et al. (Tucson: Univ. Arizona Press), 731

Lubow, S.H., Seibert, M., & Artymowicz P. 1999, ApJ, 526, 1001

Mannings, V., Boss, A. & Russell, S. (eds.), 2000, Protostars and Planets IV,(Tucson: Univ. Arizona Press), p. 1400

Marcy, G., et al. 2001, www.exoplanets.org

Mathieu, R. D. 1994, ARA&A, 32, 465

Mathieu, R. D., Ghez, A. M., Jensen, E.L.N., & Simon, M. 2000, in Protostars and Planets IV, eds. V. Mannings et al. (Tucson: Univ. Arizona Press), 703

Mayor M., & Queloz, D. 1995, Nature, 378, 355

Nelson, A. 2000, ApJ, 537, L65

Owen, T., & Bar-Nun, A. 1995, Icarus, 116, 215

Owen, T., Mahaffy, P., Niemann, H.B., Atreya, S., et al. 1999, Nature, 402, 269

Papaloizou, J.C.B., Nelson, R.P. & Masset, F. 2001, A&A, 366, 263

Schneider, J. 2001, www.obspm.fr/planets

Takeuchi, T., & Artymowicz P. 2001, ApJ, in print

Ward, W.R. 1997, ApJ, 482, L211

Weidenschilling, S.J., Marzari, F. 1996, Nature, 384, 619

Weinberger, A. J., Becklin, E. E., Schneider, G., Smith, et al. 1999, ApJ, 525, L53

Wolszczan A., Frail D.A. 1992, Nature, 355, 145