Part II

Progress in the theory of planet formation

Planetary Systems in the Universe - Observation, Formation and Evolution
Proceedings IAU Symposium No. 202, ©2004 IAU
Alan Penny, Pawel Artymowicz, Anne-Marie Lagrange, & Sara Russell, eds.

Modes of Gaseous Planet Formation

Alan P. Boss
Carnegie Institution of Washington, Department of Terrestrial Magnetism, 5241 Broad Branch Road, N. W.,
Washington, D. C. 20015-1305, U.S.A.

Abstract. The discovery of gas giant planets around nearby stars has launched a new era in our understanding of the formation and evolution of planetary systems. However, none of the over four dozen companions detected to date strongly resembles Jupiter or Saturn: their inferred masses range from sub-Saturn-mass to 10 Jupiter-masses or more, while their orbits extend from periods of a few days to a few years. Given this situation, it seems prudent to re-examine mechanisms for gas giant planet formation. The two extreme cases are top-down or bottom-up. The latter is the core accretion mechanism, long favored for our Solar System, where a roughly 10 Earth-mass solid core forms by collisional accumulation of planetesimals, followed by hydrodynamic accretion of a gaseous envelope. The former is the long-discarded disk instability mechanism, where the protoplanetary disk forms self-gravitating, gaseous protoplanets through a gravitational instability of the gas, accompanied by settling and coagulation of dust grains to form solid cores. Both of these mechanisms have a number of advantages and disadvantages, making a purely theoretical choice between them difficult at present. Observations should be able to decide the dominant mechanism by dating the epoch of gas giant planet formation: core accretion requires more than a million years to form a Jupiter-mass planet, whereas disk instability is much more rapid.

1. Introduction

Prior to 1995, theoretical work on planet formation was largely limited to the problem of the origin of the Solar System – very little attention was directed toward trying to understand how the planet formation process might operate around other stars. Perhaps the most notable exception was the pioneering work by Wetherill (1996) on terrestrial planet formation in the case of stars and protoplanetary disks with varied masses and with different assumptions about the location of any gas giant planets in the system. As a result of this theoretical single-mindedness, we now have a fairly mature theory of terrestrial planet formation, two very different hypotheses about how the gas giant planets (Jupiter and Saturn) formed, and a few suggestions regarding the formation of the ice giant planets, Uranus and Neptune.

However, since 1995 the direction of the field of planetary origins has been profoundly altered by the discovery of the first extrasolar planets, forcing us to look outward and to try to understand how these alien planets and planetary systems could have formed and evolved. Theorists must now confront their favorite theories not only with the Solar System's coterie, but also with the widening and often surprising circle of extrasolar planets. As a result of this observational prodding, the general theory of planet formation is likely to evolve rapidly in the next decade. We begin with a brief review of the observational progress to date, and then turn to the implications for the theory of the formation of gas giant planets.

2. Extrasolar Planets

The first definitive discovery of an extrasolar planet orbiting a solar-type star was that of 51 Pegasi's $\approx 0.5 M_{Jup}$ companion (Mayor & Queloz 1995). Subsequent discoveries have been made at such a rapid rate that published review articles are quickly outdated (e.g., Marcy & Butler 1998 listed only eight planet candidates). Fig. 1 depicts the 50-odd planetary candidates announced through August 2000.

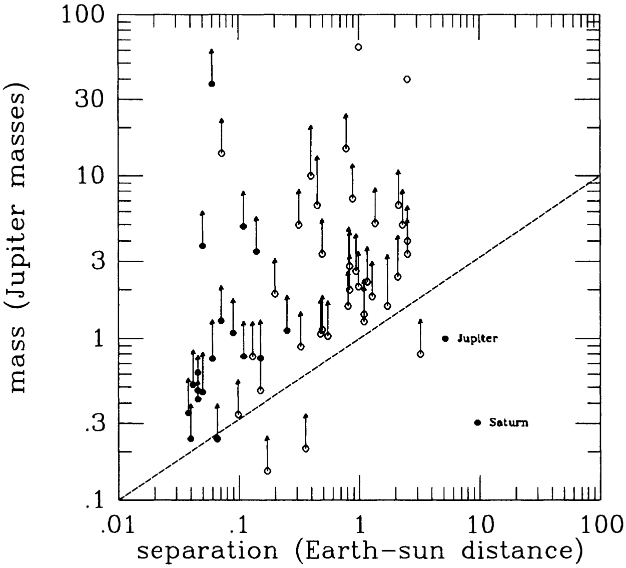

Fig. 1. Discovery space for extrasolar planets and brown dwarf stars. The oblique dashed line illustrates the sensitivity limit for radial velocity detections at a given accuracy. Three of the most recent discoveries fall below this line (79 Ceti: Marcy et al. 2000; HD 83443: Mayor et al., this volume; Eps Eri: Cochran et al., this volume), demonstrating ongoing progress in reducing observational noise. All of these objects were found by the radial velocity method, which yields only a lower limit on the companion's mass. Filled circles represent roughly circular orbits, while open circles represent significantly eccentric orbits.

A number of significant surprises are evident in Fig. 1. First, 51 Pegasi's planet has an orbital period of 4.23 days and a semimajor axis of 0.05 AU, compared to Jupiter's values of 12 years and 5.2 AU, respectively. Several more "hot Jupiters", and even "hot Saturns", have now been found, and extrasolar gas giant planets seem to be distributed throughout the range of semimajor axes from about 0.04 AU to 3 AU. Second, the planet candidates with semimajor axes larger than about 0.1 AU tend to have considerably more eccentric orbits than Jupiter ($e \approx 0.05$), a characteristic initially thought to be more applicable to binary stars than to planets. Third, many of these very low mass companions have minimum masses that are considerably more massive than Jupiter, by a factor of order 10 or so. A few companions have minimum masses of about 30 M_{Jup} or more and may be brown dwarf stars, which are expected to be rare in orbits around these mostly solar-type primary stars. Very few companions have minimum masses in the range of 10 M_{Jup} to 30 M_{Jup}, which implies that this range is the dividing point between planets and brown dwarf companions to solar-type stars. Theoretical estimates place the absolute minimum mass of a brown dwarf star at about 10 M_{Jup}, a value that appears to be roughly consistent with the observations to date.

The planetary nature of the companions with minimum masses below about 10 M_{Jup} is supported by three recent discoveries. First, the Upsilon Andromedae triple planet system (Butler et al. 1999) is highly significant because there are no stellar systems known to consist of a massive primary orbited by three smaller stars. Second, the discovery of the first transiting planet, on a short period orbit around HD 209458 (Charbonneau et al. 2000; Henry et al. 2000), provided a further argument that at least some of these objects are gas giant planets: the transiting planet's mass is $\approx 0.7 M_{Jup}$, and its radius of $\approx 1.4 R_{Jup}$ (Mazeh et al. 2000) is in good agreement with theoretical expectations for a hot Jupiter (Burrows et al. 2000). Finally, the spectroscopic transit of HD 209458 observed by Queloz et al. (2000) shows that the planet revolves in the same direction as the star rotates, and its orbit lies close to the star's equatorial plane.

The discovery of the transiting planet around HD 209458 is consistent with the expectation that if the rotational axes of the primary stars are randomly distributed in angle, and hence presumably so are the orbital planes of the planets, about 10% of the hot Jupiters should show transits. While some of the planetary candidates may turn out to be more massive brown dwarf or low mass stars with chance pole-on orientations, it appears likely that the bulk of the objects in Fig. 1 have true masses less than 10 M_{Jup}. Astrometric detections will be needed to determine the true masses of planets that do not transit.

3. Gas Giant Planet Formation Theories

All modern theories of gas giant planet formation envision the process occurring in the protoplanetary disks that are now known to accompany the star formation process (Beckwith, this volume; Stapelfeldt, this volume). There appear to be only two extreme possibilities: starting small, and growing larger (core accretion), or starting large, and perhaps losing mass later (disk instability). Core accretion is the generally preferred mechanism, though the disk instability mechanism has resurfaced.

3.1. Core Accretion

The terrestrial planets are widely believed to have formed in the inner solar nebula through the collisional accumulation of successively larger, solid bodies micron-sized dust grains, kilometer-sized planetesimals, lunar-sized planetary embryoes, and finally, Earth-size planets (Wetherill 1990). The same process presumably occurred in the outer solar nebula, leading to the growth of roughly 10 $M_\oplus$ cores, which are large enough to begin to accrete massive gaseous envelopes from the disk gas (Mizuno 1980). The cores are expected to grow through runaway accretion (Lissauer 1987), where the largest bodies grow the fastest because self-gravity increases their collisional cross-sections.

The time scale for core accretion depends strongly on the initial surface density of solids. Pollack et al. (1996) showed that with $\sigma_s = 10$ g cm^{-2} at 5.2 AU, a standard model requires 8×10^6 years to form Jupiter. When σ_s is decreased to 7.5 g cm^{-2}, the time required jumps to over 5×10^7 years. When σ_s is increased to 15 g cm^{-2}, the time falls to less than 2×10^6 years, but the core mass then far exceeds the possible values for Jupiter. At 9.5 AU (Saturn's orbit), growth is even slower. These times tend to exceed estimates of disk lifetimes of a few million years (Wolk & Walter 1996) or less (Bally et al. 1998). Assuming that Jupiter and Saturn could form fast enough to accrete gaseous envelopes, Uranus and Neptune might then be explained by their having formed too slowly to have captured significant disk gas. The formation of the ice giant planets is poorly understood (Levison et al. 1998) and may require rethinking, such as formation of extra cores between Jupiter and Saturn followed by outward migration to their present locations (Thommes et al. 1999).

Core accretion models have now been extended to several of the extrasolar planets (Figs. 2 and 3; Bodenheimer et al. 2000), though the difficulties encountered in forming gas giant planets well inside 5.2 AU imply that these planets probably formed farther out than their current distances and then experienced inward orbital migration. Migration is likely to result from interactions between the planet and the gaseous portion of the disk, either before or after the planet becomes large enough to open a gap in the disk (Lin et al. 1996; Ward 1997; Artymowicz, this volume). Gravitational interactions between planets that formed too close together is also a possibility (Weidenschilling & Marzari 1996), and would lead to eccentric orbits for the surviving planets.

Core accretion clearly can lead to large core masses and to non-solar bulk compositions, and forming a Jupiter after 10^6 years can be consistent with the absence of a major planet in the asteroid belt (e.g., Wetherill 1996). However, in addition to the time scale problem, there are other difficulties for core accretion. The gas giant planets are now thought to have core masses (Guillot et al. 1997) that may be too small to induce gas accretion. A 10 $M_\oplus$ core may migrate inward to the protosun in $\sim 10^4$ years via disk interactions (Ward 1997; Papaloizou & Larwood 2000), well before a gaseous envelope can be accreted. Assuming the core can avoid the migration problem and accrete an envelope, the subsequent opening of a disk gap will slow the further growth of the planet (Bryden et al. 1999). Finally, avoiding the loss of a gas giant planet's gaseous envelope may require embedding the planet in a disk massive enough to undergo gravitational instability (Wuchterl et al. 2000).

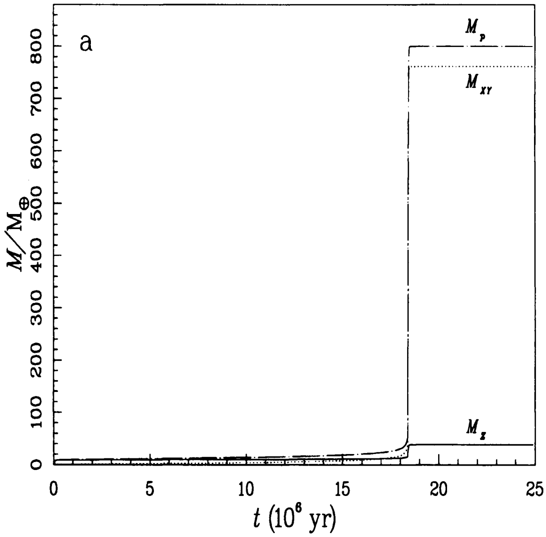

Fig 2. Core accretion model for the *in situ* formation of the gas giant planet orbiting 47 UMa at a distance of 2.1 AU in a massive protoplanetary disk. In this figure as well as Fig. 3, the solid line shows growth of the solid core with time, dotted line gives the mass of the gaseous envelope, while the dot-dashed line is the total mass. In this model, the accretion rate of solids is calculated as a function of time but is typically $\sim 10^{-7} M_\oplus$/yr. Nearly 20 Myr is required to form the planet. [Adapted from Bodenheimer et al. 2000.]

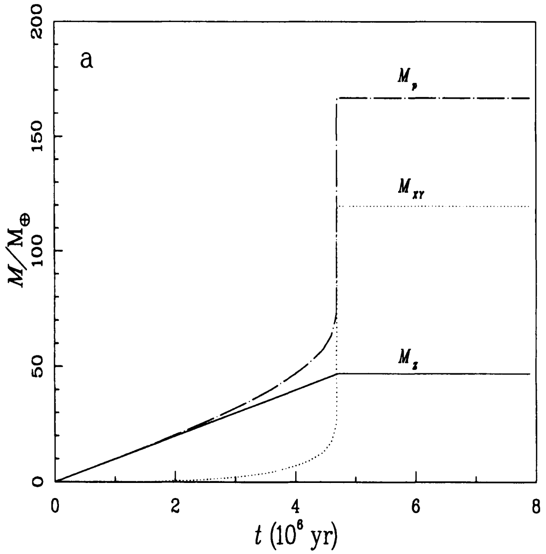

Fig 3. Core accretion model for the *in situ* formation of the gas giant planet orbiting 51 Peg at a distance of 0.05 AU in a standard protoplanetary disk. The accretion rate of solids is assumed to be constant at $10^{-5} M_\oplus$/yr. Nearly 5 Myr is then required to form the planet. [Adapted from Bodenheimer et al. 2000.]

3.2. Disk Instability

Considering these concerns regarding core accretion, it seems worthwhile to consider the alternative mechanism, disk instability, where gas giant protoplanets form rapidly through a gravitational instability of the gaseous portion of the disk (Cameron 1978). The disk instability mechanism was discarded in the 1980s primarily because of the inferred large core masses for Jupiter and Saturn, but these estimates are now much lower (Guillot et al. 1997). Furthermore, it is likely that a core can form in a giant gaseous protoplanet by sedimentation of dust grains to the center, prior to contraction of the protoplanet to planetary densities and temperatures (Boss 1997). Disk instability and core formation would then occur nearly simultaneously within $\sim 10^3$ years, with contraction to planetary densities occurring in $\sim 10^5$ years. Disk instabilities have also been neglected because of the belief that gravitational instability will lead only to rapid transport of mass and angular momentum in the disk rather than to the formation of self-gravitating clumps (Cassen et al. 1981; Laughlin & Bodenheimer 1994).

If disk instabilities can produce gas giant protoplanets, then several properties of the extrasolar planets might be explained. Disk instability works best in relatively massive disks, so it should be able to form fairly massive extrasolar planets. The instability can proceed in a disk with a surface density at 5.2 AU comparable to that required in the standard core accretion models (Boss 2000). Protoplanets might then form with initially eccentric orbits (Fig. 4), removing the need to pump up their eccentricity by subsequent interactions. The disk instability mechanism avoids problems with orbital migration with respect to the disk and gap-limited mass accretion because the clumps form directly in the disk, without requiring prior formation of a core that can migrate with respect to the disk, and the clumps quickly open a disk gap, preventing motion with respect to the gas, but after the protoplanet's mass is already captured. Recent work suggests that rapid gas giant planet formation by disk instability does not unduly impede terrestrial planet formation and helps to limit growth in the asteroid belt (Kortenkamp & Wetherill 2000).

Disk instability has its own share of problems, however. The instability may require a trigger to produce clumps, such as the pile-up of gas in a magnetically-dead zone, episodic infall onto the disk, or a close encounter with another star. Even if clumps form, the circumstances under which they might survive to become gas giant planets remain to be understood. Disk instability may have trouble forming sub-Jupiter-mass planets, unless one invokes tidal stripping of the protoplanet's envelope during a phase of inward orbital migration. Tidal stripping would also seem to be needed in order to produce non-solar bulk compositions, along with the ongoing accretion of planetesimals (the latter is also required for the core accretion mechanism). Finally, disk instability would seem to be a very poor means to form the ice giant planets, so Uranus and Neptune would still have to form by core accretion.

Much of the current focus deals with the disk's thermodynamics. Models assuming locally isothermal behavior can lead to the formation of clumps (Boss 1997, 1998a; Nelson et al. 1998; Armitage & Hansen 1999), whereas assuming locally nonisothermal behavior can stifle the formation of clumps (Pickett et al. 2000; Nelson 2000). 3D radiative transfer is needed to improve the disk's thermodynamical description.

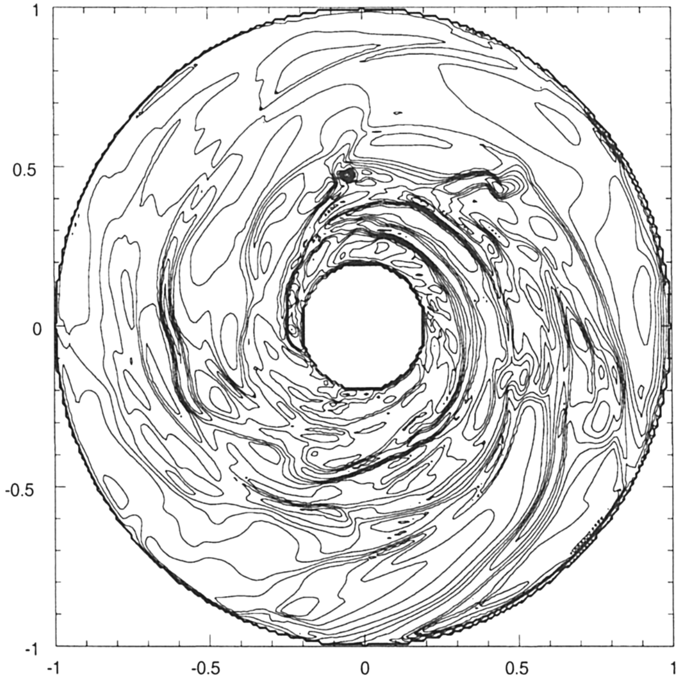

Fig. 4. Midplane density contours from a high spatial resolution, 3D hydrodynamical model of a gravitationally unstable disk after 374 years of evolution. The initial disk is marginally unstable ($Q_{min} = 1.3$) with a mass of $0.091 M_\odot$ within a total radius of 20 AU. Each contour represents a factor of two change in density. The maximum density of 4.3×10^{-7} g cm^{-3} occurs in a relatively long-lived clump located near 12 o'clock. [Adapted from Boss 2000.]

4. Determining the Epoch of Gas Giant Planet Formation

While theoretical work may not yield a consensus about the mechanism of gas giant planet formation any time soon, there is at least one powerful observational test. Core accretion requires at least a million years to form a planet, whereas disk instability should form planets around the very youngest stars, so by dating the epoch at which young stars first show signs of Jupiter-mass companions, the relative importance of each mechanism should be clarified. This epoch could be determined either by searching for astrometric wobbles (Boss 1998b) or by direct imaging of gaseous protoplanets or the spiral arms and gaps that they produce in the disk. The Space Interferometry Mission (SIM) should be capable of the former, while the Atacama Large Millimeter Array (ALMA) could accomplish the latter. We may just have to wait a decade or two for the answer.

References

Armitage, P. J., & Hansen, B. M. S. 1999, Nature, 402, 633

Bally, J. et al. 1998, AJ, 116, 854

Bodenheimer, P., Hubickyj, O., & Lissauer, J. J. 2000, Icarus, 143, 2

Boss, A. P. 1997, Science, 276, 1836

Boss, A. P. 1998a, ApJ, J., 503, 923

Boss, A. P. 1998b, Nature, 393, 141

Boss, A. P. 2000, ApJ, J., 536, L101

Bryden, G. et al. 1999, ApJ, 514, 344

Burrows, A. et al. 2000, ApJ, 534, L97

Butler, R. P. et al. 1999, ApJ, 526, 916

Cameron, A. G. W. 1978, Moon Planets, 18, 5

Cassen, P. M. et al. 1981, Icarus, 48, 377

Charbonneau, D. et al. 2000, ApJ, 529, L45

Guillot, T., Gautier, D., & Hubbard, W. B. 1997, Icarus, 130, 534

Henry, G. W. et al. 2000, ApJ, 529, L41

Kortenkamp, S. J., & Wetherill, G. W. 2000, Icarus, 143, 60

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

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

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

Lissauer, J. J. 1987, Icarus, 69, 249

Marcy, G. W., & Butler, R. P. 1998, ARA&A, 36, 57

Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, ApJ, 536, L43

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

Mazeh, T. et al. 2000, ApJ, 532, L55

Mizuno, H. 1980, Progr. Theor. Phys., 64, 544

Nelson, A. F. 2000, ApJ, 537, L65

Nelson, A. F. et al. 1998, ApJ, 502, 342

Papaloizou, J. C. B., & Larwood, J. D. 2000, MNRAS, 315, 823

Pickett, B. K. et al. 2000, ApJ, 529, 1034

Pollack, J. B. et al. 1996, Icarus, 124, 62

Queloz, D. et al. 2000, A&A, 359, L13

Thommes, E. W., Duncan, M. J., & Levison, H. F. 1999, Nature, 402, 635

Ward, W. R. 1997, Icarus, 126, 261

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

Wetherill, G. W. 1990, Ann. Rev. Earth Planet. Sci., 18, 205

Wetherill, G. W. 1996, Icarus, 119, 219

Wolk, S. J., & Walter, F. M. 1996, AJ, 111, 2066

Wuchterl, G., Guillot, T., & Lissauer, J. J. 2000, in Protostars and Planets IV, V. Mannings, A. P. Boss, S. S. Russell, eds. (U. Arizona: Tucson), 1081