

Formation of Giant Planets

David J. Stevenson

Caltech, 150-21, Pasadena CA 91125

Abstract. This is a descriptive and non-mathematical summary of giant planet formation theories. There are two end-member models, core accretion and disk instability. In the core accretion model, several to ten Earth masses of solid (ice and rock) accumulate and this promotes subsequent gas accretion. In the disk instability model, a self-gravitating sphere of gas forms (somewhat analogous to star formation) and a core may arise through condensation and rainout from its low density envelope. The core accretion model may have a time-scale difficulty because the formation of the embryo and infall of gas is not fast relative to the time at which the gaseous disk is removed. The disk instability model suffers from the current theoretical inability to follow the development of these putative instabilities through to the formation of a planet. Observational data on core masses in Jupiter and Saturn do not clearly favor one model. However, the existence and nature of Uranus and Neptune strongly suggest that the formation of appropriate embryos occurs. Moreover, there is considerable flexibility in the elapsed time required to form Jupiter and Saturn by the core accretion model, including times of only a few million years, compatible with disk lifetimes. This suggests that core accretion remains the favored mode in our solar system. It is possible that both mechanisms operate in general and that many extrasolar planetary systems make use of the disk instability mode. Future theory and observations are essential for deciding this issue.

INTRODUCTION

Giant planets are often defined as those bodies that are mostly hydrogen but are distinguished by stars through their inability to undergo fusion. For the purposes of this brief review, I will also include discussion of bodies such as Uranus and Neptune (sometimes called ice giants) because they may provide crucial insight into the formation process. The central question is obvious: How did these bodies form? There are two end-member models that people consider, those where the process has a lot in common with star formation and involves the instability of gas masses, and those where the accumulation of solids precedes the addition of gas. I refer to these as the disk instability and core accretion models, respectively. The disk instability mechanism might be thought of as closely related to the Jeans collapse process for forming low mass sub-stellar companions.

This brief review is descriptive and non-mathematical. Recent more thoroughly documented reviews cover dynamical aspects [1], extrasolar aspects [2] and cosmochemical aspects [3] very well, but do not offer the descriptive comparison of models that I provide here. We are not yet at the point where we can directly observe giant planet formation in other systems. The central question of giant planet formation must therefore be tackled through consideration of the outcome: How is the formation mechanism expressed in the structure and chronology of the observed system? In our solar system, we would seek the answer to this question by looking at the structure of the planets, their dynamical relationships (orbits, etc.), and their satellite systems. But we must also look at the time

CP713, *The Search for Other Worlds: Fourteenth Astrophysics Conference*,
edited by S. S. Holt and D. Demings

© 2004 American Institute of Physics 0-7354-0190-X/04/\$22.00

scales of processes, inferred by direct observation of materials (isotopic information) or analog systems (star formation regions). This turns out to be a central but unresolved issue for the success of models. Of course, we should be guided by evidence of other planetary systems, even though the time scale constraints may be less secure in those cases. Is the formation mechanism universal or could it vary from system to system? In this review, I will also discuss what we need to make further progress.

THE STANDARD PICTURE

Although many details are poorly understood, a standard picture has emerged for the star and planet formation process, supported by observation and by theory [4,5,6]. In this picture, here discussed for stars of order a solar mass, collapse of an interstellar cloud of gas and dust proceeds on a short time scale (a few hundred thousand years), leading to a disk that is required by accommodation of the angular momentum budget of the infalling material. A process of angular momentum redistribution within the disk feeds a central protostellar mass. The nature of this process is not well understood but can be thought of as a “viscosity”. It causes redistribution of angular momentum (and therefore mass) within the disk on a time scale of one to ten million years, comparable to the total lifetime of the disk. The disk is plausibly of order 100 AU in extent, and typically with a thickness (or scale height) that is about 0.1 of the local radius. The disk mass may be of order 0.1 solar mass, equivalent to a hundred Jupiters. The surface density of the disk declines with radius, but slowly enough that the mass tends to be mostly in the outer parts of the disk. If this disk is dense enough and cold enough then instabilities may occur within it, as it accumulates, especially while the central mass is still low. These are essentially Jeans instabilities but the Keplerian shear as well as the usual competition of pressure and gravity mediate them. If the disk is not very near the stability threshold, then the characteristic time-scale for the development of instabilities can be of order the orbital time, 10 to 1000 yrs.

Independent of this, small (kilometer-sized) solid bodies aggregate quickly, on time scales of a million years or less [7,8]. It is thought that larger bodies may also form quickly, by so-called runaway accretion. This process relies on the small relative velocity of planetesimals and the consequent high gravitational collision cross-section of the largest embryos in the swarm. This cross-section can be as much as a thousand times the geometric cross-section, so that time scales of accumulation can be reduced by a corresponding factor. This lowers the formation time of the largest solid embryos from about a billion years to about a million years or even less at Jupiter orbit, assuming plausible surface densities of solids. The outcome of runaway accretion is to form a set of planetary embryos that have accumulated most of the mass within an orbital radial distance of a few Hill sphere radii. This corresponds to Mars-sized bodies in the terrestrial planet region and Earth mass to ten Earth mass bodies in the giant planet region. In the core accretion scenario of giant planet formation, these embryos are the seeds for accumulating large masses of gas. Massive embryos can also undergo rapid orbital migration in a hundred thousand years or even less by exchanging angular momentum with the much more massive disk. This can cause a rapid inward spiral of

the planet and is a natural explanation for the small orbital radii discovered for many extrasolar planets. Apparently, this process was largely avoided in our solar system. At some point, estimated to be at a few million to ten million years after initial cloud collapse, the gaseous disk is removed through the combination of inflow onto the central star and outflow back into the interstellar medium. This event must postdate the formation of Jupiter and Saturn.

GIANT PLANET COMPOSITION

Detailed reviews of giant planet structure include [9,10,11]. The material available to build planets can be divided into three classes: “Gases”, “ices” and “rocks”. Gas refers to primarily hydrogen and helium, the most abundant elements in the universe. They also happen to be materials that do not condense as solids or liquids under conditions that are encountered during planet formation. The material deep within a planet such as Jupiter can be called a liquid metal but it has a much higher entropy than the value at the conventional critical point of molecular hydrogen where liquid and gas merge. Ice refers to the compounds that form from the next most abundant elements: oxygen, carbon, and nitrogen. These compounds are primarily hydrides (water, methane and ammonia) but also include carbon monoxide, carbon dioxide, and molecular nitrogen among others. The ice label does not mean that these constituents are necessarily found as solid materials, but that is indeed the form in which they would provide the building blocks for planets. However, condensation requires low temperatures, corresponding to the orbit of Jupiter and beyond in our solar system. Ice is roughly two orders of magnitude less abundant than gas, by mass. “Rock” is even less abundant and refers to everything else: primarily magnesium, silicon and iron, and the oxygen that would naturally combine with these elements (far less than the total oxygen because of its much higher abundance). These elementary considerations enable us to understand why it is improbable that we would ever encounter a Jupiter mass planet that is made mostly of iron, say. There would typically be insufficient iron around during the formation of a planetary system to imagine doing this. Our inclination to interpret Jupiter mass extrasolar planets as gas balls (even when there is no information on the size or composition) is not mere solar system prejudice but guided by cosmochemical principles. While it is tempting to argue that planets forming close to stars might tend to have a composition dominated by materials that condense at high temperature, we probably cannot exclude formation of gas balls at small radii. We certainly cannot exclude orbital migration as an important process in the structure of many planetary systems, probably including most of those extrasolar systems thus far discovered, even if it had only a modest role to lay in our own solar system.

There is no doubt that Jupiter and Saturn are mostly hydrogen and helium. No other material of sufficiently low density can explain the global properties of these bodies. However, Jupiter is much closer to cosmic (or primordial solar) composition than Saturn. Roughly speaking, Jupiter and Saturn have similar total amounts of heavy elements (all elements heavier than hydrogen or helium), but Jupiter is over three times the mass of Saturn. We have constraints on the interior structure of these bodies that

arise primarily from the observed gravity field (including the gravitational moments caused by planetary rotation) but also from deep atmosphere composition, magnetic field, heat flow and laboratory and theoretical equations of state. Jupiter may have a dense core; Saturn almost certainly has a dense core. It may seem surprising that we can be more precise about Saturn than about Jupiter, since better data exist for Jupiter. However, the putative core of Jupiter is a tiny fraction of the total mass, perhaps about a percent (i.e., three Earth masses), and it accordingly has a very small effect on planetary structure. The common practice of placing a separate core of heavy elements at the centers of these planets is governed by simplicity, rather than by observation. To varying degrees, the “core” could have a fuzzy boundary with the overlying hydrogen-rich envelope. Uncertainties in the hydrogen equation of state continue to be a major source of uncertainty in the interior models.

Both of these planets are enriched in heavy elements throughout (and separate from the presence or absence of a core). In Jupiter, this enrichment is probably about a factor of three relative to cosmic abundance and is readily observed in those volatile components that do not condense out in the observable atmosphere. We cannot be sure that this is the enrichment for water, which condenses out deep and was not observed to be enriched in the presumably dry region that Galileo probe sampled. Water is the most abundant carrier of oxygen and therefore presumably the most abundant heavy material in Jupiter. However, interior models support about ten Earth masses of heavy elements mixed throughout the hydrogen, consistent with this factor of three. It is particularly interesting that this factor of three is even seen in the heavy noble gases, including argon. The threefold enrichment of argon suggests delivery to Jupiter of material that condensed at very low temperatures, probably around 40K, since there is no known way of incorporating argon into solid bodies in large amounts at higher temperature. These planets also supported in situ formation of a satellite system. The Galilean satellite system is particularly impressive and may contain important clues to the last stages of giant planet formation [12]. Ganymede and Callisto are roughly half water ice, and Callisto has most of this ice mixed with rock. It follows that conditions must be appropriate for the condensation of water ice at the location where Ganymede formed, and conditions at Callisto must have allowed formation of that body on a time scale exceeding about 0.1 million years, so that water ice would not melt and lead to a fully differentiated structure.

Uranus and Neptune are far less well understood than Jupiter or Saturn. However, there is no doubt that they are mostly ice and rock, yet also possess two or so Earth masses of gas each. The atmospheres have solar hydrogen to helium ratios (though with large uncertainty because this determination is based on the pressure induced absorption features of hydrogen, a method that has been unreliable for Jupiter and Saturn). The amount of hydrogen extractable from the ices is in principle about 0.2 of the total mass (assuming the hydrogen was delivered as water, methane and ammonia) and this is marginally close to the hydrogen mass required by interior models. Moreover, there is the possibility that methane would decompose into carbon and hydrogen at extreme pressures. However, the atmospheres of Uranus and Neptune are highly enriched in methane (thus limiting massive decomposition of this compound to very deep regions, if any, and there is no experimental or theoretical evidence for extensive decomposition of water or ammonia under the conditions encountered inside these bodies. Consequently,

it is not plausible to derive even one Earth mass of predominantly hydrogen gas from the breakdown of hydrogen-bearing ice or rock, even leaving aside the dubious proposition that such decomposed hydrogen would rise to the outer regions of the planet. This gas appears to have come from the solar nebula. Uranus and Neptune must have formed largely in the presence of the solar nebula, a very stringent constraint on the formation of solid bodies.

It is often supposed that the presence or absence of a core in Jupiter (say) can be placed in one-to-one correspondence with the presence or absence of a nucleating body that caused the inflow of gas to form the much more massive envelope. However, there is no neat correspondence between mode of giant planet formation and current presence of a core. One could imagine a core instability model even if there is no core remaining, because the core might become mixed into the overlying envelope by convective processes [11]. One could also imagine making a core in the low-density protoplanet phase by rainout. Making a core by rainout once the material is dense and degenerate is far less likely because the high temperatures and dilution make it thermodynamically implausible. The one exception is helium, which can rain out because of its relatively high abundance and extremely tightly bound electronic states relative to the metallic state of the hydrogen. Even helium is only modestly raining out in Jupiter. It seems likely that whatever model one favors for giant planet formation, it should allow for the formation of a core, since Saturn probably has a core and one must in any event explain Uranus and Neptune. It would be special pleading to attribute a different origin for Jupiter than for the other giant planets. The merits of the two end-member models for giant planet formations should therefore be assessed primarily by their dynamical and time scale predictions, though with attention paid to the nature of the cores that they predict.

CORE INSTABILITY MODEL

The central feature of this model is the existence of a core (ice and rock) mass that does not allow for a hydrostatic equilibrium solar composition envelope that merges smoothly with any plausible solar nebula. The failure of hydrostatic equilibrium can then allow for solutions in which there is a steady inflow of gas. In the simplest picture, there is accordingly a critical mass for the solid embryo above which one may form a giant planet. The giant planets are attributed to solid embryos forming in the outer solar system and reaching this critical mass, but failing to do so in the inner solar system. Cameron advanced the idea, but the classic work on this concept was by Mizuno [13]. Popularity for the idea stemmed in large part from the “coincidental” similarity of the critical mass predicted by theory (about ten earth masses) and the range of estimates of actual core masses in Jupiter and Saturn. These core masses were in turn related to the development of planets beyond the “snow line” (the innermost radius in the nebula where water ice condenses), where a higher surface density of solids is likely to exist. Subsequent work has somewhat reduced the attractiveness of the core instability idea for several reasons. First, the actual core masses are in doubt and may not match the requirements for the model. Second, the required critical core mass depends on many parameters and is therefore quite model dependent. We now appreciate that this concept

of a critical mass is fuzzy, because its value depends on the dynamical conditions through a dependence on mass inflow rate (i.e., luminosity). Third and perhaps most important, the total elapsed time for the process to run to completion may be many million years and therefore possibly too long relative to the lifetime of the gas disk (cf. [1]).

The most detailed core instability model is Pollack et al. [14]. This model incorporates a particular scenario for the growth of solid embryos and then follows the inflow of gas and solids onto that embryo (using the simplification of spherical symmetry). The model has three stages. Phase I is the formation of the many earth mass embryos by runaway accretion from a planetesimal swarm. Phase II is the longest and corresponds to the slow growth of a gas and solid component, mediated by the cooling time of the growing protoplanet. Phase III begins when the solid and gas masses are comparable and is a rapid inflow of gas, leading quickly to the final mass of the planet. The model does not explicitly consider the mechanism responsible for the truncation of the gas inflow, which could arise by the formation of a tidally truncated gap in the disk. Phase I is modeled in much the same way that many groups now model the formation of the terrestrial planets, apparently successfully. In this game, success for the terrestrial planets can be measured quite well because we have quantitative clocks (isotopic systems) that tell us about that chronology [7, 8]. It is quite likely that the elapsed time for this process is quite accurately estimated at a million years or less for the embryo that led to the formation of Jupiter. In order to get an embryo of order ten Earth masses it is necessary to assume a disk surface density that exceeds by a factor of three or so the “minimum mass solar nebula” (essentially the value needed to explain the total heavy elements in the giant planets). Equivalently, any such model will always make more embryos than are needed to explain the giant planets. One has to assume that the excess embryos are somehow eliminated. At the end of this phase, the embryo has a gas mass of order an Earth mass or so, roughly comparable to the final state of Uranus or Neptune (and thus potentially offering an explanation for bodies of this kind). Phase II takes the longest time and therefore merits the most scrutiny. In this phase, the zone of solid embryo growth is largely exhausted of solid material and there is a gentle arrival of additional solids and gas. Here, one must question both the assumptions that dictate the solid mass inflow as well as the parameters that dictate the inflow of gas. The former depends on the implicit assumptions that there is complete isolation of neighboring zones of embryo growth, probably an artificiality of the model used. The latter depends on the cooling ability of the proto-giant planet and is accordingly sensitive to the opacity of the envelope. The models used assume opacity not much less than that of the interstellar medium. It is possible that the correct opacity is an order of magnitude or more lower, because much of the solid material has already aggregated into large bodies. Of course, there are collisions and condensation that can create additional grain opacity, but there is no reason to expect an interstellar value. The “standard” Pollack model takes 6 or 8 million years for this phase but one could envisage lowering this by a factor of two or three. This might eliminate the time scale problem that mainly motivates the alternative disk instability model. Phase III is rapid and is a positive feedback because the zone of accumulation of gas grows while the protoplanet contracts under gravity. Although this stage may have important consequences for the final properties of the planet, it does not pose a time scale problem. In summary, the core accretion model is perceived to have problems primarily because it might take too long and might require an excessive initial

embryo mass. However, these problems are soft in the sense that they are somewhat specific to a particular formulation of the model and not generic to all models in this category. There is a large amount of recent and promising work in this area, most of which is unpublished but has been reported at conferences.

DISK INSTABILITY MODEL

Although there are many papers on disk instabilities, especially by Boss (see references in [1] and [15]), there is not as complete an understanding of how this process would lead to observable giant planets as that offered by the core instability model. The fundamental problem here is that the quite sophisticated 2D or 3D disk evolution models can only follow the development of the system for typically times of tens to hundreds of orbits, far less than the actual time of formation of planets. Although the models may include many relevant processes, including shocks and radiation, they may lack the resolution or input physics needed to include all the processes that might provide for disk evolution. Of course, it could be argued that since disk instabilities are “fast”, it does not matter. In this strict and narrow sense, published models may be self-consistent. It does not follow that they are necessarily relevant. The approach to disk instability may be very slow (essentially an accretion time scale) and so the development of an instability may also be slow or even non-existent, if it manifests itself as wave disturbances that redistribute angular momentum and mass. The slow “viscous” processes may actually be competitive with the gravitational instabilities. Even if self-gravitating clumps form, they may not survive over hundreds to thousands of orbits in the Keplerian shear field. This is a very short time-scale compared to others of relevance to the problem of disk evolution and planet formation.

Recently, Boss [15] has extended the disk instability model to a scenario for forming planets such as Uranus and Neptune. He suggests that it might be possible to explain these planets as the remnant of initially Jupiter-mass planets that are stripped of their gaseous envelope by photoevaporation, caused by nearby OB stars. A problem with this model is that it requires a somewhat fortuitous outcome: The observed properties of Uranus and Neptune are very much what you would expect when you form a ten Earth mass solid in the solar nebula. The additional two or so Earth masses of hydrogen is not an accident- it is what the standard theory predicts for that stage [12, 13]. It is unlikely that one could produce an outcome that is fortuitously correct by stripping.

WHAT IS THE CORRECT PICTURE?

If disk instabilities are capable of forming giant planets, it seems likely that they have prevailed in at least some systems. The history of astronomy has taught us that if something can happen, it probably will, at least part of the time. They may even be the most common formation mechanism for giant planets and responsible for the numerous examples of giant planets in the current extrasolar planet catalog. However, the system that we can study most closely is our own and in that singular case, the evidence seems

at present to point towards core instability. Despite time scale concerns, there is the likelihood that there exists a reasonable set of parameters and conditions in which the mechanism can work. There is an approximate correspondence between the mass that is required for the embryo and the allowable core masses that planetary models predict. Very importantly, the mere existence of Uranus and Neptune point strongly towards the formation of appropriate embryos, notwithstanding the provocative model of Boss [15]. Uranus and Neptune appear to be excellent candidate embryos for gas giants, and failed to become gas giants only because the nebula was largely removed at the time they approached their final mass. Of course, there is not yet a good model for the formation of Uranus and Neptune! While it is true that the formation of Uranus and Neptune remains poorly understood, we cannot deny the empirical evidence. They exist and cannot be easily explained any way other than by accumulation from smaller bodies. There is sometimes the tendency in science to seek a single explanation for a phenomenon. In the case of giant planets, this would seem to be a false goal because giant planets can come in many forms. We know of two mechanisms that could work. The physical principles are not seriously in doubt for either model, but the precise set of circumstances needed are not well established. Probably, both mechanisms exist, and the phase space of parameters that allow them to work are quite large, perhaps overlapping. The challenge is to decide which mechanism prevails for each system we encounter, and how this is related to parameters such as metallicity, angular momentum budget, proximity of other (possibly high energy) events such as supernovae, mass of central star, and so on.

THE FUTURE

Despite the lack of a close connection between formation mechanism and presence of a core, there is a need to understand better the nature of the giant planet interiors. It is likely that this will test the relative merits of the formation models. For this, we need more experimental and theoretical work on the hydrogen equation of state. It is humbling to admit that despite decades of work, this simplest of systems remains poorly understood. We also need a Jupiter polar orbiter that collects data on gravity and magnetic field close in to the planet. Future missions should also establish the all important water abundance in the deep atmosphere, either through remote sensing techniques (microwave sounding) or through dropping one or more probes into the atmosphere. Eventually, both techniques are needed. The giant planet formation story for our solar system is surely connected to our existence and to the existence of terrestrial planets. Despite the quite successful application of computer simulations to terrestrial planet accumulation, it is essential to fully incorporate the presence of these giant planets into these simulations [16]. It is likely that a fully consistent picture of the origin of Earth depends on the formation mechanism for Jupiter. The explosion of data on extrasolar planets can also be used to test giant planet formation models. If we can establish the location for their formation and orbital migration, then we can probably use this as a diagnostic for the relative merits of the two models. Observations of young systems might provide direct evidence of the formation mechanism, especially for the disk instability model where there is some possibility of directly observing the large

non-axisymmetric disturbances that these models predict. It should be stressed however that early observational evidence is unlikely to tell us whether we are observing a giant planet in the making. The challenge of actually observing newly forming giant planets is both achievable and highly exciting. The most important constraint may prove to be Uranus and Neptune, and their analogs in other systems (albeit not yet discovered.) If planets of this kind prove to be common then this will point strongly to a core instability model. The disk instability mechanism may coexist, of course, but we will have at least established empirical evidence that large solid bodies can accumulate on appropriate time scales.

REFERENCES

1. A. P. Boss, Formation of gas and ice giant planets, *Earth Planet Sci Lett.*, **202** 513-523 (2002).
2. P. Bodenheimer and D. N. C. Lin, Implications of extrasolar planets for understanding planet formation, *Ann Rev earth Planet Sci.*, **30** 113-148 (2002).
3. J. I. Lunine, A. Coradini, D. Gautier *et al.*, to appear in *Jupiter*, ed F. Bagenal *et al.*, Cambridge Univ. Press (2004).
4. F. H. Shu, F. C. Adams and S. Lizano, Star formation in molecular clouds - observation and theory, *Ann Rev Astron Astrophys*, **25**, 23-81 (1987).
5. R. B. Larson, The physics of star formation, *Reports on Progress in Physics*, **66**, 1651-1697 (2003).
6. J. J. Lissauer, Planet Formation, *Ann Rev Astron Astrophys*, **31**, 129-172 (1993).
7. G. J. Wasserburg, Isotopic abundances: inferences on solar system and planetary evolution, *Earth Planet. Sci. Lett.*, **86**, 129-173 (1987).
8. A. N. Halliday and D. Porcelli, In search of lost planets - the paleocosmochemistry of the inner solar system, *Earth Planet Sci Lett.*, **192**, 545-559 (2001).
9. T. Guillot, Interiors of Giant Planets Inside and Outside our Solar system, *Science*, **286**, 72-77 (1999).
10. W. B. Hubbard, A. Burrows, & J. I. Lunine, Theory of Giant Planets, *Ann Rev Astron Astrophys*, **40**, 103-136 (2002).
11. T. Guillot, D. Stevenson, W. Hubbard, and D. Saumon, The Interior of Jupiter, to appear in *Jupiter*, ed F. Bagenal *et al.*, Cambridge Univ. Press (2004).
12. D. J. Stevenson, Jupiter and its Moons, *Science*, **294**, 71-72 (2001).
13. H. Mizuno, Formation of the Giant Planets, *Prog. Theor. Phys.*, **64**, 544-557 (1980).
14. J. B. Pollack, O. Hubickyj, & P. Bodenheimer *et al.*, Formation of the Giant planets by Concurrent Accretion of Solids and Gas, *Icarus*, **124**, 62-85 (1996).
15. A. Boss, Rapid Formation of Giant Planets by Disk Instability, *Astrophys. J.*, **599**, 577-581 (2003).
16. H. F. Levison and C. Agnor, The role of giant planets in terrestrial planet formation, *Astron J*, **125**, 2692-2713 (2003).