Citation

Alibert, Y.; Broeg, C.; Benz, W.; Wuchterl, G.; Grasset, O.; Sotin, C.; Eiroa, Carlos; Henning, Thomas; Herbst, Tom; Kaltenegger, Lisa; Léger, Alain; Liseau, Réne; Lammer, Helmut; Beichman, Charles; Danchi, William; Fridlund, Malcolm; Lunine, Jonathan; Paresce, Francesco; Penny, Alan; Quirrenbach, Andreas; Röttgering, Huub; Selsis, Frank; Schneider, Jean; Stam, Daphne; Tinetti, Giovanna and White, Glenn J. (2010). Origin and formation of planetary systems. Astrobiology, 10(1) pp. 19–32.

URL

https://oro.open.ac.uk/25521/

License

None Specified

Policy

This document has been downloaded from Open Research Online, The Open University’s repository of research publications. This version is being made available in accordance with Open Research Online policies available from Open Research Online (ORO) Policies

Versions

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding
Origin and Formation of Planetary Systems

Y. Alibert,1,2 C. Broeg,1 W. Benz,1 G. Wuchterl,3 O. Grasset,4 C. Sotin,4 Carlos Eiroa,5 Thomas Henning,6 Tom Herbst,6 Lisa Kaltenegger,7 Alain Léger,8 Réné Liseau,9 Helmut Lammer,10 Charles Beichman,11 William Danchi,12 Malcolm Fridlund,13 Jonathan Lunine,14 Francesco Paresce,15 Alan Penny,16,17 Andreas Quirrenbach,18 Huub Röttgering,19 Frank Selsis,20 Jean Schneider,21 Daphne Stam,22 Giovanna Tinetti,23 and Glenn J. White16,24

Abstract

To estimate the occurrence of terrestrial exoplanets and maximize the chance of finding them, it is crucial to understand the formation of planetary systems in general and that of terrestrial planets in particular. We show that a reliable formation theory should not only explain the formation of the Solar System, with small terrestrial planets within a few AU and gas giants farther out, but also the newly discovered exoplanetary systems with close-in giant planets. Regarding the presently known exoplanets, we stress that our current knowledge is strongly biased by the sensitivity limits of current detection techniques (mainly the radial velocity method). With time and improved detection methods, the diversity of planets and orbits in exoplanetary systems will definitely increase and help to constrain the formation theory further. In this work, we review the latest state of planetary formation in relation to the origin and evolution of habitable terrestrial planets. Key Words: Planet formation—Gas giants—Ice giants—Terrestrial exoplanets—Habitability. Astrobiology 10, 19–32.

1. Formation of Planetary Systems

Planet formation is closely connected to star formation and early stellar evolution (see, e.g., Bodenheimer, 1997; Mannings et al., 2000; Wuchterl et al., 2000; Boss, 2003). Stars form from collapsing clouds of gas and dust. The collapse leads to the formation of a central body, the protostar, which contains most of the mass of the cloud, and a circumstellar disk, which retains most of the angular momentum of the cloud. In the Solar System, the circumstellar disk is estimated to have had a mass of a few percent of the Sun’s mass. The planets form from the material in the circumstellar disk,
which is in this stage also referred to as the protoplanetary disk.

1.1. Giant planet formation theories: core accretion versus disk instability

Currently, there are two main theories on the formation of giant gaseous planets in planetary systems:

- core accretion and
- disk instability.

Most of the work on giant planet formation has been performed in the context of the core accretion mechanism, so its strengths and weaknesses are better known than those of the disk instability mechanism, which has only recently been subjected to serious investigation. In this model, it is presumed that planetesimals form by coagulation and merging of dust particles, in a process that is not totally clear (see, e.g., the recent models of Johansen et al., 2007). If sufficient material is available, such planetesimals can grow by mutual collisions up to terrestrial planet masses and even beyond, and form giant planetary embryos. If such embryos become larger than around 1 Earth mass, they can gravitationally bind some of the surrounding gas (e.g., Perri and Cameron, 1974) and thus form a gas envelope.

Planet growth then proceeds by the concurrent accretion of solids and gas. Numerical calculations have shown that the accretion rate of solids during this phase is typically of the order of $10^{-3} \, M_{\text{Earth}}/\text{year}$, while the gas accretion rate is several orders of magnitude lower than that. This process ends when the feeding zone of the planet becomes severely depleted, which generally occurs before the core has reached its so-called critical mass (the core mass necessary to trigger a rapid and large accretion of gas, giving birth to a giant planet), of the order of $10 \, M_{\text{Earth}}$. Planet growth, therefore, results from the slow accretion of a gaseous envelope. With the accumulation of the envelope’s mass, the feeding zone of the planet also increases. This allows the accretion of more planetesimals. The accretion rates of solid and gas turn out to be relatively constant during this phase, with the gas accretion rate exceeding the solid accretion rate by less than an order of magnitude.

This phase lasts until the core mass has reached a critical value, the critical mass, at which time the system’s evolution proceeds extremely fast by runaway gas accretion, as the envelope is no longer able to maintain hydrostatic equilibrium. The mass of the planet increases correspondingly. The timescale for the formation of a giant planet in this core accretion scenario is an extremely sensitive function of the disk’s surface density (e.g., Pollack et al., 1996). With typical assumed solid surface densities (comparable to the ones expected in the protosolar nebula), formation timescales for Jupiter-class planets are found close to 8–10 Myr. If the surface density is reduced by 75%, the formation timescale is increased to nearly 50 Myr.

Realistic models of planet formation should explain the diversity between the Solar System and exoplanetary systems by varying the assumptions about the protoplanetary nebulae. Variations of nebula properties result in variations of the core growth rates caused by a coupling of the dynamics of planetesimals and the contraction of massive planetary envelopes, as well as changes in the hydrodynamical accretion behavior of the envelopes caused by differences in nebula density, temperature, and orbital distance. The general theoretical understanding of planet formation is at the present stage limited by the following facts:

- it is unclear if and how planetesimals form given the presently available assumptions, approximations, and knowledge of the protoplanetary nebula conditions;
- the nebula that forms the planetary building blocks and embryos might also affect their subsequent loss from the system via various planet-nebula interaction processes;
- the duration of the final stages of planet growth and the nebula lifetimes are uncertain;
- the effect of stellar central mass on planet formation is poorly known. The pioneering study of planet formation around a diversity of stars has been performed by Nakano (1987, 1988a, 1988b). The so-called Kyoto model of Solar System planet formation is applied to stars of 0.5–10 solar masses ($M_{\text{Sun}}$), which approximately corresponds to effective temperatures of about 4,000–30,000 K and luminosities of 0.04–4000 solar luminosities ($L_{\text{Sun}}$) on the stellar main sequence. Planet-forming regions are identified by requiring conditions for appropriate condensed material to be present. Nakano concluded that there should be an upper limit for the masses of stars that can be accompanied by planets and one for stars accompanied by planets with surface temperatures nearly equal to that of Earth.

Core accretion, as derived from the work of Pollack et al. (1996) and briefly presented above, requires several million years or more for a gas giant planet to form in a protoplanetary disk like the protosolar nebula. On the other hand, the disk instability model predicts that the gaseous portion of protoplanetary disks undergoes a gravitational instability that leads to the formation of self-gravitating clumps, within which disk grains coagulate and settle to form cores (Mayer et al., 2002; Boss, 2003).

- Disk instability can form a gas giant protoplanet in a few hundred years.

Disk instability, however, has previously been thought to be important only in relatively massive disks. New three-dimensional, locally isothermal, hydrodynamical models without velocity damping have shown that a disk instability can form Jupiter-mass clumps, even in a disk with a mass of about 0.091 $M_{\text{Sun}}$ within 20 AU. This mass is low enough to be in the range inferred for the solar nebula. The clumps form with initially eccentric orbits, and their survival will depend on their ability to contract to higher densities before they can be tidally disrupted at successive periastrons (Mayer et al., 2002; Boss, 2003). Because the disk mass in these models is comparable to that apparently required for the core accretion mechanism to operate, the models suggest that disk instability could obviate the core accretion mechanism in the solar nebula and elsewhere.

However, it remains to be seen whether disk instability leads to long-lived clumps in models with more detailed thermodynamical treatments than has been the case for the currently used locally isothermal models. Disk instability
models with three-dimensional radiative transfer in the diffusion approximation, detailed equations of state, and dust opacity routines are in progress in order to investigate these questions (Boss, 2004). In addition, the formation of lower-mass planets (below the one of Saturn) seems difficult in this model.

Other arguments against the disk-formation model have been made in view of the Galileo probe’s isotope measurements in Jupiter’s atmosphere (Lunine, 2003). The disk instability model would have produced a gas giant with solar composition at Jupiter’s orbit, but Jupiter’s atmosphere is not of solar composition.

Moreover, the nitrogen isotope ratio in Jupiter’s atmosphere tightly constrains the source of rocky and icy bodies that enriched Jupiter during formation and tends to favor the core formation model. However, one cannot rule out that Jupiter formed in a different way from most exoplanets discovered so far. Therefore, by increasing the parameter space of known exoplanets and putting additional constraints on their structure and composition, future observations from, for example, the James Webb Space Telescope, the Atacama Large Millimeter Array, or the European Extremely Large Telescope will be helpful to test such theories.

2. Basic Principles of Planet Formation

Pre-planetary disks are rotating structures in quasi-equilibrium. The gravitational force is balanced in the radial direction by the centrifugal force augmented by the gas pressure, while in the vertical direction it is balanced by the gas pressure alone. The gravitational force is mostly related to that of the central star. The self-gravity due to the disk itself remains weak in comparison.

Typical nebula densities are more than 2 orders of magnitude below the Roche density, so compression is needed to confine a condensation of mass \( M \) inside its tidal or Hill radius \( R_H = a(M_{pl}/3M_{star})^{1/3} \) at orbital distance \( a \), where \( M_{pl} \) and \( M_{star} \) are the mass of the planet and the host star, respectively. A local enhancement of self-gravity is needed to overcome the counteracting gas pressure.

- The nucleated instability (or core accretion) model relies on an additional gravity field of a sufficiently large solid core (condensed material represents a gain of 10 orders of magnitude in density and therefore self-gravity compared to the nebula gas).
- The disk instability may operate on length scales between the short-scale pressure support and the long-scale tidal support.
- An external perturbation could compress an otherwise stable disk on its local dynamical timescale, for example, by accretion of a clump onto the disk or a close encounter with a stellar companion.

Giant planet-formation theories may be classified by how they provide this local density enhancement into the above three points. If the gravity enhancement is provided by a dynamical process as in the latter two cases, the resulting nebula perturbation is compressionally heated; the matter is optically thick under nebula conditions so that the internal heat cannot radiate rapidly. Giant planet formation would then involve a transient phase of tenuous giant gaseous protoplanets that would be essentially fully convective and contract on a timescale of typically about \( 10^6 \) years (Bodenheimer and Pollack, 1986). However, there are three major problems in planet-formation theories:

- First: the qualitative problem of planetesimal formation, the process of which is not clear today.
- Second: the qualitative problem of migration that could become a quantitative one when migration-rate estimates are too high.
- Third: The purely quantitative formation timescale issue, which may be solved by improving the physics included in planet-formation models. This is the case, for example, when including the consequences of planetary migration within the protoplanetary disk (Alibert et al., 2005a). Furthermore, if the dust present inside the planetary envelope settles down to the planet’s core, this may reduce the opacity and the formation timescale (e.g. Hubickyj et al., 2005).

2.1. Nebular stability

Protoplanetary nebulae that have a solar composition and a mass just sufficient to accommodate the Solar System’s inventory of condensable elements within a few percent of a solar mass are stable. Substantially more massive disks that result from the collapse of cloud cores are self-stabilizing due to the transfer of disk mass to the stabilizing central protostar (Bodenheimer et al., 1993). Moderate-mass nebula disks might evolve that can develop a disk instability, which leads to strong density perturbations so that giant gaseous protoplanets might form when such an instability has developed into a clump (e.g., Decampli and Cameron, 1979; Bodenheimer, 1985; Bodenheimer and Pollack, 1986).

Only condensates that were present initially would rain out to form a core, while material added later by impacts of small bodies after formation of the giant gaseous protoplanets would be soluble in the envelope (Stevenson, 1982). To account for the bulk heavy element compositions of Saturn and Jupiter, planetesimal accretion has to occur after the giant gaseous protoplanets have formed their cores.

2.2. Nucleated instability and core envelope accretion

Planet formation in the context of the core-nucleated instability hypothesis is the consequence of the formation and growth of solid building blocks. According to the planetesimal hypothesis, planets grow within circumstellar disks via pair-wise accretion of kilometer-sized solid bodies. Sufficiently massive planetesimals embedded in a gravitationally stable protoplanetary nebula can capture large amounts of gas and become the cores of giant planets as shown in Fig. 1 (Wuchterl, 1995; Tajima and Nakagawa, 1997; Ikoma et al., 2001).

- The onset of formation of massive envelopes is characterized by the so-called critical and crossover masses. Typical values for the critical mass (largest static envelope for a growing core) are in the range of \( 7-15M_{Earth} \) for standard (minimum mass) assumptions about the protoplanetary nebula.
2.3. Gaseous envelopes and Neptune-class planets

Planetesimals in a stellar nebula are small bodies surrounded by gas. A rarefied equilibrium atmosphere forms around such objects. Detailed atmospheric models with radiative and convective energy transfer (Mizuno, 1980) showed that cores of typically about 10 $M_{\text{Earth}}$ are needed to gravitationally bind a comparable amount of nebula gas. Such envelopes were found to continue mass accretion up to a few tens of $M_{\text{Earth}}$ if the contraction of their envelopes was calculated for quasi-hydrostatic models (Bodenheimer and Pollack, 1986).

Static and quasi-hydrostatic models rely on the assumption that gas accretion from the nebula onto the core is subsonic and the inertia of the gas and dynamical effects such as dissipation of kinetic energy do not play a role, so that the motion of gas is neglected. To check whether gas motion remains slow during accretion, hydrodynamical investigations are necessary. A linear adiabatic dynamical stability analysis of envelopes evolving quasi-hydrostatically revealed that motions would remain slow and, hence, the gaseous envelopes could grow relatively slowly to giant planet masses (Tajima and Nakagawa, 1997). Nonlinear, convective radiation-hydrodynamical calculations of core-envelope protogiant planets (Wuchterl et al., 2000) found two pathways for further evolution, as follows:

- First: typical for lower nebula densities that lead to mostly radiative envelopes and produce Neptune-class planets with relatively small hydrogen-helium envelopes around a large core.
- Second: typical for higher nebula densities that lead to mostly convective envelopes (Wuchterl, 1993; Ikoma et al., 2001) and ultimately massive giant planets like Jupiter and Saturn (Wuchterl, 1995).

The two pathways are separated by the transition of the outer protoplanetary envelopes from radiative to convective energy transfer. An approximate condition separating the two pathways depending on the midplane nebula density. Protoplanets that grow under nebula conditions above that density have larger envelopes for a given core. They also feature a reduced critical mass and accrete envelopes that are more massive than the core.

The first case of these scenarios is interesting for missions that search for habitable planets, because Neptune-class planets with large cores and relatively small hydrogen-helium envelopes may lose this gaseous layer due to hydrodynamic loss.

If the orbits of such planets are located in an inner system, they may lose their hydrogen envelopes due to heating of X-rays and extreme UV, and evolve to a new type of terrestrial planet with a secondary atmosphere, which can be studied by terrestrial planet-finding missions such as the Terrestrial Planet Finder Coronagraph (TPF-C) or Darwin/terrestrial Planet Finder Interferometer (TPF-I). (Kuchner, 2003; Lammer et al., 2003a, 2003b, 2009; Vidal-Madjar et al., 2003; Tian et al., 2005; Penz et al., 2008).

Finally, Boss (2003) applied the disk instability model to the formation of outer giant planets (Neptune and Uranus) and investigated whether clumps with a few Jupiter masses could result from disk instability in the outer solar nebula. If the clumps he found develop later into giant planets, they would have to form a core subsequently and then lose $\sim 97\%$ of their mass (by photoevaporation driven by assumed massive stars that neighbor the early Sun) to become, in the end, Uranus-class planets (Wuchterl et al., 2000).

2.4. Migration of giant planets

The vast majority of the currently known exoplanets have orbital radii much smaller than that of Jupiter. This phenomenon may be fully explained as an observational bias, and exoplanets in wider orbits may well be detected with time. The presence of these close-in extrasolar giant planets (CEGs), however, has to be explained and understood, in particular for a terrestrial planet-finding mission like Darwin, because the formation and evolution of CEGs will undoubtedly influence that of terrestrial planets in a planetary system.

Close-in extrasolar giant planets are generally thought to have started their formation at larger distances from their star than where they are observed today and to have migrated toward their star during their formation (Lin and Papaloizou, 1986; Lin et al., 1996; Ward, 1997a, 1997b; Tanaka et al., 2002; Alibert et al., 2005a, 2005b, 2006; Mordasini et al., 2009a, 2009b). The main arguments against the in situ formation of CEGs are the following:

- Close to the star, it would be impossible to form a protoplanet because the protoplanet would be pulled apart by the stellar tides.
- The temperatures in the very near vicinity of the star are too high for grain condensation; thus there will be no material to form the core of a giant planet.
Even at distances up to about 1 AU, where grains do condense, the disk will not contain enough solid material to form a giant planet’s core or enough gas to form its gaseous envelope.

An additional argument for extrasolar giant planets that migrate through a protoplanetary system is that some of the known CEGPs have highly eccentric orbits, which suggests that, after they formed, they were subjected to close encounters with other giant planets (see, e.g., Weidenschilling and Marzari, 1996; Lin and Ida, 1997; Levison et al., 1998) or perturbations from companions. Note that such dynamical interactions are of a very different physical nature compared to the processes that lead to planetary migration inside a protoplanetary gaseous disk.

Indeed, planet migration in a protoplanetary disk results from the interaction between the planet and the gaseous disk. Theoretical models predict two types of migration scenarios:

- **Type I migration**: the planet has a mass that is small compared to the disk’s mass and cannot open up a gap in the gaseous disk. Due to an imbalance of torques from the inner and outer parts of the disk, the planet loses angular momentum and migrates through the disk toward the star. The drift rate of migration is linear with the mass of the planet until the disk density is significantly perturbed by the moving planet. At that point, the migration rate drops. Type I migration is a fast process, \( \sim 10^5 \) years—much faster than giant planet formation: the currently available type I migration rates are so short that all planets may actually be destroyed by the central star long before the disappearance of the gaseous disk.

Even the most recent analytical calculations in two or three spatial dimensions performed by Tanaka et al. (2002), which resulted in longer migration timescales than those originally derived by Ward (1997a), still do not allow the survival of many of the growing planets. This is difficult to reconcile with the population of presently known exoplanets (e.g., Mordasini et al., 2009a, 2009b). Considerably longer type I migration timescales can be found in calculations by Nelson and Papaloizou (2004), who suggested that, at least for low-mass planets \((M_{\text{pl}} < 30 M_{\text{Earth}})\), turbulent magneto-hydrodynamic disks could considerably slow down the net inward motion of embedded planetary bodies. These considerations seem to indicate that the actual migration timescales may in fact be considerably longer than originally estimated by Ward (1997a) or even by Tanaka et al. (2002). This is confirmed by recent models that take into account energetic effects in a protoplanetary disk. These models indeed predict much lower migration rates (e.g., Kley and Crida, 2008; Paardekooper and Mellema, 2008), whose order of magnitude is similar to that assumed by Mordasini et al. (2009a, 2009b), in order to reproduce the properties of exoplanets in a population synthesis approach.

- **Type II migration**: the planet has a mass that is large enough to disturb the structure of the gaseous disk significantly. The planet opens up a gap in the disk and drifts with the disk and the gap toward the star. Type II migration timescales are found to lie between 0.1–10 Myr.

It is an open question as to how the migration is stopped such that planets do not migrate into the star. It may be that it takes several generations of planets and the time of dissipation of the gaseous disk determines which generation survives. In this case, migration just stops when the gaseous disk has dissipated.

In more recent studies (Alibert et al., 2004; Ida and Lin, 2004; Mordasini et al. 2009a, 2009b), two cases for type II migration were considered. For low-mass planets (with mass negligible compared to that of the disk), the inward velocity is determined by the viscosity of the disk. When the mass of the planet is comparable to the disk’s mass, migration slows down and eventually stops. In the simulation of Alibert et al. (2006), the migration type switches from type I to type II when the planet becomes massive enough to open a gap in the disk, which occurs when the Hill radius of the planet becomes greater than the density scale height of the disk.

Since all relevant timescales (planet formation, disk evolution, and migration) are of the same order of magnitude, it appears difficult to obtain a self-consistent picture while omitting one of these processes.

### 2.5. Giant planet formation with migration

To infer the effect of planetary migration and disk evolution on planet formation, Alibert et al. (2004) specified an initial disk profile \( \Sigma \propto r^{-2} \), where \( r \) is the distance to the star, and a given viscosity parameter, yielding a typical evolution time of the disk of a few Myr. The assumed surface density profile yields isolation masses \( \Sigma \) that are independent of the distance to the star. The gas-to-dust ratio is equal to 70 for midplane disk temperature below 170 K, and 4 times higher for high temperatures. It has been found from these simulations that the effect that gap formation has on formation timescales appears to be low, at least until the runaway accretion phase. Two models have been considered, one without migration and disk evolution (assuming planet formation at 5.5 AU) and one that takes into account migration and disk evolution. The corresponding planetary embryo is assumed to start its formation at 8 AU, this value being chosen so that the planet reaches critical mass at 5.5 AU, where the **in situ** model is computed.

Figure 2a shows the mass of planetesimals and the mass of gas accreted by the planet as a function of time. It should be noted that the mass of accreted planetesimals does not correspond to the core mass, since some fraction of them are destroyed while traversing the envelope and never reach the core.

As in Pollack et al. (1996), the formation timescale is essentially determined by the time period necessary to reach the runaway accretion phase, which occurs shortly after the crossover mass (mass of core equals mass of envelope), \( M_{\text{cross}} \), has been reached.

- Allowing for migration and disk evolution, Alibert et al. (2004) obtained a formation time of about 1 Myr, that is,

\(^2\)The isolation mass is the mass of a planetary core once it has accreted all the material inside its feeding zone. Its does not depend on the distance to the sun for a disk surface density \( \Sigma \propto r^{-2} \).
30 times faster than in an identical model, in which migration and disk evolution has been switched off.

The main reason for this speed-up process is that, owing to migration, the feeding zone is not as severely depleted as in Pollack et al. (1996), and then the long time needed by the core to reach critical mass and start runaway gas accretion is suppressed.

- Taking migration into account, the moving planet always encounters new planetesimals; thus its feeding zone is never emptied.

To illustrate this important point, Fig. 2b shows the initial and final disk profiles for the gaseous and the solid component.

2.6. Formation of giant planets at closer orbital distances

The model results of Alibert et al. (2004, 2005a) also indicate that it is possible to form giant planets that resemble some of the short periodic giant exoplanets that have been discovered so far. The example shown in Fig. 3 assumes a density profile $\Sigma \propto r^{-3/2}$ normalized so that the mass of the disk between 0.5–50 AU is 0.05 $M_{\odot}$ and the disk photo-evaporation rate is $4 \times 10^{-9} M_{\odot}$/year. The other parameters are kept the same.

Figure 3a shows the evolution of the mass of the gaseous envelope, the mass of accreted planetesimals, as well as the mass of the disk for calculations starting with a planetary embryo at 15 AU. The distance to the star as a function of time is shown in Fig. 3b. The crossover mass is reached in this simulation after about 3 Myr; and, shortly after, due to gap formation, the accretion rate of gas reaches its maximum value, which decreases with decreasing disk mass (Alibert et al., 2004).

The formation process of a giant planet with a mass of $1.6 M_{\text{Jup}}$ located at about 2 AU ends after 6 Myr, when the disk disappears. The final planet is characterized by a core of about $7 M_{\text{Earth}}$ and an envelope of about $500 M_{\text{Earth}}$, which itself contains about $38 M_{\text{Earth}}$ of heavy elements. These $38 M_{\text{Earth}}$ result from the accretion of about $28 M_{\text{Earth}}$ in the form of planetesimals and about $10 M_{\text{Earth}}$ due to accreted gas with an assumed solar metallicity.

Thus, the migration of the planet can be divided into two main phases:

- Before about 2.3 Myr, the planet undergoes type I migration, at which time a gap opens and migration switches to type II.
- Shortly after about 4 Myr, the mass of the planet becomes non-negligible compared with the disk mass, and migration slows down and eventually stops when the disk has disappeared.

These model simulations show that the formation of giant planets, at least the first phase until runaway gas accretion, can be significantly sped up if the effect of migration is taken into account. The speed-up due to migration has been found to be robust against changes in various parameters (Alibert et al., 2004). The assumed size of the planetesimals plays a critical role, as already noted by Pollack et al. (1996). Assuming planetesimals with a size of about 10 km instead of about 100 km leads to runaway accretion after only about 0.3 Myr! The formation of giant planets through the core accretion scenario may, therefore, proceed over timescales in
good agreement with disk lifetimes without the need to consider disks significantly more massive than the minimum-mass solar nebula.

2.7. Hydrodynamical models for giant planet formation

Models involving migration caused by disk-planet interactions are favored by many researchers for the formation of short-period giant planets (e.g., Lin et al., 1996, 2000; Trilling et al., 1998; Ward and Hahn, 2000; Alibert et al., 2004, 2005a, 2005b, 2006; Mordasini et al., 2009a, 2009b). However, Guillot et al. (1996) showed that giant planets may also be stable over the lifetime of a solar-mass star even if they are formed as close as 0.05 AU. Simulations that apply hydrodynamical models for the origin of giant planets support the possible formation of giant planets very close to their host stars, provided enough mass is available that close to the star.

Giant planets that orbit their stars with orbital periods of a few days and distances of a few hundredths AU may form by accretion induced by a core of a few $M_{\text{Earth}}$. Detailed convective radiation hydrodynamical calculations of core-envelope growth at 0.05 AU, indeed, show gas accretion beyond 300 $M_{\text{Earth}}$ for core masses around 10 $M_{\text{Earth}}$ (see Figs. 2 and 5 in Wuchterl, 1996 and 1997, respectively). Figure 4 shows the evolution of the luminosity of a short-period giant planet during the first 100 Myr.

Hence, short-period giant planets may form in situ if sufficient mass of gas and dust is available in their feeding zones. That requires a constant replenishment of the feeding zone with matter from the outer regions of the disks.

By assuming such a replenished feeding zone, Broeg and Wuchterl (2007) showed that even planets having huge cores such as HD149026b with an inferred core mass of 67 $M_{\text{Earth}}$ and a total mass of 114 $M_{\text{Earth}}$ (see Sato et al., 2005) can be formed in situ in close proximity to the host star.

3. Terrestrial Planet Formation and Water Delivery

The primary perturbations on the Keplerian orbits of kilometer-sized and larger bodies in protoplanetary disks are mutual gravitational interactions and physical collisions (Safronov, 1969). These interactions lead to accretion and in some cases erosion and fragmentation of planetesimals. The most massive planets have the largest gravitationally enhanced collision cross sections and accrete almost everything they collide with. If the random velocities of most planetesimals remain much smaller than the escape speed of the largest bodies, these large planetary embryos grow extremely rapidly (Safronov, 1969).

![Figure 3](image1.png)

**FIG. 3.** (a) Solid line: mass of accreted planetesimals. Dashed line: mass of H/He. Heavy solid line: mass of the disk. (b) The kink around 2.3 Myr signals the change from type I to type II migration (courtesy of Y. Alibert, C. Mordasini, and W. Benz).

![Figure 4](image2.png)

**FIG. 4.** Luminosity of a short periodic giant exoplanet during the first 100 Myr according to a fluid-dynamical model. The two maxima correspond to peaks in the accretion of planetesimals and gas, respectively (courtesy of G. Wuchterl).
A few large bodies can grow much faster than the rest of the swarm in a process known as runaway accretion (We-therill and Stewart, 1989; Kokubo and Ida, 1996). Planetary embryos may accrete most of the solids within their gravitational reach, so that the runaway growth phase ends. Planetary embryos can continue to accumulate solids rapidly beyond this limit if they migrate radially relative to planetesimals as a result of interactions with the gaseous component of the disk (Tanaka and Ida, 1999). A typical planetary embryo in the terrestrial planet zone of the Solar System could have the size and mass of Mars, that is, about half the radius and a tenth of the mass of Earth.

The eccentricities of planetary embryos in the inner Solar System were subsequently pumped up by long-range mutual gravitational perturbations; collisions between these embryos eventually formed the terrestrial planets (Wetherill, 1990; Chambers and Wetherill, 1998). However, timescales for this type of growth in the outer Solar System are at least about 10^9 years (Safronov, 1969) and are longer than the lifetime of the gaseous disk (Lissauer et al., 1995).

Unless the eccentricities of the growing embryos are damped substantially, they will eject one another from the star’s orbit (Levison et al., 1998). Thus, runaway growth, possibly aided by migration (Tanaka and Ida, 1999), appears to be the way for solid planets to become sufficiently massive to accumulate substantial amounts of gas while the gaseous component of the protoplanetary disk is still present (Lissauer, 1987).

3.1. The role of the snowline and giant planets in water-delivery scenarios to terrestrial planets

From the point of view of the accretion of the terrestrial planets, there is an important difference between the accretion of material through planetesimals as opposed to planetary embryos. Because planetesimals have a small individual mass, the accretion of a significant amount of material requires collision with the growing planet of a large number of planetesimals. The large numbers involved ensure that the accretion process is governed by statistical laws. For instance, two terrestrial planets on similar orbits will accrete comparable masses from the same population of planetesimals.

The opposite is true for the mass accreted through planetary embryos. One or a few collisions with embryos are enough to deliver a large amount of mass to the growing planet. The collisional history of the embryos is therefore governed by small statistical numbers. Accretion becomes a stochastic process; and, as a consequence, two terrestrial planets on similar orbits can accrete very different amounts of mass from planetary embryos. Water can be accreted by terrestrial planets via both planetesimal and planetary embryos. The level of hydration of planetesimals and embryos depends on the heliocentric distance at which they are formed.

A threshold distance is the so-called snowline, beyond which water condenses as ice grains. There is a large uncertainty in the orbit location of the snowline in the solar nebula (Raymond et al., 2004). The standard notion of a snowline around 4–5 AU can explain the rapid formation of Jupiter in a high-density environment immediately past the snowline. However, volatile-like asteroids are also found as close as 2–2.5 AU.

Models of protoplanetary disks by Sasselov and Lecar (2000) around T Tauri stars result in snowlines as close as 1 AU to the central stars, depending on the stellar luminosity and the rate of accretional heating within the disk. After these quantities evolve with time, the snowline as shown in Fig. 5 can migrate with time to other orbital locations (Hueso and Guillot, 2003).

- Objects formed beyond the snowline have a composition similar to that of comets and thus contain a large amount of water (40–80% according to various estimates of the gas/dust ratio of comets).
- Objects formed closer to the star than the snowline should not be completely anhydrous.

Meter-sized icy snowballs formed at the snowline could drift inward by gas drag and be incorporated by growing planetesimals within 1–2 AU from the snowline (Cyr et al., 1998). Thus, only far inward from the snowline would water be very scarce.

Hueso and Guillot (2003) modeled planetesimal formation in a pragmatic approach by including important processes in planetesimal formation in a balanced way. A simple evolving nebula model is combined with radial transport of gas, condensation and evaporation, and the drift and growth of dust and planetesimals.

It was found that kilometer-sized planetesimals may form within their approximate, but synoptic, model; and an increase was found in size and number of the planetesimals somewhat interior to the snowline (Hueso and Guillot, 2003). It is believed that, at the time of planetesimal formation in the Solar System, the snowline was at about 5 AU. Comets, now stocked in the Oort Cloud, the scattered disk, and the Kuiper Belt, should all have formed beyond this limit. Carbonaceous chondrites, presumably pieces of C-type asteroids formed in the 2.5–4.5 AU region, contain 5–10% of their mass as water. Inside 2.5 AU, asteroids are much drier. Ordinary chondrites, presumably fragments of S-type asteroids formed in the 2.0–2.5 AU region, contain 0.1% of their mass as water. Enstatite chondrites, probably linked to E-type asteroids at
The correlation between water content and heliocentric distance indicates that the planetesimals in the terrestrial planet region should have carried a negligible amount of water.

Therefore, the terrestrial planets should have accreted water from farther out, from planetesimals, embryos, or both, which formed near or beyond the snowline and moved to the inner Solar System through complicated dynamical paths (mutual scattering, scattering by giant planets, resonant effects that excited their eccentricity and decreased their perihelion distance, etc.). It should be noted that, though this view has been accepted by the astronomical community, it is not unanimously shared by the community of geochemists (see, for instance, Drake and Righter, 2002).

There is currently no explanation as to why planetesimals at 1 AU should have had a higher water content than S-type asteroids at 2 AU. According to dynamical simulations, comets, which are planetesimals that formed beyond the snowline, should have carried less than 10% of the total amount of water to Earth (Morbidelli et al., 2000). This is consistent with the D/H ratio of Earth’s water being very different from that of comets. Note that a recent study by Genda and Ikoma (2008) showed that the D/H ratio of Earth’s water might be strongly influenced by a primordial hydrogen atmosphere. Mass fractionation during subsequent hydrogen mass loss could have enriched deuterium by a factor of 9. Therefore, the D/H discrepancy is not a good argument against nebula origin of water.

The reason for the low contribution of comets to Earth’s water budget, despite the large total mass of the initial comet population (presumably of the order 100 M_{Earth}), is Jupiter:

- A giant planet consumes very rapidly all the bodies that come to cross its orbit. As a consequence, the probability that a short periodic comet would hit Earth over its dynamical lifetime is about 1 in a million.

The situation could be very different in planetary systems without gas giants.

- In a system where Jupiter and Saturn have a lower mass, like Uranus or Neptune, the collision probability of comets with Earth would be orders of magnitude higher.

Therefore, for planetary systems with no giant planets in the outer system, comets should be the dominant delivery source of water to terrestrial planets in the habitable zone.

In the case of Earth, water should have been accreted, in the absence of a significant cometary contribution, through planetesimals and embryos formed in the outer asteroid belt (2.5–4.5 AU), which is consistent with Earth’s water D/H ratio being equal to that of carbonaceous chondrites. The amount of asteroids that belong to planetesimals formed within the snowline is difficult to quantify with confidence. It critically depends on the model assumed for mass depletion and primordial orbital excitation of the asteroid belt.

The model currently most successful in reproducing the observed properties of the asteroid belt is that of Petit et al. (2000), which is an elaboration of an original idea of Wetherill (1992). The basic idea assumes that planetary embryos were present in the asteroid belt and the combined action of the embryos and Jupiter forced all the embryos and more than 99% of the asteroids out of the asteroid belt.

Using this model, Morbidelli et al. (2000) estimated that the asteroids could have barely delivered the required amount of water to Earth. Furthermore, most of the asteroid contribution should have occurred at a time when Earth was still undergoing substantial growth due to giant collisions with planetary embryos, so that presumably the retention of the delivered water was inefficient. Therefore, Morbidelli et al. (2000) proposed an alternative possibility of hydrated planetary embryos.

### 3.2. Hydrated planetary embryos

In this scenario, the bulk of Earth’s water was delivered by one or a few hydrated planetary embryos. In recent simulations, two-thirds of the terrestrial planets accreted at least one embryo originally placed in the outer asteroid belt (Morbidelli et al., 2000). From the embryos’ mass, and assuming a water content similar to that of carbonaceous chondrites, Morbidelli et al. (2000) estimated that Earth could accrete some 3–6 \times 10^{25} g of water, about 10–20 times the amount of water currently on our planet. The presence of a large amount of water at some time during Earth’s formation is compatible with recent geochemical models (Abe et al., 2000) and allows an inefficient retention of the water during the violent phases of Earth’s formation.

- This model can explain why terrestrial planets can start their geochemical evolution with very different water budgets, because accretion of water from planetary embryos is a stochastic process.

Assuming that no planetary embryo hit Mars, which would be to say that the only available water was accreted from asteroids and comets, Lunine et al. (2003) estimated that Mars received only a tenth of Earth’s water, with a D/H ratio twice as large. This is compatible with the constraints obtained from analysis of martian meteorites.

The idea that water has been delivered by planetary embryos has stimulated more-recent investigations. Levison and Agnor (2003) and Raymond et al. (2004) confirmed that the accretion of embryos by terrestrial planets from distant regions is a quite generic process. It happens in a large number of giant planet configurations and even in the absence of giant planets.

- The most important parameter that inhibits the process of water delivery to terrestrial planets is the eccentricity of gas giants in the outer system (Chambers and Wetherill, 2001).

With an eccentric Jupiter, the dynamical lifetime of embryos in the outer asteroid belt is strongly reduced; consequently, their probability of hitting Earth drops proportionally.

In a recent study, Raymond et al. (2004) applied dynamical simulations of terrestrial planet formation and water delivery. Their simulations included planetary embryos, which are Moon- or Mars-sized protoplanets and planetesimals, by assuming that the embryos formed via oligarchic growth. They investigated the volatile delivery as a function of a Jupiter-class planet at orbital distances between 4 and 7 AU, the position of the snowline and the density in solids of the planetary nebula. All simulations produced 1–4 terrestrial
planets, which varied in mass and volatile content, inside an orbital distance of about 2 AU. In 44 simulations, 43 terrestrial planets between 0.8 and 1.5 AU were formed. Eleven planets thereof had their orbits inside the habitable zone between 0.9 and 1.1 AU of a Sun-like GV star.

Further, Raymond et al. (2004) suggested that terrestrial planets may also accrete icy planetary embryos, formed beyond the snowline. This is particularly plausible if the giant planets of the system (unlike Jupiter) are placed several AU beyond the snowline. If not, the icy embryos are rapidly eliminated by the giant planets before having a chance to hit a terrestrial planet. The accretion of an icy embryo delivers a substantial fraction (even several 10%) of the terrestrial planet’s mass in the form of water. This would be a mechanism for the origin of so-called ocean planets (Léger et al., 2004), which involves formation of an icy planet beyond the snowline and the planet’s migration to the habitable zone by tidal interactions with the disk.

It seems from these studies that terrestrial planet formation, or the process by which Moon- to Mars-sized bodies are cleaned up to form a few Earth-sized planets, is stochastic and leads to planets that are similar to the terrestrial planets in the Solar System in terms of size and orbital location.

These models generate a high abundance of terrestrial planets around other stars; hence terrestrial-finding missions like Darwin ought to assume a high likelihood of terrestrial planets around solar-type stars for which giant planets are not so close that they induce orbital instability.

3.3. Internal structures of terrestrial and icy planets

Comparative studies between Earth, Venus, Mars, and the large satellites of the giant planets in the Solar System can be used for model simulations of the internal structure of terrestrial exoplanets (Léger et al., 2004; Sotin et al., 2007; Grasset et al., 2009). The input parameters are the relative abundances of Mg and Fe relative to Si, the total mass of the planet, the amount of iron in the silicates (Mg\(^{\#}\)), and the total water amount.

The main elements that compose terrestrial planets are O, Si, Fe, and Mg. Other elements such as Ca, Al, S, Na, are much less abundant and can be replaced by their major counterparts in silicate phases (i.e., Mg and Fe). In addition, water may be included in the case of water-rich planets.

For the interior of terrestrial planets, it is valid to assume iron-rich cores, covered by silicate mantles, and eventually icy layers. The metallic cores are thought to be liquid and composed of an iron-sulfur alloy. Although the presence of an inner solid iron core is possible, the difference in mass would be small (on Earth, the mass of the inner core is less than 1.5% of the total planetary mass). The silicate mantle is composed of two layers: below 23 GPa, it is made of olivine (Mg, Fe\(_2\))SiO\(_4\) and pyroxene (Mg, Fe\(_2\))Si\(_2\)O\(_6\), while above 23 GPa it is composed of perovskite (Mg, Fe)SiO\(_3\) and magnesiowüstite (Mg, Fe)O. If an icy layer is considered, it is assumed to be made of pure water.

The size of the metallic core depends mostly on the Fe/Si ratio and the Mg\(^{\#}\) number (Mg\(^{\#}\) = Mg/[Mg + Fe]). Once the Fe/Si ratio is fixed, to be that of the star, the size of the core is determined by Mg\(^{\#}\) because it fixes the amount of iron relative to Magnesium in silicate phases: the larger the Mg number, the larger the core. The Mg/Si ratio determines the relative amount of pyroxene and olivine (perovskite and magnesiowüstite) in the lower and upper mantle. Finally, the water amount imposes the thickness of the hydrosphere. For a given planetary mass, elementary ratio of the host star, and amount of water, one can determine the radius of both terrestrial and icy planets by computing the density distribution as a function of depth.

An internal structure is provided once a good estimate of the density profile within the planet is known. At each depth, the pressure is estimated from the hydrostatic equilibrium, and temperature is fixed by adiabatic profiles. Then, a different equation of state can be used for each layer in order to compute the density at each depth. By an iterative process, it is then possible to get the unique internal structure with respect to the input parameters (Sotin et al., 2007; Grasset et al., 2009). An adiabatic gradient is assumed, which means that heat transfer is achieved by convection in each layer. When applied to Earth, both the density profile and the pressure profile are almost similar to the preliminary reference Earth model (Dziewonski and Anderson, 1981). In addition, the model computes a radius for Earth of about 6326 km, rather than 6371 km, which is less than 1% error (Sotin et al., 2007).

Figure 6 shows the relationship between planetary radius and mass as a function of the amount of water. In the moderate mass range between 1 and 10\(M_\text{Earth}\) and for terrestrial planets only (without water), this corresponds to the power-law \(R/6326 = (M/M_\text{Earth})^{0.288}\). The value of 0.288 is significantly lower than the value of \([IT]\) that would be obtained if compressibility is not taken into account. For larger masses, as for different amounts of water, a more general scaling law has been proposed in Grasset et al. (2009).

- One key prediction of the internal structure models of Sotin et al. (2007) or Grasset et al. (2009), the mass-radius relation, can be tested by accurate radius determinations of super-Earths during transits (observed, e.g., by the CoRoT and Kepler missions) and ground-based follow-up mass determinations.

It can be seen in Fig. 6 that a terrestrial exoplanet with 10\(M_\text{Earth}\) which could be the upper limit for Earth-like planets, would have a radius less than twice that of Earth. Figure 7 shows the mean density and the surface gravi-

![FIG. 6](image-url) Radius of an Earth-like or icy planet versus its mass. As discussed in the text, the main elements used in the calculations are Fe/Si, Mg/Si, Mg\(^{\#}\), and the water amount (small numbers on the curves in wt %) (Sotin et al., 2007; Grasset et al., 2009).
tional acceleration of terrestrial planets as a function of planetary mass. Figure 7 shows that the gravitational acceleration for a terrestrial planet with 10 $M_{\text{Earth}}$, compared to Earth increases by almost a factor of 3. This implies that, if the planetary components are identical to Earth, the atmospheric altitude can be divided by a factor 3.

Observed masses of some terrestrial exoplanets at distances < 10 pc by the forthcoming SIM planetary quest mission along with their planetary radii observed by future space missions like Darwin or TPF-I can be used to study aspects of planetary interiors, such as the geodynamic simulation of core sizes.

Accurate assessment of core sizes of terrestrial exoplanets can be used to estimate the strength of intrinsic magnetic moments, which are important for the formation of atmosphere-protecting magnetospheres.

4. Planet Formation and Its Implications for Terrestrial Planet–Finding Missions

In the standard planetesimal model, all planets form via an intermediate stage of terrestrial (or solid) planets. Giant planets form if, and when, sufficient amounts of gas and solids are available in the protoplanetary nebula, which allows for further growth. Alternatively, travel of protoplanets over significant distances from their formation places, driven by violent migration or planet formation via disk instability, should be considered.

If planets form by the disk instability mechanism, planet formation does not necessarily involve planetesimals. In this case, terrestrial planets might not exist, except in rare cases like the Solar System.

If planetesimals are needed during planet formation to produce terrestrial planets, while gas giants form quickly by way of disk instability and migrate significantly, then terrestrial planetary embryos would likely be swept up by migrating gas giants. Survival of terrestrial planets would require careful timing of migration and disk dissipation, with the likely result of a low terrestrial planetary yield. In the case of migration type II, major implications for terrestrial planet formation appear:

- First: terrestrial planets may only form if the timescale for migration of giant planets is longer than the lifetime of the gas in the disk.

The migration velocity of low-mass planets is proportional to the size of the protoplanets (e.g., Tanaka et al., 2002), and an Earth-like planet can only form if the final stages of its accretion took place after dissipation of the nebula. Simulations performed by Tanaka and Ida (1999) showed that, in a minimum-mass nebula, a protoplanet with 1 Mars mass at 1 AU could survive migration. Final growth to larger planetary masses may be achieved after the nebula dissipation through the gravitational scattering and merging of residual planetesimals.

- Second: the migration scenario would lead to an accumulation of planetary building blocks orbiting at close orbital distances from the star.

If migration were to stop within close vicinity of the star, truncation of the disk due to the stellar magnetic field or direct tidal interaction with the star (Lin et al., 1996) could occur. This would provide a considerable amount of material to build a giant planet at very close orbital distances.

If planets form through planetesimals, but the time window to nebula dissipation is narrow, the survival of terrestrial planets would depend on the ratio of planet losses due to violent migration3 to the timescale of nebula dispersal.

---

Footnote:

3Violent migration: > factor 10 change in semimajor axis.
Terrestrial planet formation would then depend on disk lifetimes that could depend on the star-birth environment in the original cluster via photoevaporation or disk-star interactions during stellar encounters.

If the standard model of formation via planetesimals is valid, the terrestrial planet yield will depend on the disk-surface density with a higher giant/terrestrial planet ratio for more-massive disks with higher surface densities. Properties of the distribution of protoplanetary disk masses would then determine the frequency of terrestrial planets. If, for example, the Solar System turns out to be in the tail of this distribution, terrestrial planets may be rare. Giant planets would likely require more massive disks; hence, systems with giant and terrestrial planets would be less frequent if the required disk masses were below the peak of the disk distribution.

- In consequence, it is important to characterize the potential host systems for terrestrial planet-finding missions like Darwin and provide a target sample that is likely to bracket the diversity of planetary systems to contain a sufficient number of terrestrial planets.

Abbreviations

CEGP, close-in extrasolar giant planets; TPF-C, the Terrestrial Planet Finder Coronograph; TPF-I, the Terrestrial Planet Finder Interferometer.

References


Address correspondence to:
Yann Alibert
Observatoire de Besançon
Institut UTINAM
Besançon
France
E-mail: alibert@obs-besancon.fr