Oxford Research Encyclopedia of Planetary Science

Planet Formation

Morris Podolak

Subject: Planet Formation Online Publication Date: Apr 2018
DOI: 10.1093/acrefore/9780190647926.013.108

Summary and Keywords

Modern observational techniques are still not powerful enough to directly view planet formation, and so it is necessary to rely on theory. However, observations do give two important clues to the formation process. The first is that the most primitive form of material in interstellar space exists as a dilute gas. Some of this gas is unstable against gravitational collapse, and begins to contract. Because the angular momentum of the gas is not zero, it contracts along the spin axis, but remains extended in the plane perpendicular to that axis, so that a disk is formed. Viscous processes in the disk carry most of the mass into the center where a star eventually forms. In the process, almost as a by-product, a planetary system is formed as well.

The second clue is the time required. Young stars are indeed observed to have gas disks, composed mostly of hydrogen and helium, surrounding them, and observations tell us that these disks dissipate after about 5 to 10 million years. If planets like Jupiter and Saturn, which are very rich in hydrogen and helium, are to form in such a disk, they must accrete their gas within 5 million years of the time of the formation of the disk. Any formation scenario one proposes must produce Jupiter in that time, although the terrestrial planets, which don't contain significant amounts of hydrogen and helium, could have taken longer to build. Modern estimates for the formation time of the Earth are of the order of 100 million years.

To date there are two main candidate theories for producing Jupiter-like planets. The core accretion (CA) scenario supposes that any solid materials in the disk slowly coagulate into protoplanetary cores with progressively larger masses. If the core remains small enough it won't have a strong enough gravitational force to attract gas from the surrounding disk, and the result will be a terrestrial planet. If the core grows large enough (of the order of ten Earth masses), and the disk has not yet dissipated, then the planetary embryo can attract gas from the surrounding disk and grow to be a gas giant. If the disk

dissipates before the process is complete, the result will be an object like Uranus or Neptune, which has a small, but significant, complement of hydrogen and helium. The main question is whether the protoplanetary core can grow large enough before the disk dissipates.

A second scenario is the disk instability (DI) scenario. This scenario posits that the disk itself is unstable and tends to develop regions of higher than normal density. Such regions collapse under their own gravity to form Jupiter-mass protoplanets. In the DI scenario a Jupiter-mass clump of gas can form—in several hundred years which will eventually contract into a gas giant planet. The difficulty here is to bring the disk to a condition where such instabilities will form. Now that we have discovered nearly 3000 planetary systems, there will be numerous examples against which to test these scenarios.

Keywords: planet formation, disk instability, core accretion, pebble accretion, cosmogony

Our solar system contains three basic types of planets. Those closer to the Sun—the *terrestrial planets* Mercury, Venus, Earth, and Mars—are all roughly earthlike. None are larger than the Earth, and all consist of a core composed mostly of iron, a mantle composed of rock (mostly silicates), and an atmosphere whose mass is a negligible fraction of the total mass of the planet. Beyond Mars lie the *gas giants* Jupiter and Saturn. These planets are, respectively, some 300 and 100 times more massive than the Earth, and are mostly composed of a mixture of hydrogen and helium in the same proportion as is found in the Sun. The other elements are enriched over their solar proportions by a factor of a few, and although their contribution to the total mass is not negligible, they remain a minor component. Finally, at the outer edge of the solar system are the so-called *ice giants*, Uranus and Neptune. These planets are roughly 15 times more massive than the Earth, and are composed of roughly equal parts each of rock, ice, and a hydrogen-helium mix.

For many years it has been a major challenge to understand the connection between these different planetary classes and their respective places in the solar system. The challenge has become more acute with the discovery of planetary systems around other stars. It is now known that our planetary system is but one example among many, and the variation among these systems is very broad. What processes determine the mass of a planet and its composition? How are these processes related to the planet's distance from the host star? These questions are at the heart of the study of planet formation.

In principle, the way to answer these questions is to start at the birthplace of stars, the interstellar clouds, and apply the known laws of physics. This should indicate how the materials arrange themselves. Unfortunately, many different laws have to be applied simultaneously. In the first place, the cloud will have some random, nonisotropic, shape. This means that the force of gravity will be different in different parts of the cloud. This will lead to pressure gradients, and gas flow. But in addition, the gas will contain solid condensate grains. If they are small enough they will be dragged along by the gas. If they are larger, they will tend to resist the gas drag and move in response to the gravitational forces that are applied by the surrounding material. The sizes of the grains will depend on how often they collide and on whether they stick or fragment upon colliding. This, in turn, will depend on the speed of the collision. The composition of the grains will depend on the initial composition of the cloud, and the temperatures and pressures within. But since the grains control the opacity in the cloud, they themselves affect the cloud temperatures. If magnetic fields are present, they further complicate the motion of the gas.

Ultimately the evolution of a region of the gas cloud into a stellar system requires the solution of a complex system of equations drawn from thermodynamics, hydrodynamics, radiative transfer, plasma physics, chemistry, and mechanics. Solving the full set of equations is beyond the scope of current computing power, and so approximations must be made. The earliest approximations concentrated on one or another of the many processes involved, and viewed the rest as some sort of background. In this way it was possible to learn something about how each relevant process behaved in the context of planet formation. As our understanding of the physics grew, and computing power increased, it became possible to model more complex processes, and we are now gaining a better understanding of how the various components fit together to make a planet, and how they influence the final appearance of the body.

Early Theories

The earliest observations of the planets showed that all the known planets of our solar system orbit the Sun in the same direction. In addition, these orbits all lie in nearly the same plane. This circular motion impressed those speculating on planetary origins. If we assume a simple system, say a universe filled with dilute material (a gas?), how is it possible to derive circular motion from it? The very earliest ideas included the suggestion by Rene Descartes, in 1633, that the universe was originally full of vortices, and that some of the material in these vortices somehow contracted to form the planets. This would explain why the planets all move in the same direction around the Sun. In 1749 Georges-Louis Leclerc, Comte de Buffon suggested that the planets were formed from the debris scattered when a giant comet collided with the Sun. Here too the planets would all be expected to orbit the Sun in the same direction. Although both these theories have serious difficulties, variations of them remained popular until well into the 20th century.

Today, however, the prevailing theory is based on the nebular hypothesis proposed by Emanuel Swedenborg in 1734. This idea was further developed by Immanuel Kant in 1755, and independently by Pierre-Simon Laplace in 1796. Laplace's picture of an extended gaseous “atmosphere” surrounding the Sun and filling the volume of what is now the solar system is very much in line with modern views. Modern astrophysicists would argue as follows: Because more than 99 percent of the mass of the solar system lies in the Sun, it seems reasonable to view the planets as being the “leftovers” from the process of star formation. The environment of the newly forming Sun will give us the background conditions for the formation of planets. Thus understanding the context in which planets form requires an understanding of star formation.

Disk Formation

Spiral galaxies are observed to include large clouds of interstellar gas. Such clouds can contain between a hundred thousand and several million solar masses of material. This material, mostly gas, is subject to a number of influences, including magnetic fields and stellar radiation pressure, but the two major factors are gravity, which causes the gas to collapse, and the kinetic motion of the gas molecules which causes the gas to expand. If a perturbation is applied to cause the gas to contract, the self-gravity will increase and favor further contraction, but at the same time the gas will heat up and try to expand, so there is a balance between the tendency to contract and the tendency to expand. For low masses this results in a stable situation. The increase in self-gravity is offset by the increase in temperature, and the gas clump remains stable. But if a particular region contains enough mass, a small contraction will increase the self-gravity enough so that it will overcome the associated heating, and the region will continue to contract. This mechanism, called the *Jeans instability*, is the first stage of star formation, and was described by Sir James Jeans in 1902.

The details of the collapse of such an object are complex. They involve calculating the hydrodynamic response of the gas to the forces of gravity, radiation pressure, magnetic fields, thermal gradients, and rotation. However the general behavior can be understood by imagining the gravitational contraction of a rotating sphere. The gravitational force will be radial, inward, and equal in all directions, but the centrifugal force due to rotation will be stronger near the equator than near the poles. Because of this the contraction along the spin axis will be greater than the contraction in the equatorial plane and a disk-like structure will result. Such disks are indeed observed surrounding young stars.

Redistribution of Angular Momentum

Since the gravitational force in the disk varies with distance from the center, regions of the disk that are at different distances from the center will rotate at different angular speeds. The gas will also be turbulent. There are a number of possible sources for this turbulence. One mechanism involves the magnetic field. If the gas is ionized, magnetic field lines will tend to be tied to the gas, and since different regions of the gas move at different speeds, these lines will be stretched and twisted. The forces that result can cause turbulence in the gas (Balbus & Hawley, 1991). Other sources of turbulence come from instabilities due to the fact that the pressure and density gradients are not necessarily in the same direction (*baroclinic instability*) or due to strong thermal gradients. This turbulence gives the gas a much higher viscosity than would be expected from simple gas kinetic considerations (Shakura & Sunyaev, 1973).

The high viscosity creates a coupling between adjacent regions in the disk. Typically the inner parts of the disk move faster and this viscous coupling between adjacent streamlines of gas will transfer angular momentum from the inner, faster streamlines, to the slower outer ones. The net result is to transfer angular momentum outward in the disk. As the material of the inner disk loses angular momentum, it falls toward the center of the disk, and a central mass—the protostar—is formed. The outlying material, out of which the planets will eventually form, will contain most of the angular momentum. This is consistent with the fact that more than 95 percent of the solar system's angular momentum is in the planets despite the fact that they contain less than 1 percent of the total mass.

Chemical Composition

The composition of the gas depends on its age. The oldest gas, which dates from the time of the Big Bang, is a mixture of approximately 75 percent by mass hydrogen, 25 percent helium, and a trace of heavier elements. Most of the elements other than hydrogen and helium were produced by stellar nucleosynthesis and mixed back into the background gas. Our sun is a second-generation star and was formed in an environment with a composition given by Table 1. The table shows the abundance, by number, of the most common elements in our solar system relative to 10^6 silicon atoms.

The materials that can be formed from these elements can be divided into three different classes depending on their volatility. Some materials will be gases under almost all conditions of pressure and temperature relevant to the disk. These materials will be referred to as *gases*. Other materials will be solid for all but the very highest temperatures in the disk, and they will be referred to as *rocks*. Finally, there will be materials of intermediate volatility. For some conditions in the disk they will be solid and for others they will be gaseous. We will refer to such intermediate materials as *ices*. Note that these terms refer to the class of material, not to its actual phase.

Elements such as helium, neon, and argon are chemically inert, and will be gases under all relevant conditions. Since hydrogen is much more abundant than all the other elements, most of the hydrogen will simply combine with itself to form H_2 and will also be a gas. Elements such as silicon and other heavier elements will tend to form refractory materials such as Fe_2SiO_3, $MgSiO_3$, Al_2O_3, etc., and will be rocks. The remaining oxygen, carbon, and nitrogen will mostly encounter hydrogen atoms, and will form compounds such as H_2O, CH_4, and NH_3. These molecules, in turn, form ices and will be either solid or gas depending on the local pressure and temperature. For typical pressures in the disk, H_2O will freeze at around 170K, NH_3 at around 110K, and CH_4 at around 40K. Here it is important to point out two caveats:

The first is that although pure CH_4 will only freeze at extremely low temperatures, it may be trapped by condensing water molecules as an occluded ice (Bar-Nun et al., 2013). Thus it may be trapped in solids in a temperature regime where it is too warm for CH_4 to freeze. The disk can thus be viewed as a being composed of mostly H_2 and He gas with a small admixture of solid grains. The composition of these grains will vary somewhat with distance from the Sun due to changes in composition as a function of temperature, but the major composition change will occur when the temperature drops low enough so that the different ices, in particular H_2O, can freeze out. As can be seen from Table 1, oxygen is the most abundant element after helium, so that the mass of solids available after water ice condensation is significantly greater than the mass available before condensation. The place where the temperature first drops low enough to allow a particular ice to condense is called the *ice line* for that particular molecule.

Table 1. Abundances of the Most Common Elements in the Solar System. Shown is the number of atoms relative to 10^{12} hydrogen atoms. The numbers are taken from Lodders (2010)

Element	**Number Per 10^{12} H Atoms**
H	1.00×10^{12}
He	8.41×10^{10}
C	2.46×10^{8}
N	7.24×10^{7}
O	5.37×10^{8}
Ne	1.12×10^{8}
Na	1.95×10^{6}
Mg	3.47×10^{7}
Al	2.88×10^{6}
Si	3.39×10^{7}

S	1.45×10^{7}
Ar	3.16×10^{6}
Ca	2.04×10^{6}
Cr	4.47×10^{5}
Fe	2.88×10^{7}
Ni	1.66×10^{6}

The second caveat is that under certain circumstances carbon will tend to combine with oxygen rather than the more abundant hydrogen. Ordinarily since encounters between atoms in a gas are governed by statistics, and hydrogen is 2000 times more abundant than oxygen, a carbon atom is 2000 times more likely to encounter a hydrogen atom than an oxygen atom. However, the bond between a carbon and an oxygen atom is much stronger than the bond between a carbon and a hydrogen atom. If the temperature is low, the collisions are weak enough so that a C-H bond, once formed, will generally not be broken. In this case most of the oxygen atoms are free to combine with hydrogen to form H_2O, and crossing the water ice line will increase the mass of solid material by more than a factor of 2.

However, if the temperatures are high enough, even though C-H bonds are formed, the molecules will be moving rapidly enough so that intermolecular collisions will break these bonds. C-O bonds, are much stronger, and, once formed, will be much harder to break. As a result, in the solar nebula, at temperatures higher than about 1000 K, CO, rather than CH_4 will be the preferred carbon compound. If the gas flows into the hot region CO is formed. If it then flows back to the cold region, the CO should eventually revert to CH_4 and H_2O at the lower temperatures of the snow line (170 K). However this reaction is kinetically inhibited, and the CO will remain stable for periods of extremely long duration (Lewis & Prinn, 1980). Since carbon is only slightly less abundant than oxygen, if most of the carbon is in the form of CO, this will leave very little oxygen available for H_2O formation. In this case, crossing the water ice line will increase the mass of solids by only about 50 percent.

Fate of the Gas

The gas disk, with the growing protostar at its center, is a complex environment. Radiation from surrounding stars causes the outer layers of the gas to ionize. This ionized gas can trap ambient magnetic fields, and these magnetic fields can induce turbulent motions in the gas. This turbulence gives the gas a high effective viscosity and allows for a relatively rapid transfer of angular momentum within the gas. This causes the gas to eventually disappear by falling into the star. Another mechanism for gas loss is photoevaporation, whereby the UV radiation from the central star or nearby stars external to the disk ionizes the gas and heats it so that the gas can escape. In addition gas can be carried out of the system by a strong stellar wind. As a result, such disks have a relatively short lifetime. Early studies of the statistics of gas disks in young star clusters indicated that the vicinity of the star was cleared of gas on a timescale of roughly five million years. More recent work indicates that some disks might persist for times as long as 20 million years (Mamajek et al., 2009). In any case, this sets some time constraints on the timescale for planet formation. Gas giant planets such as Jupiter and Saturn, which contain large amounts of hydrogen and helium, must have formed well before the gas dissipated.

Fate of the Solid Material

In addition to the gas, the disk will have materials in the solid phase. Originally these materials will be similar to the grains seen in interstellar clouds. These interstellar grains are typically on the order of 0.1–1 microns in radius and are spread over the entire volume of the disk. The gravitational force of the central star on a typical grain can be divided into two components. The component in the radial direction is offset by the centrifugal force of the grain’s orbital motion around the star. The component in the vertical direction causes the grain to drift toward the midplane of the disk.

The speed with which the grains fall toward midplane is determined by the balance between the vertical component of the gravitational force of the star and the drag force of the gas on the grain. Small random motions of the grains cause them to collide, stick, and grow. In addition, the drift speed of grains toward midplane is larger for larger grains, so that these overtake the smaller ones on their way down and grow. These growth processes, can significantly decrease the time required for the grains to drift to the region of the midplane. After about 10^4–10^5 years the grains have grown to a size of millimeters and have formed a dust layer in the vicinity of the midplane.

The dust layer shields this region from ionizing radiation so that the gas is neutral and the magnetic field does not cause turbulence in the gas. Nonetheless, other sources of turbulence are expected to be present. If the thermal gradient in the vertical direction is large enough, that alone will be sufficient to stir the gas. In addition, with so much of the dust in a thin layer, the ratio of dust mass to gas mass is high. This leads to an interesting situation. The gas pressure in the disk decreases as one moves away from the central star. This pressure gradient applies a force to the gas that opposes the gravitational force of the star. As a result, although the dust orbits at the Keplerian velocity, the gas is supported, in part, by the pressure gradient in the disk and orbits more slowly. When the dust to gas mass ratio is high enough, the dust will drag the gas and speed it up to close to the Keplerian velocity. But in the layers just above and just below the midplane, where there is relatively little dust, the gas will move more slowly. This shear leads to an instability which is another source of turbulence. Turbulence in the gas will stir the dust up to higher levels above the midplane, so that this dust layer will have a finite thickness. At some distance r from the center of the disk the typical thickness of the dust layer will be approximately $0.1r$.

If the grains could continue to grow, they would be less influenced by the gas turbulence, and the thickness of the dust layer would remain small. The motion of the grains can be viewed as being composed of two parts: a basic circular orbit around the star and an additional random component. This random component gives the orbit its nonzero eccentricity and inclination. The thickness of the dust layer is a measure of the inclination of the dust orbits, and hence of their random motion. The collection of grains can also be viewed as a “gas” of particles, and the random motion of the grains can be associated with the temperature of this gas. Just as a clump of gas which is sufficiently massive and sufficiently cool will be subject to the Jeans instability, so too this “gas” of grains, if it is sufficiently cool, (i.e., if the dust layer is sufficiently thin) will tend to collapse into clumps of dust. This mechanism, known as the *Goldreich-Ward instability* (Goldreich & Ward, 1973), will produce kilometer-sized planetesimals. This Goldreich-Ward mechanism allows us to efficiently jump from millimeter or centimeter-sized grains, whose motion is strongly coupled to the motion of the gas, to kilometer-sized planetesimals which are much less affected by gas drag, and, in addition, have enough of a gravitational attraction to enhance their further growth.

The difficulty with this idea is the turbulence of the gas. Turbulent eddies will cause the small grains to collide at high enough speed so that they will tend to fragment rather than grow. The grains will not be able to settle into the midplane, the dust layer will remain thick (i.e., warm) and the Goldreich-Ward instability will be inhibited. This is a major barrier to further growth. A second barrier to growth stems from the pressure gradient mentioned above which causes the gas to orbit at less than the Keplerian velocity. If the turbulence keeps the dust distributed over a relatively thick layer, the ratio of gas to dust will be higher, and it will be the gas that drags the dust. As a result, the dust particles feel a headwind and an accompanying drag force that causes the particles to lose angular momentum and drift in toward the star. This effect is strongest for meter-sized bodies, and they tend to fall into the star in less than 10^5 yr. If these bodies cannot grow quickly enough, they will be destroyed before they can become large enough to decouple entirely from the gas (Weidenschilling, 1977). The problem of how cm-sized grains overcome this barrier to become km-sized planetesimals is one of the major unsolved problems in planet formation.

One popular idea for overcoming this barrier is the *streaming instability* (Youdin & Goodman, 2005). Suppose there is a temporary region of high pressure in the disk. This will lead to a change in the overall pressure gradient, and the gas will orbit more quickly at the inner edge, where the pressure gradient is smaller, and more slowly at the outer edge. Solids in this region will feel a smaller drag force at the inner edge and will slow their drift inward. This will lead to buildup of material at the inner edge. As the amount of solid material builds up, it has a back-effect on the gas, causing the gas to speed up and reducing the net drag and inward drift of the solids. The amount of solid material continues to build up in this region and filaments are formed. If the density of solids is high enough, these filaments can break up into gravitationally bound planetesimals.

Once the planetesimal reaches a size where its self-gravity becomes important, its capture cross section becomes larger than its geometric cross section. The gravitational enhancement of the capture cross section will depend on the relative speed of the encounter. If the planetesimals move quickly the effect of gravity will be to change the orbits slightly, but if the planetesimals move slowly enough, they will be attracted even at large separations. What, then, is the speed of encounters between planetesimals? This regime of planetary growth was investigated in great detail by Safronov (1972). As the planetesimals orbit the star, they perturb each other's motion via gravitational interaction. From the point of view of the largest of these planetesimals, which we can call a *protoplanetary embryo*, the others move with respect to it with some random velocity. If we assume that these random velocities are the result of gravitational interactions with this embryo, we would expect them to be related to the escape velocity from the embryo. Safronov argued that the random encounter speed was such that the kinetic energy involved was some 10 percent of the gravitational potential energy at the surface of the embryo. Half the ratio of the gravitational potential at the surface of the embryo to the kinetic energy per unit mass of the random motion of a planetesimal is often referred to as the *Safronov parameter*.

Safronov then calculated how much material the embryo would accrete in its passage around the Sun. He found that the time for the embryo to grow to its final size was proportional to

$$\tau = \left(\frac{M_{planet}\rho^2_{planet}}{36\pi}\right)^{1/3} \frac{P}{(1+2\Theta)\sum}$$

(1)

where M_{planet} is the final mass of the planet, ρ_{planet} is it's mean density, P is the period of revolution about the Sun, Θ is the *Safronov parameter*, and Σ is the surface density of solids in that region of the disk. This raises the question of what the disk looked like at that time. One way to estimate this is to assume that the disk contained just enough material to produce the planets we see today. For example, the Earth is composed mostly of iron and silicates. According to Table 1 these materials comprise only about 0.5 percent of the disk mass. So that while the Earth was being formed, in addition to the Earth mass of iron and silicates there must have been an additional 200 Earth masses of hydrogen and helium present. Similarly, by adding the missing hydrogen and helium to the observed masses of the other planets, we can reconstruct a *minimum mass solar nebula*. The exact values depend on our estimates for the composition of Jupiter and Saturn, but generally the minimum mass solar nebula is believed to contain about 0.02 solar masses of material with the surface density varying as $\sum \propto r^{-3/2}$. Thus we can use Eq. 1 to estimate the time it takes to form a planet.

Safronov found that the time it takes to build up an Earth-mass object at the Earth's distance from the Sun is around a million years. But the time it takes to form Jupiter is around 100 million years. However, since Jupiter is mostly composed of hydrogen and helium, which must have come from the gas disk, and these disks dissipate after some 5 to 10 million years, we have a timescale problem. The problem is even worse for Neptune. Here the formation time is more than twice the age of the solar system—an impossible situation! The timescale problem for this accretion process will be further discussed below.

Disk Instability Scenario

The difficulty of forming the outer planets in a reasonable time led to the suggestion of a different formation mechanism altogether. If the disk formed via the Jeans instability, and planetesimals might have formed via the Goldreich-Ward instability, perhaps planets could also be formed via some instability in the gas disk? Similar physical principles would apply. If some region of the disk would have a gas density that was high enough and a temperature that was low enough, a slight contraction would increase the gravitational binding more than it would increase the thermal excitation, and a gravitational collapse would ensue. In rotating disks, however, there is an additional mechanism that helps prevent collapse. The gas is in differential rotation around the center, and gas closer to the center rotates more quickly. As a result, any large clump of gas experiences shear stress as its inner parts move faster than the outer parts. This shear tends to tear apart the clump and prevents its collapse. The competition between the stabilizing and de-stabilizing forces is embodied in the *Toomre parameter* defined by

$$Q = \frac{c_s \kappa}{\pi G \sum}$$

(2)

Here c_s is the speed of sound in the gas, which is related to its temperature; κ is the so-called *epicyclic frequency*, and is essentially the rotation frequency of the disk; G is

Newton's gravitational constant; and Σ is again the surface density of gas in the disk. Higher temperature (larger c_s) and faster rotation (larger κ) tend to prevent collapse, while a higher surface density of gas favors collapse. The lower the value of Q, the less stable that region of the disk is, and gravitational collapse is expected for $Q < 1$.

As can be seen from Eq. 2, lower temperatures and higher gas densities both help to induce instability. Unfortunately, the required conditions are difficult to attain in standard disk models. Near the star the temperature is generally too high and the rotational shear is too strong for the instability to occur. However, numerical models of disk evolution show that far from the star conditions can be found where $Q < 1$ and an instability can occur. This is especially true if the disk is strongly perturbed, as might happen if there was a nearby supernova explosion. Recent meteoritical analysis of short-lived radioactive isotopes supports such a possibility for our own solar system (Boss, 2017). This process has been studied by a number of authors. A good review is given by Durisen et al. (2007). Current computer simulations do not have the required numerical resolution to follow the evolution of a clump from its formation in the disk to its eventual contraction into a planetary-sized object. However different stages of the process can be studied in some detail.

After the instability forms a clump of gas in the outer regions of the disk, this clump begins to slowly contract. This contraction compresses the gas which causes it to heat up. Some of this heat is radiated into space which allows for further contraction. This quasi-static contraction continues for several hundred thousand years. Eventually the center of the clump gets hot enough so that the hydrogen gas begins to dissociate. At this point the energy released by contraction is not only radiated into space, but is also channeled into breaking the molecular bonds of the hydrogen. This new sink for the gravitational energy released allows the rate of contraction to change from quasi-static to a hydrodynamic collapse and the clump quickly goes from a radius of some tens of millions of kilometers to a protoplanet not much larger than the present day Jupiter. Once the gravitational instability, the clump, forms, the entire process of contraction and collapse to a protoplanet happens quickly. Typical times are less than a million years, and so the timescale difficulties described earlier do not appear.

There are, however, other difficulties. In the first place, the body formed should have the same composition as the nebular gas itself: about 74 percent hydrogen by mass, 25 percent helium, and about 1 percent of heavier materials. These heavier materials are collectively referred to as *metals*. How does this compare with Jupiter itself? The knowledge we have of the structure and composition of Jupiter's interior comes from numerical models. These models use information about the behavior of materials at high pressure (*equation of state*), and a description of how heat is transferred inside the body to compute the mass and radius of a body with a particular composition. If the body is rotating, the centrifugal forces make it oblate, and this oblateness, or, more precisely the nonsphericity of its gravitational field, contains additional information about the interior structure and composition. By varying the amount and distribution of the different materials one can try to vary the input parameters of the model in order to reproduce the observed mass, radius, and shape of the planet. In this way we can infer the structure and composition of Jupiter.

It turns out that the model results are rather sensitive to the assumed equation of the state of hydrogen. Because of uncertainties in this equation of that state, different investigators find different values for the composition. Currently the best estimates require that between 8 and 15 percent of Jupiter's mass be comprised of metals (Wahl et al., 2017). This is much higher than the expected composition of the nebular gas. For Saturn the enhancement of metals is even higher. In addition, both Jupiter and Saturn have envelopes that are significantly enriched in heavy elements with respect to solar composition. Clearly a simple contraction of the disk gas is insufficient to explain these planets.

Very often the solution to difficulties of this sort is to simply add more physics. In this case, in addition to the contraction of the local gas, one also has to take into account the accretion of planetesimals from outside the clump. As the planetesimals orbit the Sun, these orbits are perturbed by the clump and any additional large masses present. If these perturbed orbits intersect the clump, the planetesimals encounter the resistance of the denser, contracting gas, and the resulting gas drag slows their motion. If the motion is slowed sufficiently, the planetesimals are captured, and the overall metal content of the clump is increased. In this way the metal content of Jupiter can be increased to be consistent with the values deduced from interior models.

A second difficulty is that these unstable clumps only form in the region of the disk that is far from the Sun (~10 AU or more for the case of our solar system) where the temperature is low and the rotational shear is weaker. It is unlikely that Jupiter is far enough away from the Sun for such an instability to occur, though it is possible that it formed further out and migrated to its present position (see below). Furthermore, Uranus and Neptune, which are even further from the Sun, should also be gas giants like Jupiter and Saturn. But both Uranus and Neptune are only about 20 percent hydrogen-helium, and their overall masses are too small to be the result of a simple disk instability. Yet it is precisely for these two planets that the timescale problem is most acute. Can these bodies be accommodated in a disk instability scenario?

Again the solution may lie in the addition of more physics. The solar system was not formed in isolation. It is very likely that additional young stars were present in the solar system's vicinity at the time of its formation. Indeed, as we have seen, a nearby supernova may have initiated the formation of the solar nebula in the first place. Suppose that Uranus and Neptune formed via the disk instability mechanism. They would be more massive and more hydrogen-rich than they are today. However, since they lie near the outer edge for the solar system, they would be more exposed to the intense radiation from nearby young stars. This radiation could lead to a photoevaporation of the outer layers of the planet. Much of the hydrogen and helium could be lost in this way. The result could well be Uranus and Neptune as we see them today. However, terrestrial-type rocky planets cannot be formed via this mechanism.

Core Accretion Scenario

The inherent difficulties of planet formation via the disk instability mechanism have led researchers to look for ways of overcoming the timescale problems of the accretion mechanism. The timescale for forming the outer planets can be shortened considerably once it is realized that Safronov's calculation had two serious flaws. First, consider the random encounter speeds. Safronov had argued that these must be comparable to the escape speed from the surface of the largest planetesimal, the protoplanetary embryo. This is because the gravitational stirring of the planetesimals is due to their gravitational interactions with this embryo. As a result, his estimate for the Safronov parameter in Eq. 1 was $\Theta \sim 5$. It was this small value that was partially responsible for the long timescales required to form the outer planets.

Consider the largest of the planetesimals. It interacts gravitationally with the smaller ones, and gives them random motions with speeds corresponding to, say $\Theta \sim 5$. But it continues to grow. The next time it encounters those planetesimals it is now larger, and the escape velocity from its surface is higher. But the random speeds of the other planetesimals have not changed substantially in the interim. As a result, the value of Θ will be higher, and this largest planetesimal will grow even faster. This *runaway growth* is indeed seen in numerical studies and the resulting Θ can be many orders of magnitude larger than Safronov assumed. The resulting time to grow to planetary size is reduced correspondingly.

Of course this growth is not limited to the largest planetesimal in the swarm, although it grows faster than the rest. Other planetesimals also accrete their smaller neighbors and continue to grow. At some point, although there are still many small planetesimals present, most the solid mass has accreted into a small number of larger embryos. At the Earth's distance from the Sun, these embryos, called *oligarchs*, have radii of the order of 1000 km—roughly the size of our Moon. In a particular region of the disk there will be some tens of these oligarchs and they will each stir the smaller planetesimals to approximately the same extent. At this point Θ will decrease again and further growth will proceed via collisions among the oligarchs. This combination of runaway growth and oligarchic growth substantially reduces the time to reach planetary size.

A second factor in solving the timescale problem is the fact that for the outer planets two separate processes are involved: solid accretion and gas accretion. The planetesimal picture of Safronov only applies to the solid component of the planet. The gas component is acquired by accretion of disk gas directly onto the protoplanet. As a protoplanetary embryo moves through the gas disk, its gravitational field compresses the surrounding gas, and draws it toward the embryo. The compression of the gas, as well as the continued small influx of planetesimals, provides a source of heat that keeps the gas expanded and prevents it from hydrodynamically collapsing onto the embryo. As a result, the gas slowly concentrates around the solid core. The radius of this atmosphere and its mass depend on the mass of the solid core, and on its distance from the Sun.

As the mass of the gaseous envelope increases, the total mass of the protoplanet increases as well and its gravitational influence extends over continually larger distances. In this way the protoplanet can attract planetesimals and gas from ever larger distances. Although the influx of planetesimals does not drop to zero, it does decrease with time as the local solid mass is gradually accumulated into the protoplanet, and the mass of the solid component tends toward a constant value. But the mass of the gaseous envelope continues to grow and the total mass of the protoplanet continues to increase. The compression of the gas raises the temperature at the interface between the envelope and the core. When the temperature is high enough, the gas begins to dissociate, and the gravitational energy released by the gas contraction is partly diverted to breaking hydrogen bonds. So long as the planetsimal accretion can supply enough energy, the envelope will remain extended, but when the envelope mass becomes comparable to the core mass, the luminosity supplied by planetesimal accretion is no longer sufficient. The extended gas envelope collapses unto the solid core and a gas giant planet is formed.

It should be noted that the presence of an extended gas envelope around the rocky core creates a large region where any passing planetesimal will experience gas drag, lose energy, and thus be more susceptible to capture. As a result, the very presence of the gas increases the capture cross section of the core and lowers the accretion time.

The first detailed study of this process was done by Pollack et al. (1996). They found that, for Jupiter, the formation time is roughly 8 million years, while for Saturn the formation time is approximately 10 million years. Since most gas disks disappear after about 5–10 million years, these formation times are, at best, uncomfortably close to the upper limit allowed. The formation times for Uranus and Neptune are of the order of 20 million years, but this is less of a problem, since that time refers to the formation of a Jovian-sized planet. If the gas disk were to dissipate in the middle of the formation process, we would be left with a heavy element core of some ten Earth masses, surrounded by whatever hydrogen-helium envelope the protoplanet had managed to acquire in that time. This is actually a good description of both Uranus and Neptune.

As with the disk instability scenario, the difficulties of the core accretion scenario can also be mollified by adding additional physics; in this case the physics of small grains. The long timescale for the contraction of the gaseous envelope is due to the luminosity generated by the infalling gas and planetesimals. If that heat could be released quickly enough, the envelope contraction time could be shortened. The calculation of Pollack et al. (1996) assumed that the opacity of the envelope was similar to that of the interstellar medium. In fact there are additional microphysical processes at work in the envelope. The grains, which are the major contributors to the atmospheric opacity, collide, grow, and sediment out of the upper atmosphere. This lowers the opacity by a large factor, and as a result, the envelope radiates heat much more effectively than Pollack et al. (1996) assumed. This allows the envelope to contract more quickly and therefore to attract additional gas more rapidly. This shortens the formation time significantly. More detailed calculations, including the appropriate microphysics, shortened the formation time of Jupiter to less than 1 million years. Saturn too can be formed well within the lifetime of the gas disk. Corresponding calculations for Uranus and Neptune have not yet been performed.

Pebble Accretion

In the last few years an important modification of the core accretion scenario has appeared. Again, the source of the modification is the addition of more physics to the problem. Let us consider the fate of intermediate-sized solids, *pebbles*, in the range of tens of centimeters to tens of meters. In the absence of gas, such pebbles would be gravitationally stirred by the larger planetesimals, and captured at a rate depending on the Safronov parameter. However, pebbles are large enough so that they are not easily stopped by the disk gas, and yet small enough so that they nonetheless feel substantial gas drag as they orbit the Sun. As a result, they tend to lose their orbital angular momentum and spiral into the Sun. This was originally viewed as a difficulty, because these pebbles had relatively short lifetimes in the disk. Those few that collided with the larger planetesimals were captured, while the rest were lost. If, however, there are bumps in the radial pressure distribution of the nebula, these can form regions where the drift of the pebbles is slowed. In these regions, there will be a high concentration of pebbles, and accretion of solid material may be greatly enhanced (Ormel & Klahr, 2010). In fact the growth rate of the planetesimals can become so large that it is necessary to propose a mechanism to keep them from growing too quickly. Again, the solution seems to be to add more physics. As the planetesimal grows, it perturbs the surrounding gas in such a way as to cause the pebbles to avoid the planetesimal rather than be caught in its atmosphere. Pebbles might also provide a mechanism for overcoming the difficulty of going from grains to planetesimals that was discussed above. It has been suggested (Youdin & Goodman, 2005) that under certain conditions the combination of a high concentration of pebbles and an appropriate velocity distribution in the gas could lead to a *streaming instability* which would result in a further clumping of the pebbles so that they could form gravitationally bound planetesimals.The pebble accretion mechanism shows great promise as a means of growing large planetesimals quickly, but the idea is still in its infancy. Because the motion of the pebbles is so intimately tied to the motion of the gas, while the gas, in turn, is influenced by the motion of the pebbles, detailed numerical studies are necessary before we will be better able to understand their coupled motion. An overview of pebble accretion as of 2017 can be found in the review by Johansen and Lambrecht (2017).

Exoplanets

With regard to our solar system, it seems quite clear that the inner, terrestrial-type, planets were formed by the accretion of planetsimals. The gas giants were most likely formed by the core accretion mechanism. The most recent studies show that Jupiter and Saturn can be formed before the gas disk dissipates, especially if pebble accretion is included in the calculations; Uranus and Neptune seem to have been formed by the core accretion mechanism as well, and simply grew so slowly that the disk dissipated before they could become massive enough to capture as large a gaseous envelope as Jupiter and Saturn did. So the overall structure of our planetary system seems well-understood. However our solar system is just one example. It would be desirable to extend this picture to other planetary systems.

At the end of the 1980’s two extrasolar planets were found, Gamma Cephei Ab and HD 114762 b, but their properties were so different from the planets of our own system that their classification as planets was called into question. The first confirmed exoplanets were a system of three terrestrial-type planets around the pulsar PSR B1257+12. Here too no useful comparison could be made because of the unusual nature of the host star. But in 1995 a planet was found orbiting the Sun-like star 51 Pegasi. Called 51 Pegasi b, this planet has a mass of the order of Jupiter’s, yet it orbits its host star once every four days. A giant planet this close to the host star was a complete surprise to theoreticians.

It is hard to imagine how a giant planet could be formed so close to the star either by core accretion or by disk instability. Since that time many other *hot Jupiters* have been discovered. The most popular explanation is that the planet formed as a gas giant beyond the snowline, like the gas giants in our solar system, and then migrated to its present position. Migration is most likely due to the gravitational interaction of the planet with the gas disk. Since the angular speed of an object (planet or gas) decreases with its distance from the central star, the gas between the planet and the star moves faster than the planet, and this produces a torque on the planet which adds to the planet’s angular momentum. On the other hand, the planet moves faster than the gas orbiting further out, and it applies a torque to that gas, passing its angular momentum on to it. Modeling this effect shows that the planet will experience a net loss of angular momentum and will spiral in toward the star. This so-called *type I migration* is generally invoked to explain the massive planets observed orbiting close to their host stars.

The difficulty is that type I migration is too efficient, and brings planets into the inner solar system much too quickly. Simple models have the planets falling into the host star on timescales much less than the lifetime of the disk. More careful studies, which include such effects as accretion onto the planet, and details of the gas motion, can increase the timescale, but a more effective means of preventing all of the planets from falling into the star is for them to grow. The migration of the planet is the result of an exchange of angular momentum with the disk gas. The efficiency of this exchange depends on the viscosity, density, and temperature of the gas, as well as the mass of the planet. As the gas interior to the planet transfers angular momentum to the planet the gas moves inward. As the gas exterior to the planet receives angular momentum from the planet, this gas moves outward. If the mass of the planet is not too large, the angular momentum exchange is slow enough so that the gas can readjust and the properties of the disk remain more or less constant. But if the planet is massive enough, the angular momentum exchange is so fast that the gas cannot compensate quickly enough and the gas on both sides of the planet recedes from it, forming a gap. Once the distance between the planet and the gas increases, the rate of angular momentum transfer, and hence migration, decreases. Thus migration once a gap is formed, called *type II migration*, will be much slower, and this will help prevent planets from falling into the star before the disk dissipates.

The many planets we see very close to their host stars, with very many of them within 0.05 AU of the star (Mercury is on average 0.39 AU from the Sun) indicates that planetary migration is an important process, yet here too there seems to be a timescale problem. It appears that the short timescale of type I migration is incompatible with the much longer timescale required for growth to the size where the protoplanet can open a gap and switch to type II migration. Here too, it turns out that the answer may lie in adding more physics to an already difficult calculation. The thickness of the gas disk, as well as subtler effects like the changes in temperature gradient caused by shadows of a growing protoplanet, can slow the migration and make the two timescales more compatible. In addition, it is possible that gap opening might occur at lower masses under certain conditions.

In addition to these two migration mechanisms, angular momentum exchange is also possible between a planet and the planetesimals in the disk. In addition, close encounters between planets in the early history of the solar system can cause sufficiently strong gravitational interactions to significantly change the orbits of the planets. Although the arrangement of the planets in our own solar system appears to imply that they have managed to avoid these mechanisms of planetary rearrangement, it has been suggested that originally the giant planets in our solar system—Jupiter, Saturn, Uranus, and Neptune—were more closely spaced. Gravitational interactions between the planets themselves and the planetesimals remaining after the dissipation of the gas disk caused these planets to migrate to their present positions. Indeed in some scenarios that have been modeled, Neptune was originally closer to the Sun than Uranus, and the order of the planets changed during this migration (Tsiganis et al., 2005).

The process of planet formation is really the study of star formation, but with the emphasis on what happens to the material that does not become part of the star. We are quite certain that planets are formed from the matter, solid and gaseous, that composed the protostellar disk. The solid material tends to accumulate into planetesimals. What happens after that is less clear. The core instability mechanism seems to be better able to explain the formation of the gas giant planets in our solar system. But there are exoplanets that appear to be candidates for formation via disk instability. Fomalhaut b, a Jupiter-sized body orbiting its host star in an elliptical orbit with a semimajor axis of approximately 100 AU, is one example. At such large distances from the host star, the time to accumulate a planet via core accretion might be too long. The star HR 8799 has four giant planets orbiting it at distances of 15, 24, 38, and 68 AU, respectively. Each of these planets has a mass of roughly five Jupiter masses. This system of massive planets at large distances from the host star is another candidate for formation via the disk instability mechanism.

With the discovery of thousands of exoplanet systems, we are finding planetary systems that are extreme in different respects. Kepler 90 is a system with seven planets all within 1 AU of the host star, while the TRAPPIST-1 system has seven planets all within 0.06 AU of the host star. Kepler 16b is a Saturn-mass planet that orbits a binary star. These extreme examples of the planet formation process help us understand what the limits are on the planet formation process. How do the dynamics of the gas and solids combine to create such systems?

The wealth of observational data that is being gathered will be enhanced in the near future by more detailed spectral analysis of the atmospheres of these exoplanets. These observations not only help us test our theoretical constructs, they help us see where additional physics is needed in our models. Modeling complex cosmogonic systems is a daunting task. Even the fastest computers cannot deal with the all of the physical processes in detail. The art of the modeler is in finding which processes can be neglected in a simulation, or at least replaced by a simple approximation. However, nature does not reveal her secrets easily. Computer simulations are becoming more complex as computers become more powerful. But there is still a long way to go. What are the details of grain and planetesimal growth in the gas disk, and how do these solids influence the continued evolution of the disk gas? How do the gases and solids evolve in a circumbinary disk? How do the different sized solids interact with a growing planet to affect its subsequent evolution and its migration rate through the disk? What is the role of giant collisions? How are these processes influenced by the presence of magnetic fields? These and other questions are still issues of active research. It will be many years before we have a proper understanding of the process of planet formation.

References

Balbus, S. A., & Hawley, J. F. (1991). A powerful local shear instability in weakly magnetized disks. I—Linear analysis. II—Nonlinear evolution. *Astrophysical Journal*, *376*, 214–233.

Bar-Nun, A., Laufer, D., Rebolledo, O., Malyk, S., Reisler, H., & Wittig, C. (2013). Gas trapping in ice and its release upon warming. *Astrophysics & Space Science Library*, *356*, 487–499.

Boss, A. (2017). Triggering collapse of the presolar dense cloud core and injecting short-lived radioisotopes with a shock wave. V. Nonisothermal collapse regime. *Astrophysical Journal*, *844*(2), 113–128.

Durisen, R. H., Boss, A. P., Mayer, L., Nelson, A. F., Quinn, T., & Rice, W. K. M. (2007). Gravitational instabilities in gaseous protoplanetary disks, and implications for giant planet formation. In V. B. Reipurth, D. Jewitt, & K. Keil (Eds.), *Protostars and Planets V* (pp. 607–622). Tucson: University of Arizona Press.

Goldreich, P., & Ward, W. R. (1973). The formation of planetesimals. *Astrophysical Journal*, *183*, 1051–1062.

Johansen, A., & Lambrechts, M. (2017). Forming planets via pebble accretion. *Annual Review of Earth and Planetary Sciences*, *45*, 359–387.

Lodders, K. (2010). Solar system abundances of the elements. In A. Goswami & B. Eswar Reddy (Eds.), *Principles and Perspectives in Cosmochemistry* (pp. 379–417). Berlin: Springer.

Mamajek, E. E., Usuda, T., Tamura, M., & Ishii, M. (2009). Initial conditions of planet formation: Lifetimes of primordial disks. *AIP Conference Proceedings*, *1158*, 3.

Ormel, C. W., & Klahr, H. H. (2010). **The effect of gas drag on the growth of protoplanets: Analytical expressions for the accretion of small bodies in laminar disks**. *Astronomy and Astrophysics*, *520*, A43–A58.

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., & Greenzweig, Y. (1996). Formation of the giant planets by concurrent accretion of solids and gas. *Icarus*, *124*, 62–85.

Prinn, R. G., & Lewis, J. S. (1980). Kinetic inhibition of CO and N_2 reduction in the solar nebula. *Astrophysical Journal*, *238*, 357–364.

Safronov, V. S. (1972). *Evolution of the protoplanetary cloud and formation of the earth and the planets*. Jerusalem: Keter.

Shakura, N. I., & Sunyaev, R. A. (1973). Black holes in binary systems: Observational appearance. *Astronomy and Astrophysics*, *24*, 337–355.

Tsiganis, K., Gomes, R., Morbidelli, A., & Levison, H. (2005). Origin of the orbital architecture of the giant planets of the Solar System. *Nature*, *435*, 459–461.

Wahl, S. M., Hubbard, W. B., Militzer, B., Guillot, T., Miguel, Y., Movshovitz, N., . . . Bolton, S. J. (2017). Comparing Jupiter interior structure models to Juno gravity measurements and the role of a dilute core. *Geophysical Research Letters*, *44*, 4649–4659.

Weidenschilling, S. J. (1977). Aerodynamics of solid bodies in the solar nebula. *MNRAS*, *180*(2), 57–70.

Youdin, A., & Goodman, J. (2005). Streaming instabilities in protoplanetary disks. *Astrophysical Journal*, *620*(1), 459–469.

Morris Podolak

Department of Geosciences, Tel Aviv University