.

Fundamentals of

EXOPLANETS and ASTROBIOLOGY

From the origin of the universe, stars and elements,
and the creation of solar systems and exoplanets,
to the arise and development of life

Uffe Gråe Jørgensen, February 2023.

Part 3: The formation of solar systems

for the lectures February 28 and March 2 2023
(the appendixes, page 49 to 62, are not part of the pensum)

Version 2.3, February 23, 2023

1

The Formation of Solar Systems

— an observational approach

1.1 Basic ideas about the formation of our solar system.

In the modern standard theory for the formation of the solar system, the Sun, the planets, and the rest of the solar system is formed together from a collapsing interstellar cloud. This idea is not new. In fact it is one of the oldest still surviving theories in astronomy. Important part of its development dates back to Immanuel Kant (1724-1804) and P.S.Laplace (1749-1827) during the late 18th century. Laplace described how a rotating gas cloud collapsing under its own gravitation, will form a disk under the influence of the balance between the gravitational and the centrifugal forces. He envisioned that the planets formed as a condensation of rings of gas in the disk. In the modern standard scenario, the planets do not form from rings of gas (as envisioned by Laplace and unfortunately still drawn on many modern illustrations), but rather from planetesimals of growing sizes, that sweep up the dust in their regions of the disk. If the planetesimals become large enough, they will attract the surrounding gas from the disk and hence form openings in the disk (rather than rings) where they sweeped up the gas.

The dust itself comes mainly from condensation of the gas in the cooling nebula, and will grow in size as it falls through the nebula toward its midplane, exactly like hails or snow flakes that condense from the water vapor in the Earth's atmosphere and grow while they fall toward the ground. In the standard theory, the planetesimals sweep up the dust and grow larger and larger until they have accumulated all the solid material in their region of the disk, at which time they have become approximately Moon-sized. Slowly these Moon-sized objects will collide with one another, forming even larger solid objects. In the region beyond Jupiter, water-ice was the dominant dust grain ("snow-flake"), and its large abundance assured that the core of the four giant planets quickly grew large enough that surrounding gas from the nebula would collapse onto their cores in an analogue manner to the one that formed the whole solar system.

Many basic features of our solar system can be explained by this standard scenario, including that:

- all the planets orbit the Sun in the same direction, the same plane, and in almost circular orbits,
- the inner planets are small and solid, consisting only of the relatively rare material that will condense from a gas at high temperature,
- the four outer planets are large and gaseous as should be expected if the nebula was colder here, such that water condensed from the nebula and made them grow massive enough for the surrounding gas to

collapse onto them,

- the outer planets and their moons (but not the inner planets) resemble “miniature solar systems” in their own right, as would be expected if they (or their gaseous envelope) formed from gas collapses analogue to the solar system itself,

- The best images of young stellar-like objects in starforming regions such as Orion, reveal many collapsing protostellar systems surrounded by disks that looks like what we had envisioned for the formation of our own solar system.

However, the standard theory also has problems, which needs to be solved, and which may lead to major revisions of our scenario, or even may change it completely. These unsolved problems include the questions of

- how the Sun got rid of the huge angular momentum of the original cloud, which aught to have made the present Sun spin very rapid,

- how to explain that almost all the exo-planets we have seen around all types of stars are different from the planets in our own solar system.

We will discuss major achievements and weaknesses of the standard theory in this chapter on “the origin of solar systems”, together with the basic physical concepts behind the theory, but we will postpone the discussion of the confrontation with the observations of extra solar planets to the chapter on exoplanets.

1.1.1 Why could the Sun and the planets not have formed separately?

Some of the older theories for the solar system formation envisioned that the Sun formed alone from a collapsing cloud, and later accreted or captured the planets from interstellar space. One could envision that this theory would be in conflict with the identical age of the Sun and the planets we find listed in tables of the age of astronomical objects. However, there is no direct evidence that the Sun and the planets have the same age, and the fact that the Sun and planets are usually listed with the same age is only an assumption based on the fact that we believe that they formed together. In fact, the meteorites are the only astronomical objects in the Universe, those age can be determined directly with confidence and good precision from basic physics without any model assumption. Their age is determined from radioactive decay of elements, which shows that it is $4.567\,10^9$ years (4.567 Gyr) ago that the parent body of the meteorites chemically fractionated. We have good reasons to believe that this event mark the time where all the proto-planets melted, and other arguments indicate that the melting took place very shortly after the collapse of the proto-solar nebula. However all these good arguments are based on believing that the solar system formed all together at the same time, and hence they cannot be used to argue for a common origin. All we know from measurements is that the event that separated the meteoritic parent body into what should later be iron meteorites, stony-irons, and stony meteorites, took place 4.567 Gyr ago. This age is what we usually refer to as *the age of the solar system*. The age of the oldest Moon rocks (dust) is close to this meteoritic age, as are the oldest Martian meteorites (ALH84001). The oldest solids identified on the Earth are however only slightly more than 4 Gyr (so-called zircons), and the oldest crust only 3.8 Gyr. Parameters (of most importance the so-called mixing length parameter) in the evolutionary models of the Sun are adjusted such that the present model age of the Sun *becomes* identical to the meteoritic age. The ages of any of the other planets or other solar system objects have not be determined directly.

So while the direct age determinations does not conflict with believing that the planets and the Sun formed separately, there are other strong arguments against this idea. Some of the major arguments against the idea are

- the great chemical similarity between the refractory elements in the solar photosphere and the primitive (i.e., non-processed) meteorites. If the Sun and the planets formed separately, it should be a very unlikely coincidence that they had exactly the same relative abundances (and isotopic pattern).

- the fact that the planets all orbit the Sun in the same direction and plane. If the planets were captured by the Sun, one would expect them coming from random directions in space, and hence end in random orbits. In general the orbit of captured objects are very elliptical (as the orbits of comets, which are ”captured” from their remote reservoirs), but all the planets are in almost circular orbits, and revolve around the Sun in the same direction and almost in the solar equator plane.

Therefore it is now only of historical interest to discuss “capture theories”, and we will leave them here.

1.1.2 Why could the planets not be expelled solar material?

It has been speculated by several authors through the last centuries and decades that the planets could have originated due to a perturbation of solar material when a hypothetical star (or a ”comet” as in the original form of the theory from the 18th century where the relatively small masses of comets were unknown) passed close by the Sun. However, details in the chemical abundance of primitive meteorites show us that the material which the planets formed from, comes from the same nebula that formed the Sun, but not from the Sun itself. To understand this we first need to rehearse why only carbonaceous chondrites are really primitive; i.e., representing the original material that the (refractory part of) the solar nebula was made of.

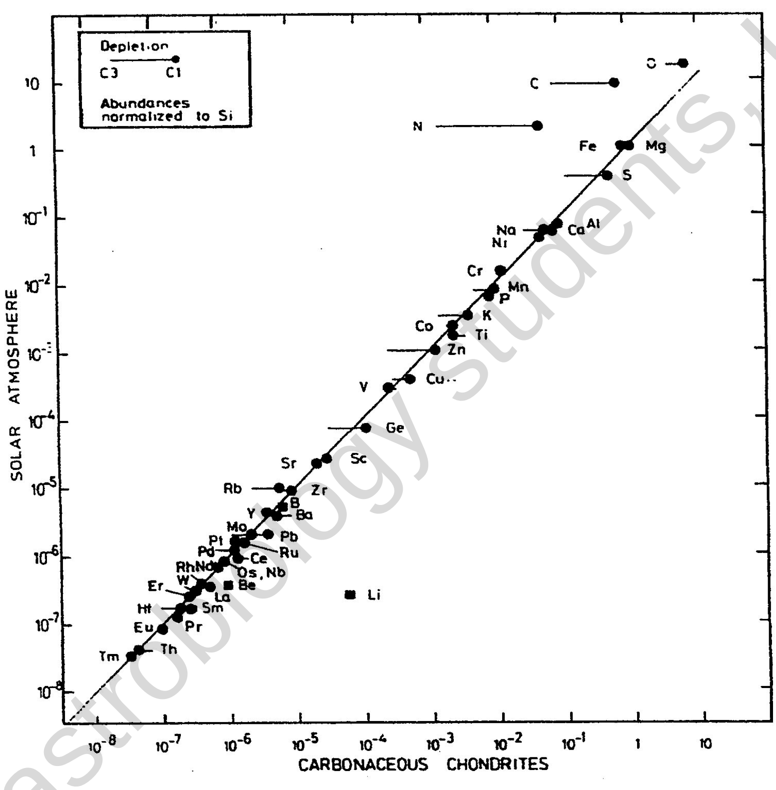

Figure 1.1. The relative abundances of elements in the solar photosphere versus the corresponding abundances in carbonaceous chondrites, all normalised to per 10^6 Si atoms. Remark the general good agreement, but also the meteoritic under-abundance of CNO elements and the pronounced solar under-abundance of lithium.

The planets have been widely fractionated after they formed. On the Earth we see a high concentration of iron group elements toward the center, and a corresponding concentration of the lighter silicates toward the surface. We explain the existence of various types of meteorites as an evidence of the same chemical fractionation process having taken place in the proto-planetary objects, or planetesimals, which became the asteroids. We envision that objects larger than a few hundred km in diameter melted shortly after their formation, which caused the heavy elements to seek toward the center, and the lighter ones to float toward the top. When such asteroids, or the precursors the planetesimals, collided with large enough energy to break the objects, pieces of the cores became the iron meteorites and pieces of the mantel and crust became the achondrites (”stony meteorites”). We will therefore not in the planets nor in the ordinary iron and achondrite meteorites find pristine solar nebula material, i.e., the original dust-grains which condensed out of the nebula and accumulated to form the planets and other solid bodies. However, not all stony meteorites are achondrites. Some are chondrites, i.e., contain chondrules (from the Greek word for seed). Chondrites are the most common type of meteorites (more than 80% of all seen falls). These have not been part of a differentiated planetesimal, but most of them are somewhat metamorphosed, and are not very suitable for identifying the original dust grains either, but a small fraction of the chondrites (about 3% of all falls) are so-called carbonaceous chondrites. Some of these have never been heated above 100°C (since they contain large amounts of chemically bound water). We might therefore find the “holy smoke” in these stones – the first solid material present in the solar nebula.

Fig. 1.1 compares the abundances of various elements in the carbonaceous chondrites with the corresponding abundances in the solar photosphere. It is seen that there is a very good general agreement between the solar and the carbonaceous chondritic material. For most of the elements the abundances are identical within the uncertainty in the estimates (which is mainly the uncertainty in the conversion from spectral line intensities in the solar spectrum to elemental abundances in the solar photosphere). From this general agreement, we convincingly conclude that the Sun and the solid bodies in the solar system were created from the same material (as already discussed above too).

Detailed comparison between the solar and the chondritic material, however, also reveal very important differences: One notices from Fig. 1.1 that the abundance of Li is very low in the solar photosphere compared to the meteoritic value. Also the abundance of deuterium is very low in the Sun (not shown in the diagram). Exactly these two elements (isotopes) ”burn” (i.e. undergo nuclear reactions) at very low temperatures, by the processes

$$\begin{aligned} {}^2D + {}^1H &\rightarrow {}^3He + \gamma \quad \text{and} \\ {}^7Li + {}^1H &\rightarrow 2{}^4He + \gamma \end{aligned} \tag{1.1}$$

From models of the energy production in the present Sun and other stars, we recognize the first reaction as the second step in the so-called pp-chain of hydrogen to helium burning. In the center of the present Sun, this reaction immediately takes place when the first step in the chain (^{1}H+^{1}H$\rightarrow$^{2}D) has produced deuterium, because the cross section for the ^{1}H+^{2}D$\rightarrow$^{3}He reaction is very high compared to the ^{1}H+^{1}H$\rightarrow$^{2}D reaction. In fact the cross sections for burning of deuterium and lithium are so high compared to other reactions, that the existing lithium and deuterium are burned already at the pre-main sequence phase of stellar formation when the star is not yet “born” (i.e., not yet has stabilized its structure as a balance between the inward directed gravitational force and the outward directed pressure gradient due to nuclear burning in its center). We know of no process that could create an over-abundance of deuterium in the meteorites (and of no likely process that could create an over-abundance of lithium), so we must conclude that it is the Sun that is under-abundant in these two elements and not that it is the meteorites showing over-abundances. The strong under-abundances of these two elements in the Sun and their “normal” (i.e., cosmic) abundances in the carbonaceous chondrites therefore tells us that the solid bodies in the solar system were not created out of material that were pushed out of the Sun after it was born, such as it could be envisioned to have happened for example if another star passed near the Sun and tidally perturbed material out of it.

We therefore conclude that we are quite certain that the Sun and the planets formed from the same cloud (i.e., the planets were not captured), but that the material the planets formed from never has been part of the Sun itself. We call this scenario of the origin of solar systems for the nebula theory, and at least our own solar system formed this way. It is most likely that all solar systems mainly formed this way, but strictly speaking we of course cannot know. As we wil discuss in the chapter on exoplanets, it seems that at least some stars have exoplanetary systems that include planets were captured after the system formed. We don’t know whether our own solar looks the way solar systems usually looks like, or whether we are here because our solar system looks in exactly a very rare (or even completely unique) way that is necessary for life to develop, but as we will see in the chapter on exoplanets, most observations now hint that there is something unusual about the planets in our solar system compared to “a typical” exoplanetary system. We are now coming close to having instruments that are capable of identifying the planet forming process around other stars, we at the brink of being able to observe exoplanets that may resemble the planets in our own solar system, and we are on the way to have theories about planet formation that almost everybody in the scientific community seems to agree on. Large observational, theoretical, and technological efforts are during these years put into solving these central questions about our place in the evolution of the universe. In the following we will first outline the status of our present observational knowledge of what we believe are solar systems under formation, before we will present the theoretical understanding of the subject, and finally look into our knowledge about planets around other stars than the Sun.

1.2 Observations of starforming regions.

Many regions in the solar neighborhood show groups of very young stars, and even stars under formation. Some of the most well known such regions were listed already by Charles Messier in his famous 18th century catalogue of “non-cometary nebulae”. Details of his object M16 became famous in 1996 when media around the world showed the Hubble Space Telescope’s astonishingly clear and beautiful new pictures of the interior of M16, which was named “pillars of creation” for the occasion (see Fig. 1.2). However, M16 is 7000 light years away, and even at the impressively high resolution of the HST images, our entire solar system would be smaller than a single pixel in these photos, so even though the pillars do take part in the creation of new solar systems, we do not see many details of what an early system may look like, but instead we must infer it from the collective interpretation of many different observations, including the pillars of creation in M16.

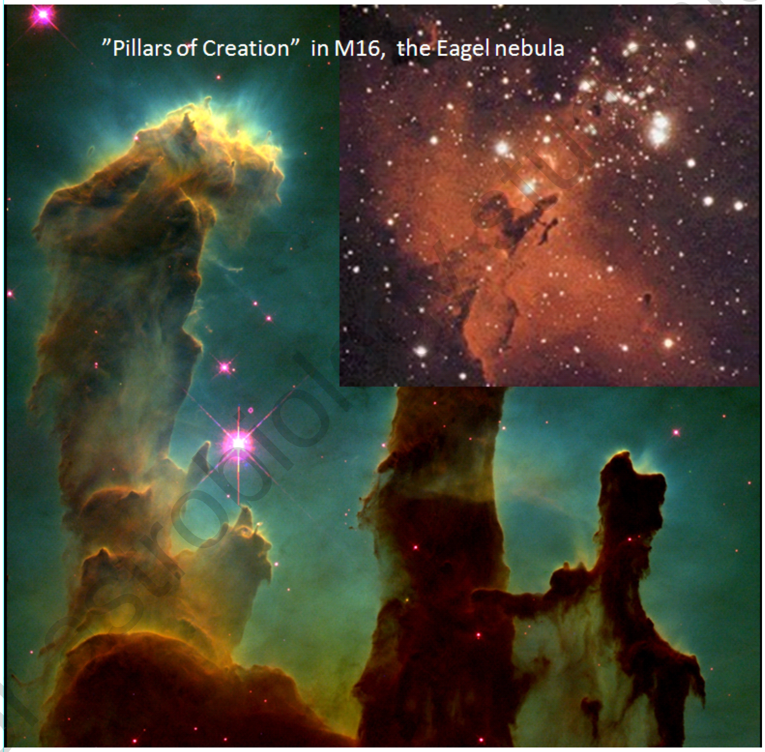

Figure 1.2. The pillars of creation: an interior region of M16 observed from Hubble Space telescope. The pillars are regions of a dense molecular cloud surrounding an expanding cavity filled with a thin hot gas of mainly hydrogen (and helium), which ionize its way out through the cloud. While an ionization front of the H II region is moving forward and evaporating the boundary layers to the molecular cloud, newborn stars are being exposed. Inset shows a larger region around the pillars, observed from the ground.

Messier's object number 42 is the Orion Nebula. It is the closest of the major starforming regions, just visible to the naked eye as a dot, 6 times closer to us than M16, and of approximately the same size ($\approx$ 20 light years in diameter). The orbit of Neptune would fill two pixels in an HST photo at this distance, so even here we will not be able to match the finer details of our models of the solar system formation against the photos of newborn stars, but we can use these fascinating pictures to guide our ideas toward the most realistic scenario of how solar systems seem to form. Such a scenarios will form our basis for understanding where the most fundamental problems are in solving the question that concerns us most here:
"Did Earth become as biofriendly as it is because of some very unusual coincidences in our solar system, or do the same conditions rather repeat themselves all over the Galaxy all the time? Are the conditions that led to our solar system identical to those that result in the star formation we find in M16 and M42 today?"

Both M16, M42 and other major starforming regions are part of huge continuous cloud complexes that stretches over thousands of light years in space, and over vast regions of the sky. They consist of many different types of interstellar clouds in mutual balance, but the ones that are particular important for the formation of stars and planets, are the so-called giant molecular clouds. They are huge gas masses of millions of solar masses and hundreds of light years across. They are called *molecular* because they are cold enough that most of the hydrogen is in the form of molecular hydrogen, H_2. Regions of these clouds can have number densities exceeding $n \sim 10^5$/cm^3. Although this is close to vacuum compared to our atmosphere (of almost $n \sim 10^{20}$/cm^3), it is very dense compared to the surrounding interstellar space. It is in molecular clouds we find the turbulence and pressure variations that will eventually make a few percents of the gas collapse into stars. As a byproduct of the collapse, a fraction of a percentage of the tiny amount that becomes stars, may eventually form planets, of which we might find another fraction of a percentage ending up as a habitable planet where life eventually may arise.

What we see as the starforming nebulae within the giant molecular clouds, are openings illuminated by a few new-born high-mass stars in the interior of the cavities. We can envision that the creation of the nebula started with a handful of massive ($\sim M > 8M_\odot$) stars, that formed during a Jeans-type collapse of interior parts of a piece of the molecular cloud. Such stars are much hotter (e.g., 30,000K) and more luminous (e.g., $10^5\ L_\odot$) than the Sun. Their UV radiation will be millions of times the solar UV luminosity, and as a result they will ionize the surrounding gas, and heat it to tens of thousands of degrees. Such hot, ionized gas consisting of mainly hydrogen (and helium), is called an HII region. It typically has a number density as low as $n \sim 100/\text{cm}^3$. We can think of it as an almost empty cavity in the surrounding more dense cloud. It's opacity against UV radiation is very small, and the UV light from the massive stars will therefore shine through it, and feed an ionization front where the "cavity" of the HII region meets the denser cool gas of the original cloud. Typically the ionization front will "burn its way" outward through the molecular cloud with a speed of a few km/s. Soon the ionization front will therefore reach a local boundary between the giant cloud complex and a lower density surrounding space, and thereby open up "a hole" in the cloud. For a while the ionisation front will keep pushing the hole more open, and we will see the illuminated interior as "a nebula", as is illustrated in Fig. 1.3 (upper panels). Messier called two of these "illuminated holes" for M16 and M42 in order to remember where they were, so that he would never again be confused by their comet-like fuzzy appearance to his eye, in his hunt for new comets. Fig. 1.2 (M16) and Fig. 1.3 (M42) show details of them in the high resolution obtainable with HST, as well as in their larger extend, visibly from telescopes on the ground.

In Orion we see the massive ionizing stars as the small group of OB stars that form a Trapezium near the center of the nebula (Fig. **??**, two lower left panels). The Trapezium cluster is just the few brightest stars of a large cluster of ∼3500 new-born stars, most of which are hidden from our eye by the large opacity of the surrounding cloud. In some million years the cloud around the cluster will have been completely pushed away, and we will see an association of stars that were born together. Many of the "hidden" stars in the Trapezium cluster can already be seen in infrared radiation, which penetrates the cloud better, as is shown in the lower middle panel of Fig. 1.3. In the beginning they will form a group of stars visible in the sky as an open star cluster such as the Pleiades. With time it will be increasingly difficult to figure out which stars were born together, as they drift away from one another, but still their common origin may be traced for a while by their almost common velocity in space, such as it has been done for example for the small β Pictoris group.

Only a few percentages – at best up toward as much as 30% in the densest regions – of the gas mass

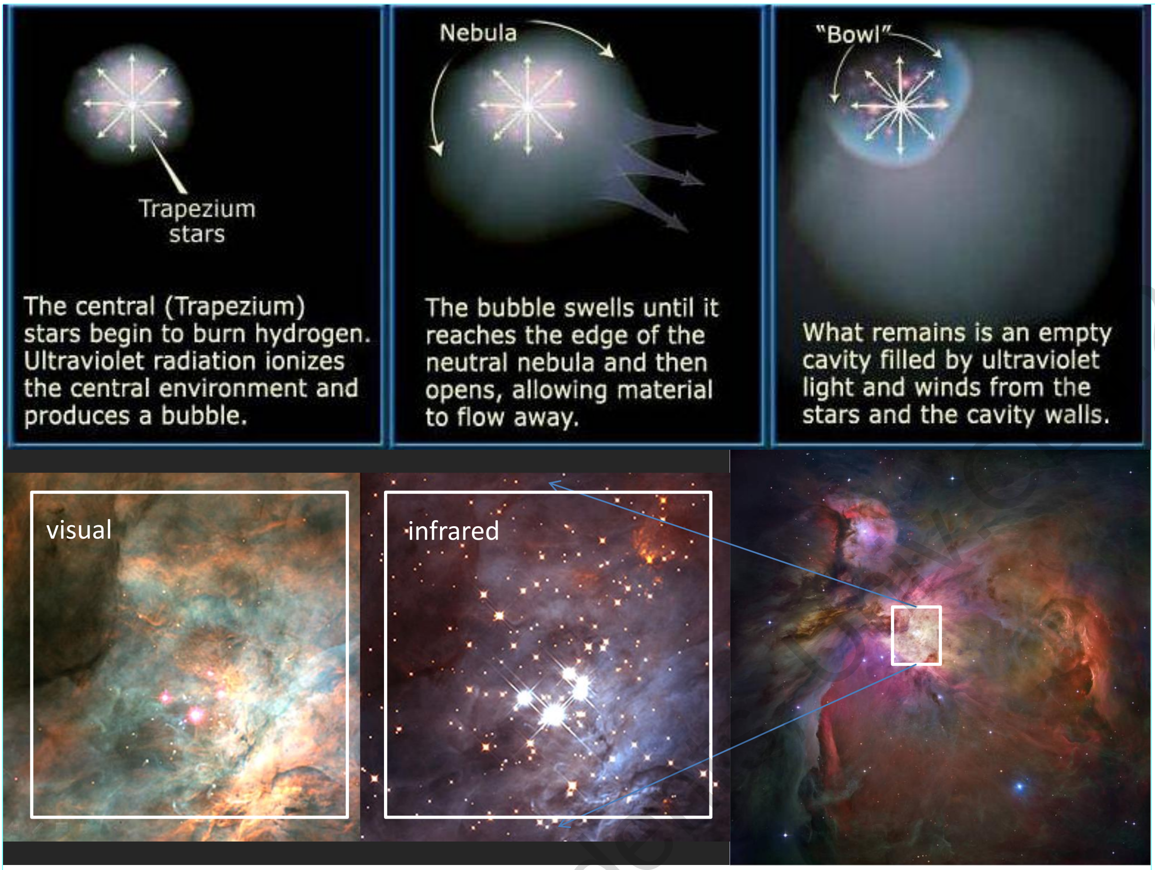

Figure 1.3. The Orion nebula as seen from the ground (lower right panel) as well as the 2.5 light years central region around the Trapezium cluster, as observed by the Hubble Space telescope in visual light (lower left) and infrared (lower middle), respectively. The three upper panels is a schematic illustration of the birth of the trapezium cluster stars inside a molecular cloud (left panel), their creation of a surrounding HII region (middle panel), and the final appearance of the Orion nebula (upper right panel) when the HII cavity opens up to our vision. It is this opening "bowl" of the schematic model that corresponds to the way we see the nebula (lower right panel) today.

actually collapses to form stars. The bulk of the self-gravitating material that keeps the cloud and the cluster gravitationally bound, will return to the ISM as gas during the final stages of the starforming process. Many clusters will therefore disperse as soon as the gas is gone. Other will loose the gravitational binding because gravitational interaction between the stars will expel some of the stars, and only few clusters will stay bound until the stars reach a mature age. Large-scale surveys indicate that only 10% of the clusters that form, will survive 10 million years, and less than 4% survive to an age of 100 million years. Although probably all stars are formed in clusters, as the Orion Trapezium cluster or in smaller clusters, most stars will appear as isolated objects, such as our own solar system, after a small fraction of their lifetime. It is still an open question where the Sun formed, and in which surroundings, but although its siblings are now scattered around the Galaxy, some of their effects on the early solar system can still be traced, and we are beginning to be able to debate whether their exact size, position, and movement relative to the Sun was essential for the rising conditions for life on our planet, which we will return to in a later chapter.

Where the ionization front heats the dense cloud, material will evaporate and blow backward into the cavity. We see this boundary as a luminous region where the interior HII region meets the surrounding dense and dark gas. As a reaction to the inward moving hot evaporating gas, a shock wave will move outward through the dense cloud material, ahead of the ionization front. The shock wave will trigger a collapse of gas clumps on its way, thereby creating a chain of new star formation along its way. While the ionization

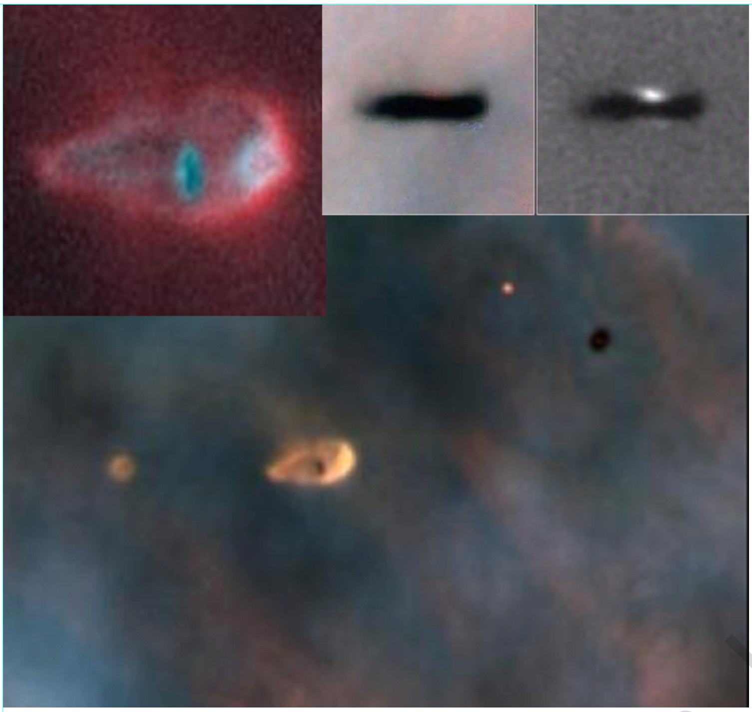

Figure 1.4. Proplyds in the Orion nebula, photographed by HST. Insert show enlarged and contrast enhanced photos of a proplyd and a protoplanetary disk.

front that follow the shock wave approaches the collapsing clumps, the clumps will start evaporate from their surface, but they are still connected to the surrounding cloud, so we will expect a continuing, very chaotic and violent infall of material during this phase. All the gas that is to become the central star of the system is to be gathered from the surroundings during this phase. In microwave and infrared radiation we can see a bit into the dense clouds, but only when the clump breaks through the boundary "wall" between the surrounding cloud and the transparent HII region, are we able to see it in visual light. We call this phase of star formation for EGG (*evaporating gaseous globule*). Several EGGs are identifiable as "dots" on the very edge of some of the "small" side-features on the "pillars of creation" in Fig. 1.2 and Fig. 1.6, as is most easily seen in the inset of Fig. 1.6. Hence, the pillars are not cloud material falling or rising into an empty cavity, but rather remnants of dense cloud material that are being eroded (evaporated) away by the expanding ionization front. The position if many EGGs on the tip of small "side-pillars" or "stalks", make us believe that these dense gas clumps are actually the reason for the pillars, as they act as temporary shields against the ionization front for the region behind them, just like a stone in a stream of water. In this sense it may be causally more correct to say that we see "the creation's pillars" rather than "pillars of creation".

Once the EGG isolates from the wall, or pillar, due to the expanding ionization front, it is fully exposed to the strong UV radiation field. The remaining envelope of the EGG will then quickly ($\sim 10^4$ years) evaporate into the HII region, and gradually the interior of the cloudlet (the EGG) will become visible. While the uppermost layers of the gas envelope evaporate, deeper layers are now ionized and pressed backward by the general gas flow in the HII region, into a droplet formed cavity, a bit like the upper atmosphere of an unprotected planet in the solar wind. This phase of final eroding of the gas envelope is called a proplyd (PROto-PLanetarY-Disk). In Fig. 1.2 of M16 one can glimpse several EGGs as well as proplyds. In M42, which is closer by, the proplyds can be seen in more detail, as is shown in Fig. 1.4.

During the EGG phase, the first concentration of mass toward the center will have taken place, maybe concentrating 1/100 $M_\odot$ inside the inner few AU. We will usually call this gas concentration a protostar, although there is no strict definition for when the hot luminous gas is a cloud, and when it is for the first time to be called a protostar. Since it is a figurative description of the phenomenon, people have a tendency to use the word a bit to their own liking, which can be confusing in the beginning. It doesn't make it clearer that the word is inconsistent in itself. A star is an object that shines due to a stable nuclear "burning" that has stopped the gravitational contraction. ("burning" is of course another erroneous use of the terms, but for some reason the more correct term "nuclear fusion" is not traditionally used in astronomy). The word *proto* comes from Greek (*protos*), meaning *the first*. Literally *protostar* therefore means the first gaseous objects that have stopped their contraction because of nuclear reactions in their interior. In the case of what we call a protostar, the gravitational contraction is still in full force, and the nuclear reactions have not started yet – in this sense it is neither *proto* nor *star*. It is the energy from the gravitational contraction that makes the gas shine during its protostellar phase. A more correct term would therefore be "a pre-star-phase", "a pre-stellar cloudlet", or just "a pre-star". In the following we will of course continue to (mis)use the word proto-star in the way it is usually done.

The gas accretion onto these objects is far from smooth and regular, but rather highly irregular as clumps of gas is accreted and falls through the disk and onto the central object (the protostar). This first phase of star formation is therefore a very violent process, and historically it was first observed in what appeared to be an otherwise normal star in the constellation of Orion. Since this object was seen to change luminosity, it was named according to the convention for variable stars, and called FU Orionis. Objects that behave like FU Orionis are therefore now called FU Orionis stars. They typically increase strongly in luminosity over a period of the order of a year, and thereafter decrease again over periods of decades or longer. One of the most well studied FU Orionis stars, V1057 Cyg, increased with 6 magnitudes over a period of a few years, and changed in appearance from that of a K-dwarf star to that of an F-supergiant. Most likely FU Orionis stars are protostars where the increase in luminosity is due to outbursts of energy in connection with huge clumps of material that falls through the disk and are accreted onto the protostar. FU Orionis may therefore be another manifestation of the EGG-proplyd phase, historically discovered with other types of instruments in other physical environments, and may as such be a short phase of evolution that all low-mass stars go through.

The luminosity $L_{\rm accr}$ of the very early protostar of mass M_* and radius R_* is driven by the gravitational infall rate $\dot{M}$,

$$L_{\rm accr} \sim GM_*\dot{M}/R_*.$$

The infall rate during the FU Orionis phase can therefore be estimated from observations of the luminosity, and is found to be $\dot{M} \approx 10^{-5}\mathrm{M}_\odot$ per year, pointing at a lifetime of the FU Orionis phase of $\sim 10^5$ years. Also number statistics indicate that their lifetime must be of the order of 100,000 years. In a starforming region as Orion, the timescale for which mass can flow from the molecular cloud, through the disk, and onto the protostar, is determined by the time difference between the start of the collapse to the evaporation of the cloudlet surrounding the protostar and disk. In large starforming regions such as M16 and M42 there is typically of the order 0.1 pc between the shock wave (i.e. the place where the collapse is initiated) that propagates into the molecular cloud and the ionization front itself, and the chock propagates with a speed of $v \sim 1$ km/s. It will therefore take $t \approx 0.1$ pc / 1 km s^{-1} $\approx 10^5$years (i.e., the same timescale as the lifetime of the FU Orionis stars) from the cloudlet starts collapsing until it is isolated from the surrounding feeding cloud as a separating EGG or a young proplyd. During the late EGG phase or early proplyd phase the accreting protostar will become visible for the first time though its envelope (Fig. 1.4).

We can envision that also our own solar system began its existence as a collapsing cloudlet that gathered most of the Sun's mass by infall from a surrounding larger cloud during a period of about 100,000 years, until it became an isolated object for the first time, resembling a proplyd or an FU Orionis star. The fact that FU Orionis stars are strongly and very irregularly variable, indicates that the accumulation is a very chaotic and irregular phenomenon in time. If this is a correct interpretation, we must expect that our own solar system began its existence with a short ($\sim 10^5$ years), very violent phase, where large irregular clumps of infalling and erupting material were flowing through the disk in which the planets were soon to form. During this first phase it was therefore still too early for a "silent snowing" and settling of dust grains to the mid-plane to form the first basis for solid planet formation.

While a rotating cloud contracts to form a protostar, its rotational speed will have to increase pretty much due to conservation of angular momentum. In the case of our own Sun, it should have increased its rotational velocity so much that it would spin around its axis in seconds only, if nothing other than a simple contraction

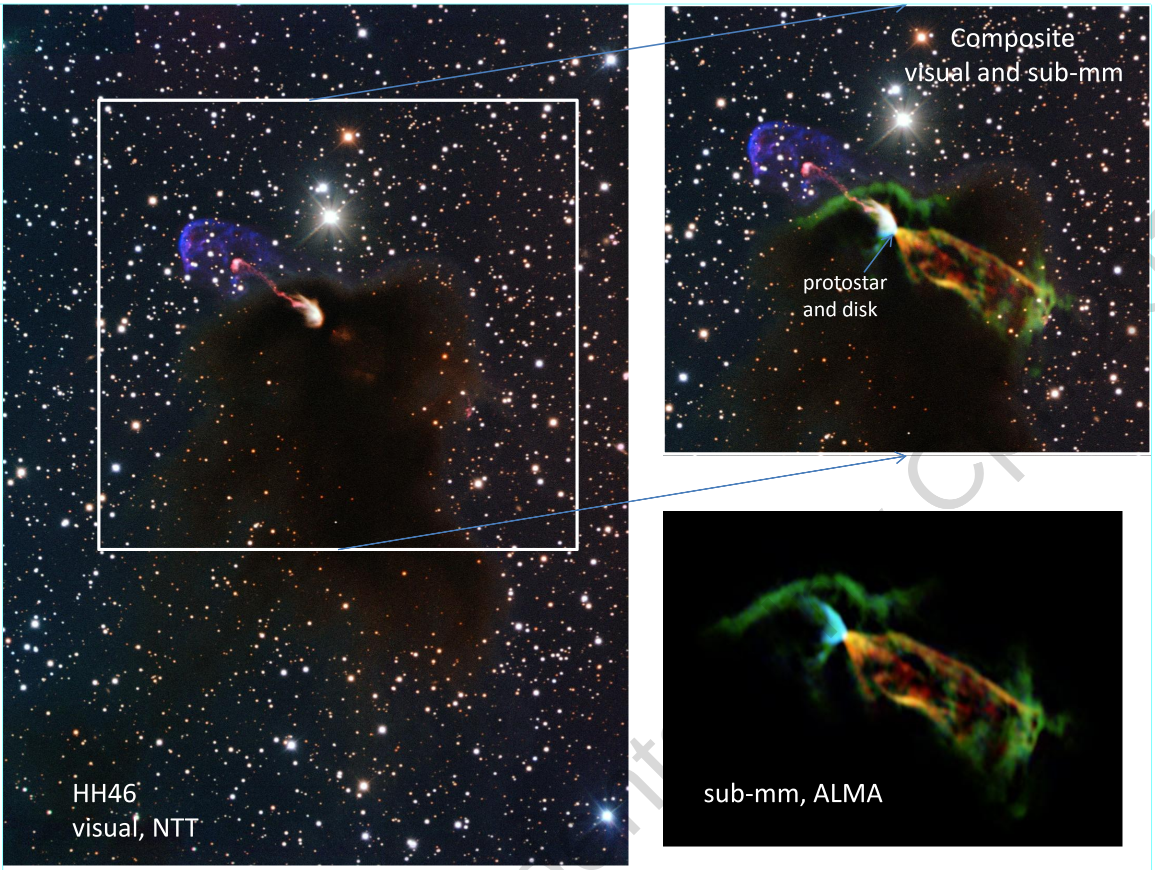

Figure 1.5. Left panel: The 0.5 light years long jet HH46/47 (red), sticking out of a dark molecular cloud, and forming a chock wave (blue) where it hits the sourounding interstellar medium. On the right end of the jet a binary protostar and a disk are hidden in a dense Bok globule. Lower right panel: A sub-mm image of the other arm of the bipolar outflow, stretching inward into the dark molecular cloud. Upper right panel: A composite of the visual and the sub-mm images.

had taken place. However, the Sun rotates very slowly today, and its specific angular momentum is much lower than that of the planets (which in this respect represent the original cloud). A most plausible way for a young star to get rid of its specific angular momentum, is to send a fraction of its material out along magnetic field lines. The mathematical details of this complicated rotation dilemma and its coupling to magnetism, will be discussed in detail later. Here we will just touch upon whether we actually observe anything that resembles such a mechanism. If our understanding of the formation of stars and solar systems are qualitatively correct, we should therefore expect to see jets of material being expelled at high speed along magnetic field lines in connection with the phase of early stellar evolution where material is accreted from the surrounding cloud onto the protostar under formation. Such phenomena are observationally now known as Herbig-Haro objects. Originally they were discovered as small nebula-like clumps associated with starforming regions, but more recent it has been realized that they are bipolar outflows of high-speed gas clumps that emit visual light during violent collisions with the surrounding cloud. Figure 1.5 show a particular large such object, known as HH46/47, almost a full light year long from the tip of one arm to the other of the bipolar outflow. One arm of the outflow stretches out of the dense molecular cloud (which can be dimly seen as a dark area i.e., with almost no stars visible) in the visual light image (left panel of the figure). The other arm stretches into the deep darkness of the molecular cloud, and cannot be seen in visual light, but reveals itself at sub-mm wavelengths (lower right panel). Upper right panel shows a composite image where the sub-mm image of the inward moving jet is superimposed onto the visual light image of the outward moving jet and the surrounding cloud and background stars in the field.

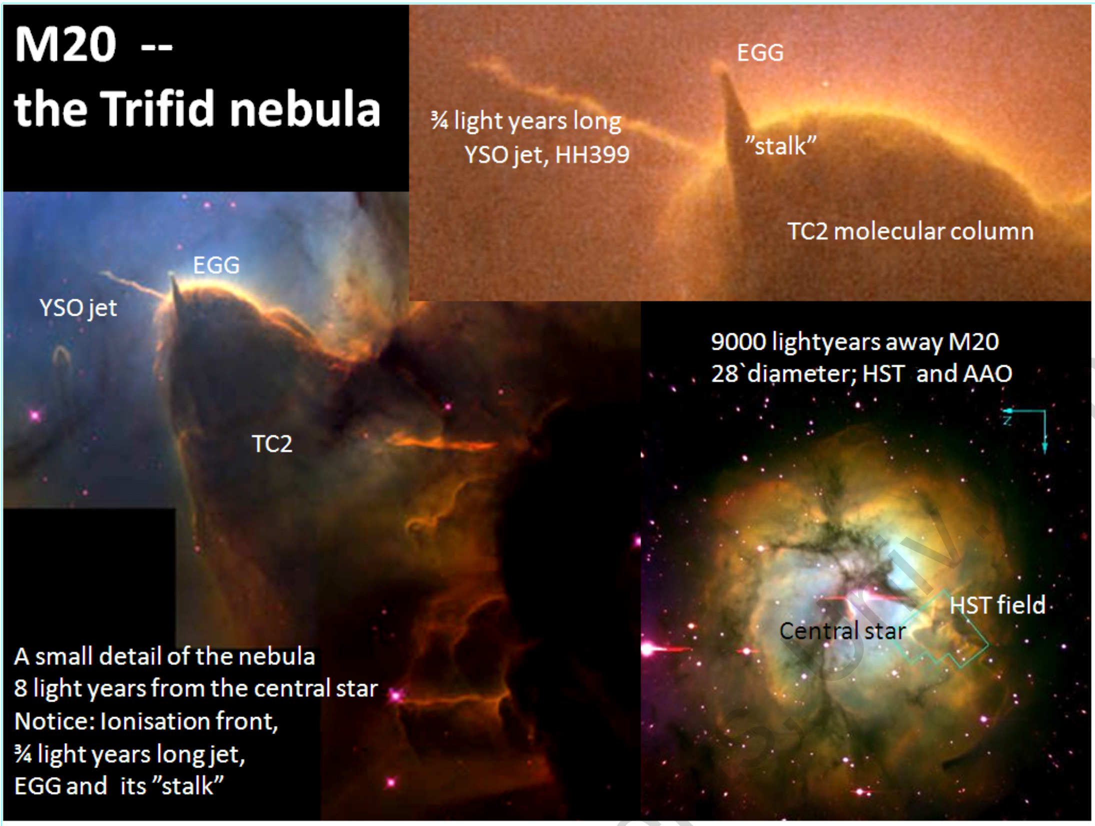

Figure 1.6. M20 – the Trifid nebula, is another major starforming region. Lower right image from Australian National Observatory show the full 20 light year large nebula. Left image from HST show details of the HII region (left) and the molecular cloud including the "pillar" TC2 (molecular column). Upper right image show details of TC2 with the 3/4 light years long Herbig-Haro object HH399 (one side of a bipolar outflow) and a "stalk" with an EGG on the top.

Typical speeds, mass and densities of Herbig-Haro objects are 100 km/s, a few Earth-masses, and 10^5/cm^3. While HH47 is particular large, and of a type that may be associated with binary stars only, analogue jets are seen in connection with all major starforming regions. Fig 1.6 show a bipolar outflow called HH399 in the starforming region M20. Next to HH399 is another "stalk" on which top sits an EGG, just like we saw it in M16 in Fig. 1.2 (and at a later evolutionary stage in Fig. 1.4). While HH399 is of the same size as HH47, also the EGG next to HH399 show a jet, although still so small that it is only visible in strong enlargement and image processing of the photo in Fig. 1.6, but it clearly shows that a jet is a normal phenomenon connected with the origin of solar systems, shedding confidence to the interpretation of our solar system model and the solution of the longstanding "solar angular momentum problem" of the early loss of the bulk of the Sun's angular momentum, that we will discuss in detail later.

After the gas envelope of the proplyd is completely blown away, also the outer parts of the disk will be exposed to the evaporating UV field in the HII region, and soon the gas disk will be too small to be seen, even in HST photos, but its existence can still be traced as an excess of infrared radiation when observing the central star. Again there is a type of objects, which historically were identified as variable stars, that resembles this phase of evolution. This time they are named after the variable star T in the constellation of Taurus, and are therefore called T Tauri stars. T Tauri variables are characterized by rapid rotation (typically between 1 and 12 days rotation periods), high chromospheric activity (including Hα emission), strong lines of lithium in their spectra, and often traces of a strong stellar wind. They are of spectral type F, G, K, and M (but considerably brighter than corresponding main sequence stars) and have masses below 2 M$_\odot$. Just as for FU Orionis objects, they are not really stars, since they shine because of contraction. Only after about

100 million years they start stable hydrogen burning as they enter the main sequence. About half of the T Tauri stars have signatures of a disk. Their corresponding higher mass (2-8 $M_\odot$) counterpart is called Herbig Ae/Be stars (or Herbig emission stars), while stars with masses even higher ($M > 8M_\odot$) have not been observed with signs of a disk.

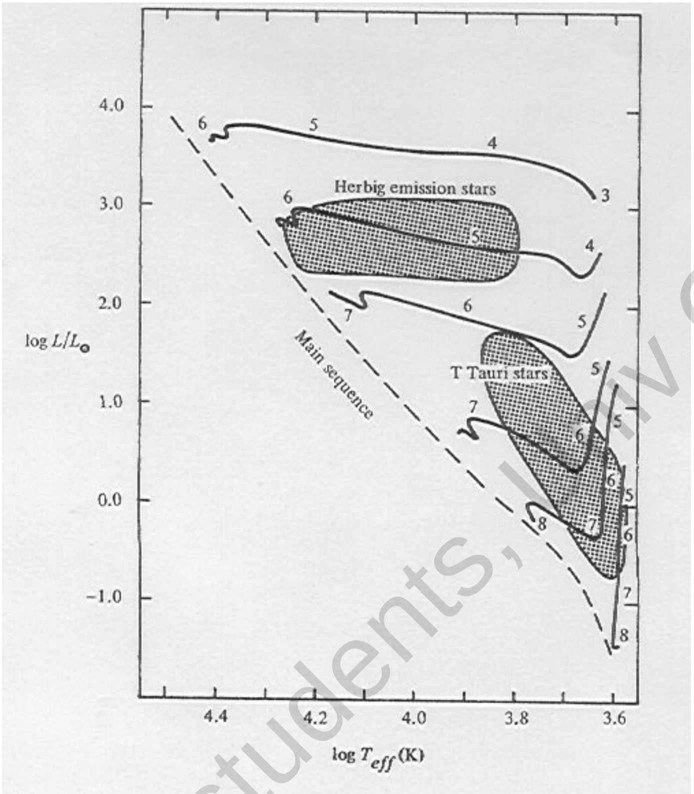

Figure 1.7. Evolutionary tracks for young star-like gas-clumps evolving toward being born as zero-age main sequence stars. The tracks from top to bottom represent models of 9, 5, 3, 1.5, 1.0, 0.5 $M_\odot$. The numbers along the tracks gives the logarithm of time in years since beginning of collapse. Shaded regions indicate observed Herbig emission "stars" (HAeBe) and T Tauri "stars" (TTS). For example one sees that a 1 $M_\odot$ gas clump is in its T Tauri phase for 10 million years, but takes 100 million years to reach the main sequence, while a 5 $M_\odot$ gas clump is a HAeBe emission object during most of its 1 million years short contraction phase to the main sequence (MS).

Fig. 1.7 shows the region of observed Herbig Ae/Be stars and T Tauri stars in an HR diagram, together with theoretical contraction tracks of gas-clumps of $0.5M_\odot$ to $9M_\odot$. Just as for the FU Orionis stars, T Tauri stars are strictly speaking not stars, but contracting gas clumps on their way to be born as stars.

Recently, it has won acceptance to try to name the young pre-stellar objects in a more logic way. Fig. 1.8 sketch the new naming convention, where one tries to classify them according to a logic evolutionary sequence that is reflected in what they look like observationally, but at the same time is meant to describe some physical, evolutionary aspects. In this description, young star-like objects (YSO) of type 0 corresponds to the deeply enshrouded object in the middle of an EGG, only visible in infrared of microwave wavelengths. YSOI corresponds to the period where most of the stellar mass is being accreted, which presumable is the FU Orionis phase or the late EGG and proplyd phase. YSOII is the first few million years of the T Tauri phase, where a pronounced disk exist around the central object that is going to become the star. YSOIII is the $\sim$10 million years long phase where the infrared visibility of the disk gradually disappears out of the observable energy distribution of the observed T Tauri star. Some times YSOII are also called classical T Tauri stars, and YSOIII are called weak-line T Tauri stars. Classical T Tauri stars most often also show Hα emission, which is interpret as radiation from gas falling inward from the disk toward the protostellar object. In weak-line T Tauri stars, the infall rate gets too low for the corresponding Hα emission to be observable, (therefore “weak-line”; generally it means that the infall rate has dropped below $\sim 10^{-9} M_\odot$ per year). From observations of young open star clusters, the classical T Tauri phase is known to last between 1 and a few million years, while the duration of the weak-line phase is $\sim$10 million years or longer. It is not obvious to which extend the disk is gone by the end of the T Tauri phase, since the Hα emission measures the infall rate and the IR radiation measures the amount of fine-grained (i.e., μm sized) dust. However, microwave observations of the CO gas component around nearby T Tauri stars are at least not contradicting the assumption that also the gas disk is gone 10 million years after the EGG phase, and as we will see later, several evidences from our own solar system seem to indicate that the gas that collapsed to form the four outer planets in our solar system was also gone 10 million years after the first collapse phases of the cloud that became our own solar system.

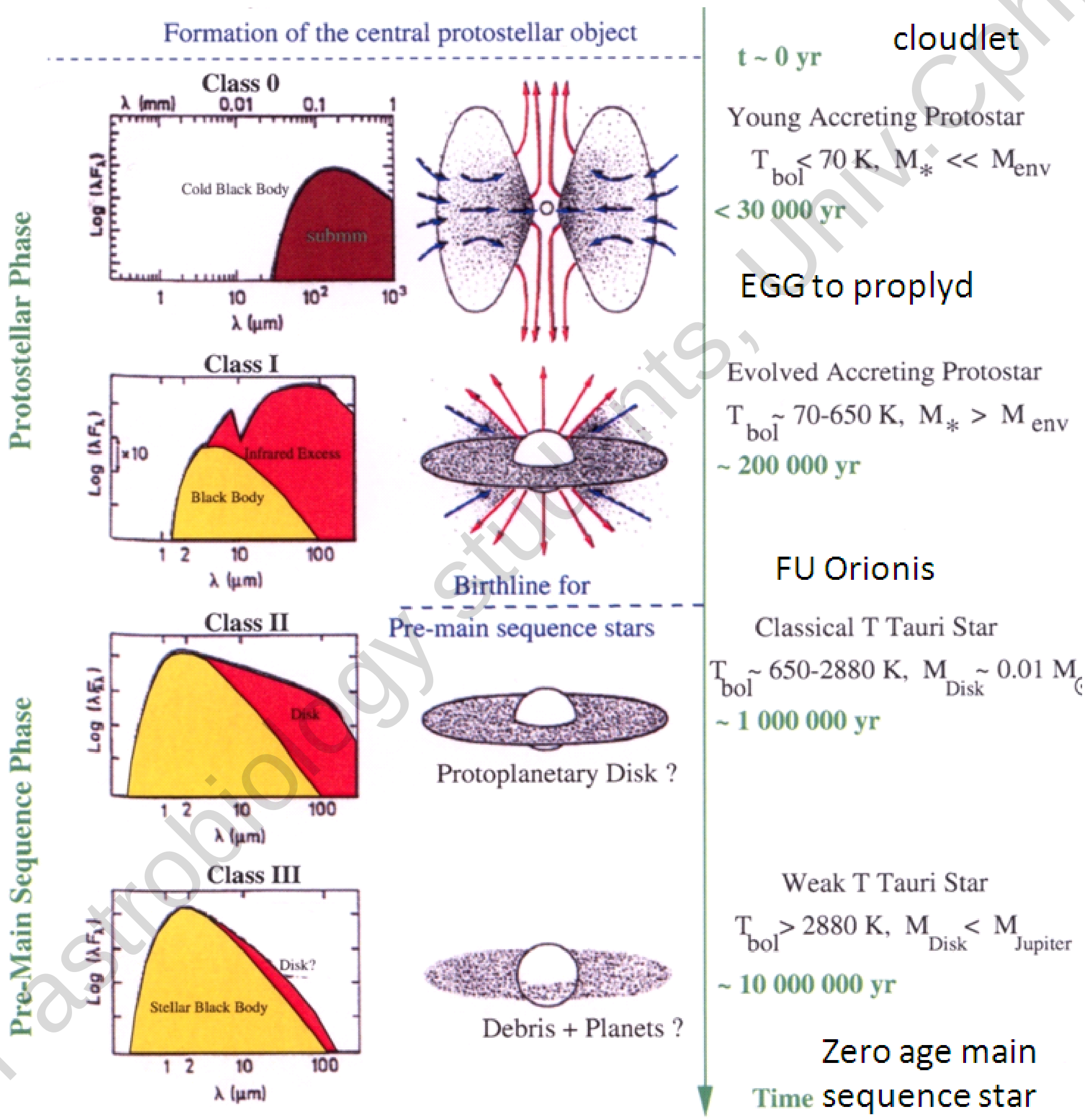

Figure 1.8. A schematic illustration of the evolution of a collapsing low-mass gas cloud through the phases YSO class 0 to III. In the right side of the panel is given the approximate corresponding names in the older classification scheme of EGG, FU Orionis, T Tauri, etc., together with the estimated typical cloud temperature and life time of the phase.

Numerical models indicate that once the proplyd envelope is gone, in an Orion-like environment, the young disk will quickly evaporated in to a distance of $\approx$ 50 AU from the central star. If such models can be proven to be a reliable representation of the early solar system, this truncation of the early disk could be the explanation for the seemingly abrupt cut off in the number of Kuiperbelt objects beyond 50 AU from the Sun, and indicate that our solar system was born in an Orion-like environment. The 50 AU boundary could also mark the place where the centrifugal force equals the gravitational force of the collapsing cloud, as will be discussed later. If planets and minor bodies formed in our solar system out to $\approx$50 AU only, possible because of the truncation of the early disk, then any population beyond 50 AU must necessarily have come there due to perturbations. This could partly be due to the outermost planets, but objects as far away as the dwarf planet Sedna, are very hard to explain in this way. Instead, perturbations by one or more of the Sun's early siblings in the cluster, could be an explanation. Such a close passage of another of the cluster stars soon after the formation of the Sun, could also form a natural explanation for why the ecliptic plane and the solar equator plane are not exactly aligned, but rather inclined 7° relative one another. Numerical simulations indicate that the distribution of the perturbed objects are very sensitive to details in the possible stellar passages, and observations of objects beyond 50 AU may therefore one day reveal details about how the stellar density conditions were in the cluster in which our solar system formed 4.5 billion years ago. An exciting perspective is that if the passing star(s) also had a planetary system, some of these planets might have been captured by our solar system, and will in that case have formed a population of planets with large orbital radii and high inclination angles, maybe one day to be identified.

Inside 50 AU, the disk of a YSO is sufficiently dense to be self-shielding against the UV radiation field, and it can persist for a few million years after its formation, before its gas is completely gone. It is far from certain how the final disk (inside 50 AU) is lost, but the common explanation from popular science books, that a strong stellar wind sets in when the protostar settles toward the main sequence, thereby blowing away the disk, is rather unsupported by scientific studies. A more likely explanation is that the outer parts of the disk gradually are photo-evaporated by the surrounding UV field, and that the inner parts spiral into the protostar and/or is lost through a Herbig-Hero-like YSO jet.

If all low-mass stars go through first a phase of EGG and possibly also a proplyd, lasting $\sim$100,000 years, and then a phase where the disk (seen as Hα emission and IR excess) gradually disappears over a few million years, then we should expect 10 to 100 times more T Tauri stars than EGG and proplyds in starforming regions. In the Orion HII nebula $\sim$150 EGGs and proplyds have been identified, while there are more than $\sim$3000 T Tauri stars, in agreement with these timescales. More than half of the Orion YSO show excess infrared radiation that can be interpreted as coming from a surrounding disk.

As a summary, it is therefore consistent with the observations presented above to assume that solar systems build up by accreting mass from the surroundings during a period of $\sim$100,000 years, then get rid of the outer envelope and disk down to $\sim$50 AU during the next $\sim$100,000 years, and finally slowly loose the rest of the disk during the next $\sim$10 million years. Within this period the dust must have grown enough in size in order to no longer reveal its existence by infrared emission, but the full sized solid planets can well have finalized their accretion only much later.

Figure 1.10 illustrate how one can envision the formation of low-mass stars with planetary systems, such as our own or such as it might take place in the Orion nebula today, inside embeded clusters in HII regions. In the first panel of the figure, one or more high-mass stars are formed, and their strong UV radiation will dig a cavity into the surrounding molecular cloud. While the associated ionization front and shock front pave their way through the cloud, low-mass stars are formed, which slowly will be exposed as clusters of T Tauri stars in the expanding cavity. Before planets have had time enough to form in their accompanying protoplanetary disks, the high-mass star(s) will explode as supernova(s), and impose their radioactive material into the forming planets. Radioactive ^{26}Al, and ^{60}Fe traced in meteorites may have come into the protosolar nebula in this way. In this scenario the supernovae were not the triggering source of the solar system formation (as previously thought), but rather a natural consequence of the same process that also formed the solar system. In the present scenario it becomes natural that the meteorites seemingly contain elements from more than one contemporary supernova explosion. Also traces of ^{10}Be seen in primitive meteorites may have their origin in connection with supernova explosions of stars in the cluster the Sun was a part of during the first million years of its lifetime. Two remaining obvious questions are however:

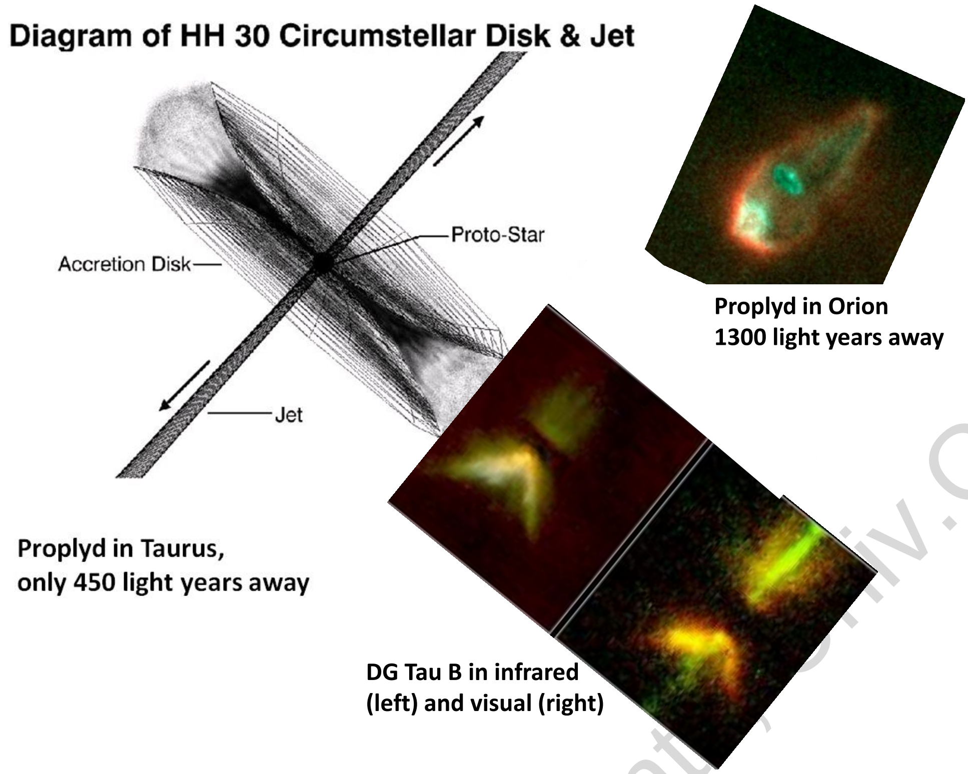

Figure 1.9. Schematic illustration of how material at the same time flows in through the disk toward a proto-star and out through bipolar outflow jets, and corresponding HST images of the same phenomenon in low mass starforming region Taurus-Auriga and high-mass starforming region Orion. Left photo of DG Tau B is in infrared light (from the IR NICMOS camera), while the right photo and the Orion photo are in visual light (taken with the WFPC2 camera). The star itself is obscured by a dense disk near the center of the images, so we don't see the star itself but only the surrounded outer disk or infalling material illuminated by the hidden star. In the visual image we additionally see the bipolar jet originating as a consequence of matter being accreted onto the star through the disk. In the Orion proplyd we also see a surrounding droplet formed envelope formed by the wind from the young high mass stars which are lacking in the small Taurus region.

"Is the Orion starforming process the standard", and "Did our own solar system follow the standard scenario?"

Not all starforming regions are large and massive as M42, M16 and M20. Taurus-Auriga is an example of a nearby rather small so-called T-association, which is a class of starforming region that typically contain only a few hundred T Tauri stars. Taurus-Auriga is 3 times closer by that Orion, only 450 light years away, and the disk forming stage of the T Tauri stars in Taurus-Auriga are therefore better resolved than those in Orion, as is seen in Fig. 1.9. However, even here the solar system would fill less than 0.5". This is the resolution of a very good telescope at a very good site on the ground, and we will therefore not easily see details better than "one solar system per pixel" here either. Fortunately, the disks in Taurus-Auriga are even larger than in M42, so we can see them as more than a dot, but have to remember that what we measure is rather something that a best corresponds to the Kuiper belt and beyond in our solar system, and not the planetary region. Here the CO gas can be traced, and it reveals Keplarian orbits with $V(r) \propto \sqrt{r}$, that the turbulence is small (≈ 0.1 km/s), and that the temperature is consistent with $T(r) \propto r^{-0.65}$ in the outer parts of the disk. Similar results for the inner planet forming parts still await better instruments.

The distribution of stellar masses in Taurus-Auriga is very different from that of M42. Most stars have masses around $0.8\,M_\odot$, there are no high-mass stars, and also very low-mass stars are absent. While the Taurus-Auriga mass distribution is quite unusual, the mass distribution of new-born stars in Orion is quite similar to the mass distribution of stars in our Galaxy in general, lending support to the idea that the bulk of stars in our Galaxy probably are formed in Orion-like clusters, and not in Taurus-Auriga-like regions. The fact that the high-mass stars are absent, also means that there is no strong UV field, no associated HII nebula, no photo-ionization front, and therefore no characteristically droplet shaped proplyds as in Orion (compare Fig. 1.9 from Taurus-Auriga with the proplyds from Orion in Fig. 1.4 and the inset in Fig. 1.9).

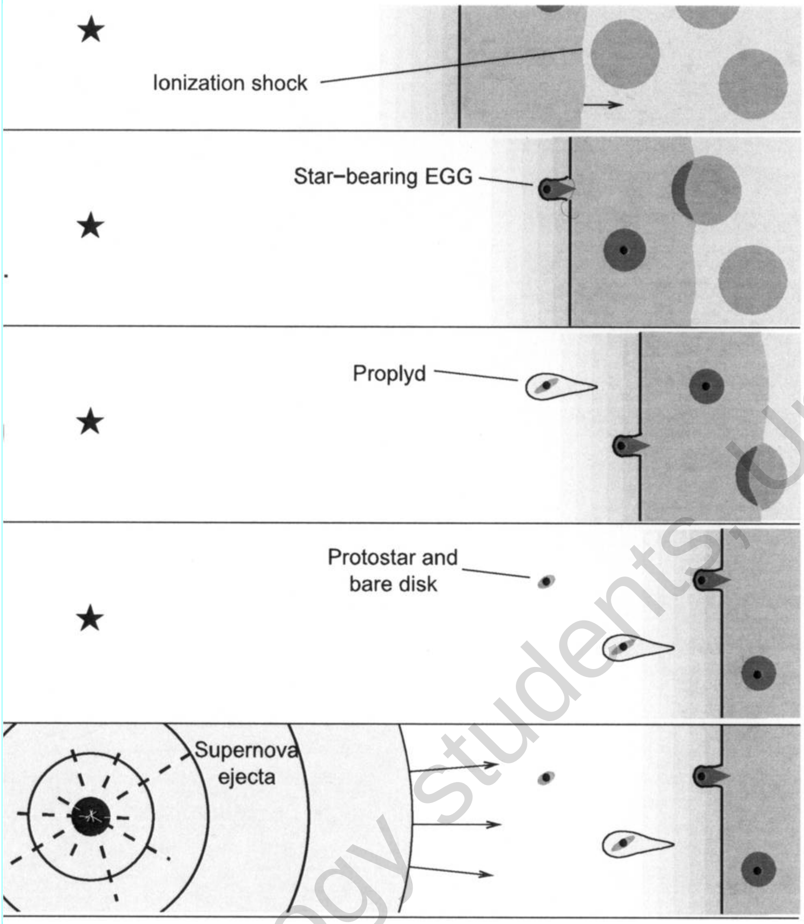

Figure 1.10. A possible sequence of starforming events in the boundary region of the outward moving shock- and ionization front, between a surrounding molecular cloud and an embedded HII cavity. This could represent the evolution of the Orion nebula, and it might also represent the cluster in which our own solar system formed. First (upper panel) one or a few high-mass stars are formed inside the molecular cloud. The UV radiation from these form an HII cavity in the cloud. An ionization front and accompanying shock front expands the cavity into the molecular cloud, hereby triggering collapse of cloudlets on its way (second panel). While the ionization front expose the collapsing cloudlets, they become visible as EGGs (panel 2 and 3), which are further eroded to become proplyds once into the open cavity (panel 3 and 4). After a while ($\sim$100,000 years) the proplyd disk is eroded down to 50 AU radius (panel 4 and 5). On this same timescale one or more of the high-mass stars will now have lived through their full life cycle, and explode as supernova(s). Radioactive material from the supernova explosion(s) spread through the HII cavity, and condensed dust grains will penetrate the forming protoplanetary disks of the low-mass stars, hereby being included into the first planetesimals (last panel).

The conditions for planetary systems under formation in Orion-like clusters and in Taurus-Auriga-like clusters, are certainly very different from one another, and yet we can only speculate about the effect the differences will have on the planets that result from each of these types of environments. Surveys of a large region of the solar neighbourhood, out to a distance of 2 kpc, indicate that about 80% of all stars formed in Orion-type clusters, and only a small fraction in Taurus-Auriga type clusters. Meteoritic evidences indicate that our own solar system formed in an Orion-type cluster. The high-mass stars (with $M > 8M_\odot$) that are present only in Orion-type clusters will have a lifetime of a few million years. Already during the T Tauri phase of the low-mass stars, the high-mass stars will therefore end their lives in supernova explosions. I.e., while the planet-forming disk is still present around the solar-type protostars under formation, the high-mass stars will already have lived through their complete life-cycle, exploded as supernovae, and spread their newly synthesized heavy elements throughout the HII region and into the planet forming T Tauri disks. Our solar system show evidence of newly synthesized radioactive elements from one, or likely more, supernova explosions that took place in the solar vicinity during the solar T Tauri phase. The radioactive energy from these elements may have played a central role for the formation and evolution of the terrestrial planets, as we will return to in a later chapter. Figure 1.10 summarize the evolution in an Orion-type cluster, and may as such represent the first stages of the formation of our own solar system.

In the extreme end of known present-day clusters, is the rich Scorpius-Centaurus OB association. This is a cluster of the same character as the Trapezium cluster in Orion, but much richer in high-mass stars. Most of the low-mass stars in the cluster are still in their T Tauri phase of formation, but ~ 20 supernovae will already have exploded in the cluster. Planets under formation around the T Tauri stars in Scorpius-Centaurus will have experienced a nearby supernova explosion and accompanying impulse of radioactive elements on average every 500,000 years. The new-born planets here could already be gleaming with radioactivity at their surface.

1.3 The standard theory for the formation of the solar system

When developing *the standard theory for planetary system formation*, observations of exoplanets and stellar disks guide the modeling. We have summarized some of the observations of starforming regions above (and we will discuss what we can learn from exoplanets in the next chapter), and drawn the conclusion that excess of infrared radiation among many young stars can be interpreted as protoplanetary disks, and even if we in most cases cannot see the disks directly, we can still get a fine understanding of the development of the disk structure from just measurements of the infrared excess, as was summarized in Fig. 1.8. Fig. 1.11 shows as function of age the percentage of stars that have detectable disks, as estimated from their excess of infrared radiation, and how much mass is accreted (from the disk and onto the star).

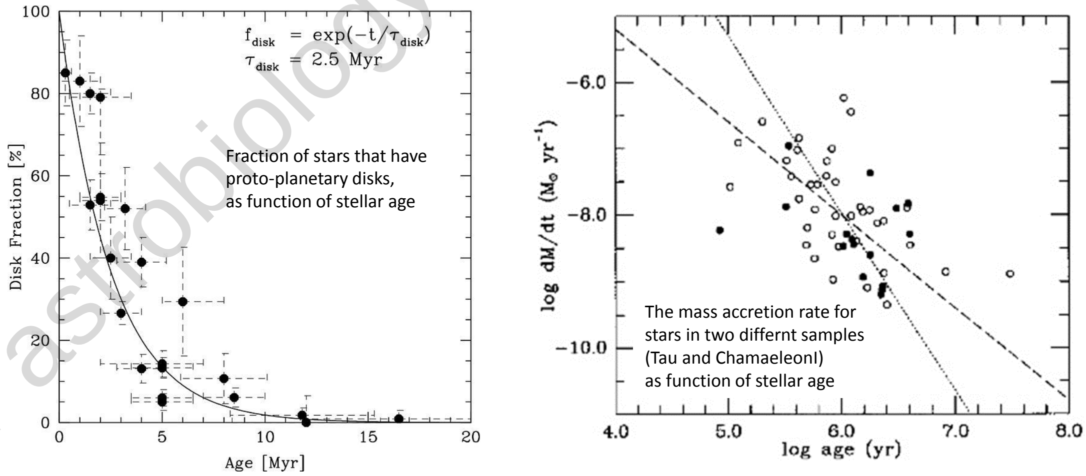

Figure 1.11. Proto-planetary disks are associated with the planetesimal formation epoch of the first few million years of the star formation. Left panel shows the percentage of nearby stars that have detectable disks, as function of stellar age. Right panel shows the accretion rate (of material from the disk to the star) as function of stellar age.

From the observations described above, and in particular Fig. 1.8 and 1.11, we are now able to summarize our basic understanding of the solar system formation. The formation started with a rotating fragment of an interstellar cloud that collapsed and flattened out due to effects of the centrifugal force. The central part of the cloud became the Sun, and the planets and the other solid bodies in the solar system formed out of the disk. In the most popular scenario, the inner planets formed from condensable material in the cloud much the same way as snow condense out of the Earth's atmosphere. The formation of the outer planets start the same way, but when the gravity of the condensed material gets high enough, the surrounding gas will collapse onto the solid planet, which will end up becoming the core of a gas planet like Jupiter and Saturn. In an alternative scenario, planets form directly as a gas collapse in the nebula. Some authors attribute only the formation of the outer planets to this mechanism, while other authors attempt to explain all the planetary formation in this way. In both scenarios, the initial phases are the same.

In the next chapter we will seek a deeper understanding of the physical mechanisms behind the process that transform an interstellar cloud into a star and an accompanying system of planets. The aim is to develop the basis for understanding how diverse planetary systems are, and hence the basic environment for the development of life, and what determines it. The basic observational features have already been summarized in Fig. 1.8 and Fig. 1.11, and can be described more or less as the following four phases:

1: **Formation of the nebula:**
Duration: ≈ Two million years in a simple free fall collapse model – could be considerably longer if so-called ambipolar diffusion across the magnetic field lines is an important process, and it could be considerably shorter if an initial inward directed velocity field is somehow imposed or if the field lines are oriented such that the collapse is directed along the lines.
Observational object: A molecular cloud core. Can usually not be seen in visual light, but might be observable in infrared or mm wavelengths. ∼YSO 0.
Physical characteristics: A part of the original molecular cloud separates from the rest and collapses inside out, with a disk forming around the central denser region. By the end of this period hydrostatic equilibrium is established perpendicular to the disk, and material is dissipating inward through the disk. In the case of the Sun the disk is called *the primitive solar nebula*, or sometimes just *the solar nebula*, and for other stars correspondingly *the stellar nebula*. The term protoplanetary nebula is sometimes used, and is in principle a good term, but potentially confusing, because it is also used to describe the first stages of the formation of a planetary nebula, which has nothing to do with planets, but is the circumstellar envelope around low-mass stars at the final stages of their evolution.

2: **Major dissipation of the nebula; formation of the Sun:**
Duration: ≈ 50,000 to 100,000 years.
Observational object: FU Orionis stars.∼YSO I.
During the FU Orionis high-state (lasting typically a few decades) inflow is $10^{-4}M_\odot$/yr, and outflow through the bipolar jets is $10^{-5}M_\odot$/yr. The low-state (i.e., lower in-flow) periods are longer. Most of the Sun's mass is accumulated during this phase, but the Sun keeps contracting after the FU Orionis stage.
Physical characteristics: A small core in hydrostatic equilibrium forms in the center of the nebula. A quasi equilibrium exist in the disk, where the amount of accreted in-falling material approximately equals the material dissipated inward. CAI (=Calcium-Aluminum-rich Inclusions, which are believed to be among the first solid material to have condensed in the solar nebula) are formed by the end of this phase.

3: **Terminal accumulation of the Sun:** Duration: ≈ 1 to a few million years. ∼YSO II.
Observational object: Classical T Tauri stars (CTT). ∼YSO II.
Physical characteristics: The accumulation is much slower than in phase 2, beginning at a rate of $\approx 10^{-7}M_\odot$/yr, quickly falling to 10^{-8} and ending at $10^{-9}M_\odot$/yr by the end of the CTT phase. Dust condense and accumulate to the mid-plane of the disk, and planetesimals are formed during this phase. Jupiter may have formed already by the end of this phase.

4: **The residual static nebula:** Duration: ≈ 3-30 million years.

Observational object: Weak line T Tauri stars (WTT). ∼YSO III.
Physical characteristics: The Sun settles slowly toward the state of stable hydrogen to helium burning (ZAMS) by the end of the WTT phase. Beyond Saturn, UV radiation removes the residual gas. Further in, possible residual gas most likely settles onto the Sun. The inner planets form from colliding planetesimals during this period, but the outer planets must have grown big enough in order to trigger a gas collapse before the gas nebula is gone.

1.4 Disks around nearby stars

Most likely, all stars form in clusters. Therefore we find the bulk of the known proplanetary nebulae in young starforming clusters, such as Orion. However, all the major starforming regions are more than 1500 light years away, and no existing facilities can resolve a disk at that distance into details which can give us the necessary observational constraints in modeling how planets form. A few individual young stars at smaller distances, however, have been resolved in more detail, foremost of them the observation of LH Tau in October 2014 when the international ALMA sub-mm facility in Chile was tested in its maximum resolution configuration. It is not completely clear what one sees in the high resolution ALMA image in Fig. 1.12, but most likely the dark regions in the ring formed disk are gaps cleared up by planets under formation. If this is the right interpretation, it shows planet formation to evolve surprisingly fast.

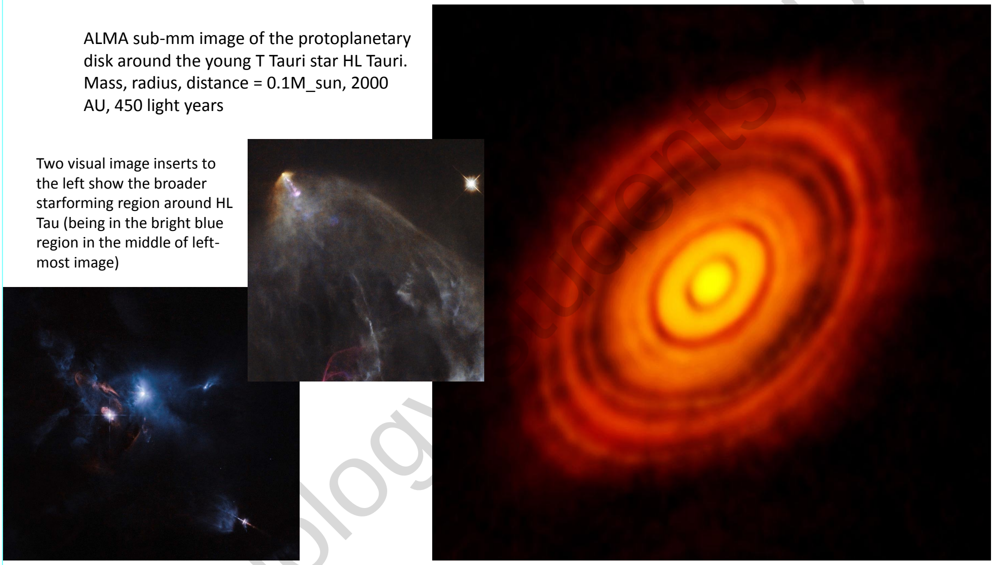

Figure 1.12. LH Tauris is a young T Tauri star, probably less than 1 million years old, surrounded by a huge protoplanetary disk which was resolved in hithero unseen details during the test observations with the ALMA antennas in October 2014 as shown here. Insets show HST images of the surroundings of HL Tau, revealing among other things the huge Herbig-Haro object HH150 originating from a strong jet from the LH Tau disk when the expelled material forms chocks with the surrounding interstellar gas.

Infrared satellites have during the latest decades identified several nearby stars with infrared excess radiation, and such excesses could be due to disks, as already illustrated in Fig. 1.8. However, high-resolution images of these stars have revealed that although the infrared excess in some cases are due to disks, the disks are usually not protoplanetary (such as LH Tauris is). Their masses are far too small to be able to give rise to planets, and often the stars are much older than the disks. Such disks are called debris disks. Debris means "remnants", but they are not remnants in the sense of material left over or still not included into the planets. Rather they are debris of some violent event that destroyed larger objects that were formed at an earlier stage in the now gone protoplanetary disk, such as asteroids and comets. Such disks are not enscrouted in dark starforming molecular clouds (as LH Tau is), because they are no longer associated with the region where they formed. A few of the disks can therefore be seen directly in scattered visual light, by placing an occulting disk inside the telescope (creating an artificial eclipse), such that the bright light from the star is reduced and the dim reflected light from the dust disk can be traced. In most cases it is, however, only possible to trace the far infrared or sub-millimeter thermal radiation from the disk. For example, a gas of T = 30 K has its maximum thermal emission at $\sim 100\mu$m. At an often used sub-millimeter wavelength of 850μm the thermal radiation from a cold disk is still strong, while the intensity of the light from a normal star is almost zero at these wavelengths.

With observations at a sufficient number of wavelengths, it is possible to fit a Planck curve to the measurements, and hereby determine the temperature of the dust that makes up the disk. When this is done, one can also measure the wavelength dependent deviation from the best-fit Planck curve, and since this deviation is strongly dependent on the size of the dust grains (and to some extend their composition and form) one gets an estimate of the dominant dust size in the disk. More refined modeling also gives a dust size distribution and other information about the system, too. In this way it has been realized that most debris disks are of temperatures as low as $\sim$50 K and consist of dust in the 10 to 100 μm range.

Our own solar system has a debris disk, too. It is called the zodiacal disk, because it stretches through the zodiacal constellations in the sky where also all the planets move. It can be seen during clear dark nights with the naked eye as a cone of light rising from the horizon, during the first hours before sunrise (during autumn) or after sunset (during spring). It looks like a huge band of dim light, called the zodiacal light, resembling the Milky Way but at an angle of 60° with the Milky Way band in the sky. The zodiacal light is sunlight reflected on dust grains in orbit around the Sun. However, several effects, which we will discuss in detail later, will remove the dust on timescales of a few million years or less, so the dust grains must continuously be replenished in order that we can still see zodiacal light on the sky. Several theories have been proposed for where the zodiacal dust grains come from, most likely is that they are debris of colliding asteroids in the inner solar system, colliding dormant comets in the Kuiper belt, and the leftover cometary dust-tails from when iceblocks come close enough for the Sun to evaporate their surface layers and press the dust out into the beautiful tails we see as a comet on the sky. Also dust blown into space from the volcanoes of the Jupiter moon Io could possibly contribute to the zodiacal dust. In this sense the zodiacal light is sunlight reflected on cometary dust tails and other dust, resolved, mixed, and on its way to swirl into the Sun.

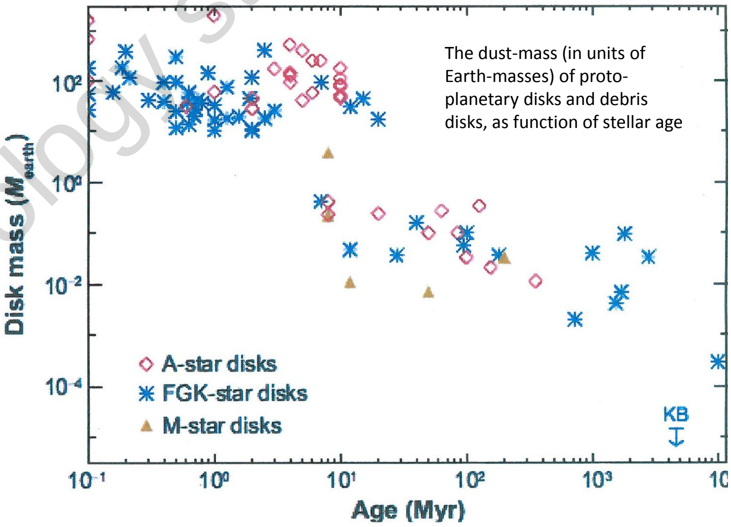

Figure 1.13. Disk dust mass of proto-planetary and debris disks around A, FGK, and M dwarfs, as function of stellar age. Remark the destinct population of proto-planetary disks of $\sim 100 M_\oplus$ and debri disks with 3 orders of magnitude less mass (and yet much lower mass of our Kuiper belt or the zodical light disk).

Although only a dozen of debris disks have actually been imaged, surveys searching for infrared excess in the stellar light, indicate that $\approx$ 50% of all nearby stars younger than 300 million years show signs of such disks. The brightest star to show a debris disk is Vega, and debris disks are therefore sometimes also called Vega-like disks. Only $\approx$ 10% of stars older than 400 million years show Vega-like disks, so whatever cause the debris disks, it is related to something that mostly take place in the early phase of solar systems formation. 700 million years after the formation of our own solar system, a swarm of comets and asteroids raged the solar system, forming all the craters we see on the Moon and all other atmosphere-less bodies in the solar system. The solar system must have had a bright debris disk that could be seen many light years out in space with instruments like those we posses today, in the millenias following this event. Since we don't see such old stars with bright debris disks, it could be that the late heavy bombardment (LHB) that ravaged the solar system 700 million years after its formation is rare, at least this late in the planetary evolution, but if the lifetime of the "LHB debris disk" was short, the 10% visible "late debris disks" may of course instead mean the opposite – that LHB-like events are common. The LHB in our solar system may have delivered the oceans on Earth, and if LHBs can be shown to be rare, it may therefore mean that oceans, and perhaps even water, are rare on otherwise habitable exoplanets.

1.4.1 The β Pictoris disk

The first image to prove that the infrared excess of any nearby star was actually due to a disk, was the now famous 1983 image of β Pictoris (Fig. 1.14 and 2.8), but its disk has hardly anything to do with a protoplanetary system under formation, such as those we see in Orion. The dust that makes up the β Pictoris disk has no more than a few times the mass of the Moon. Yet the disk has a diameter at least 40 times the orbit of Neptune, dust is depleted inside 100 AU from the star, and the innermost 30 AU is a hole. However, even though β Pictoris is not a protoplanetary disk, it still has lots to tell us about the origin of solar systems. The star is estimated to be between 8 and 20 million years old. The first thing the image tells us, is therefore that the formation of the plantesimals was over at that early age; or at least the stage that grew the solid material from dust to gravel size – there is no longer dust left to form planets from inside 30 AU. Second, there is still dust in the outer region, although very little. There are also traces of hydrogen and carbon-monoxide gas left in the disk.

It was long debated whether a seeming vertical wrap in the disk was due to a perturbing planet, orbiting within the disk. Better images (Fig. 1.14) showed that what at first had appeared as a wrap, actually were two disks inclined only 4° relative to one another, as is seen in Fig. 1.14. It is now believed that a perturbing jupiter-sized planet (see the direct image inset in the centre of Fig. 2.8 has "lifted" enough dust up from the main disk to form the secondary disk seen in Fig. 1.14. Yet an additional and detached disk is, quite mysteriously, situated at a distance of 55,000 AU from β Pictoris.

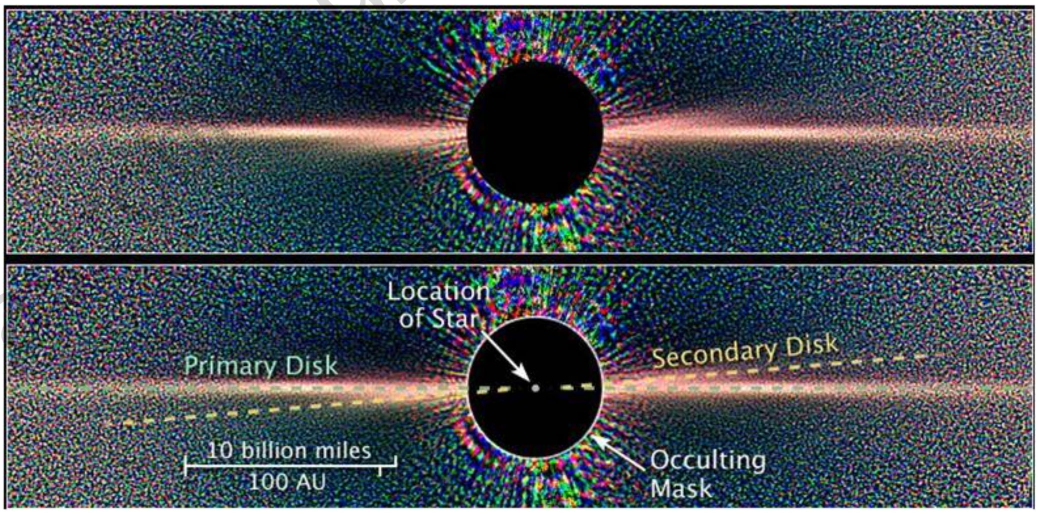

Figure 1.14. The two disks around the star β Pictoris. Both disks are seen almost edge-on, and therefore appear almost like "needles" in this high resolution image from HST.

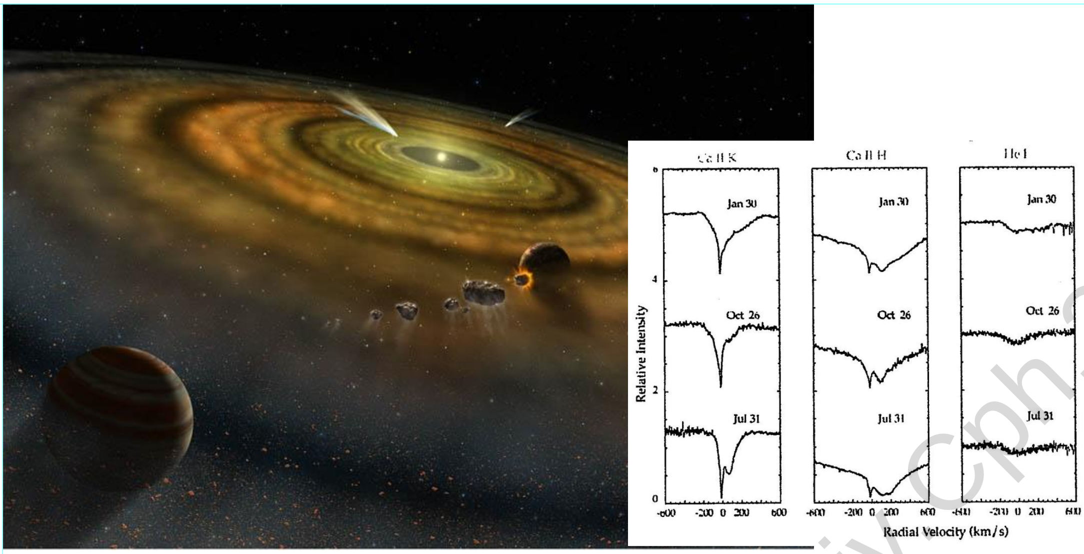

Figure 1.15. Asymmetric transient absorption lines in spectrum of β Pictoris may be indications of evaporating comets, falling in to the star on a daily basis.

An even more surprising observation, is the line shapes in spectra of β Pictoris. Because we see the βPictoris disk edge on (inclination $i > 80°$), we see absorption lines from the disk in the stellar spectrum. The absorption spectrum is composed of a stable component (with a RV identical to the 20km/s of the star) and a variable, almost always red shifted, component of a few km/s and of up to 400 km/s. This is now interpreted as families of comets crashing into the star on a daily basis. Spectral features interpreted to be originating in the inner part of the disk, have similarities to known silicate features in the spectrum of comet 1P/Halley. One could speculate whether the remote dust disk, the traces of gas, and the huge number of star-grazing comets are sides of the same phenomenon, namely traces of a dynamically very active early phase of planetary system formation. If huge ice blocks analogues to comets and Kuiper belt objects in our own solar system, are constantly colliding, it would give rise to the dust in the outer disk, and perhaps also to the small amounts of gas. The estimated mass of the CO gas is ~ 0.01 times the mass of the Moon, and the gas has a temperature of 30 K, which is the temperature where small comets at ~ 100 AU from β Pictoris would evaporate CO from their surface layers. The same comets could be perturbed inward, to eventually collide with the central star, giving rise to the red shifted asymmetries in the spectral lines shown in Fig. 1.15. Do we witness here the same phenomenon that gave rise to the late heavy bombardment in our own solar system when our solar system was 700 million years old, but here at an age of only a few millions of years? One could well speculate whether we are right now witnessing signs of the process that is forming the oceans on habitable inner planets in the newborn β Pictoris system, when colliding comets from the outer regions bring volatiles to the inner system, just as it might have happened in our own solar system 3.8 billion years ago.

1.4.2 Other nearby debris disks

β Pictoris is one of an elusive little class of nearby stars that already in the measurements by the IRAS satellite in 1983 showed strong infrared excess radiation. The disk around β Pictoris turned out to be by far the largest of them, and it was for this reason that its disk was the first to be successfully photographed in visual light. However, the three nearest ones are ϵ Eridani, Formalhaut, and Vega, and their disks are surprisingly different from one another, and from the one around β Pictoris.

ϵ Eridani is only 10.5 light years away – 150 times nearer by than the stars in Orion, and almost 7 times closer by than β Pictoris. Obviously we see details that would have been impossible to trace in the Orion-disks. A striking difference from the β Pictoris disk is that it is much smaller, by at least a factor of 10. The small size has so far made it impossible to trace it in visual light, but Fig. 1.16 show a sub-millimeter image at 850 μm, and illustrate the position and size of the disk, together with its known 1.5 Jupiter-mass planet orbiting at a distance of only 3.4 AU from the star; i.e. far inside the observed “hole” in the ring. The planet and the disk are in the same plane. The disk can be estimated to consist of dust that has an average size of $\approx 30\mu$m. As we will calculate in a later chapter, such small dust particles are forced to disappear from the disk over timescales of $\sim$10 million years or less. Since the best estimate of the age of the star is $\sim$700 million years (by far the oldest of the four), the lifetime of the disk is much shorter than the age of the star. Again, it is therefore not a protoplanetary disk out of which new planets are formed, but a debris disk – a continuously replenished disk, formed from debris of colliding comets and asteroid-like objects. Further, it is estimated to contain only 0.07 times the mass of the Moon, so obviously no planets could be created from this small amount of material.

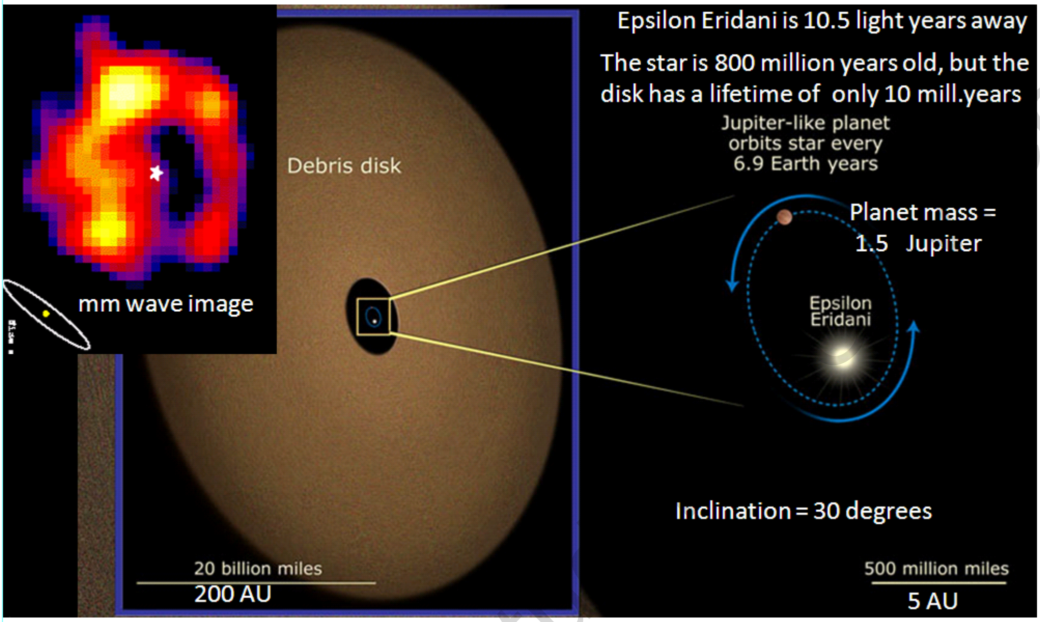

Figure 1.16. The observed disk around the nearby star ϵ Eridani (inset at left), shown together with an illustration of how the known 1.5 Jupiter mass exoplanet may orbit within the cavity in the disk.

The debris disk around ϵ Eridani may well be dust from comets, but the disk is at the outer edge of what corresponds to the Kuiper belt in our solar system, and the star is too week (1/3 of the solar luminosity) to simply evaporating cometary material at that distance. The most luminous part of the disk is two clumps at a distance of more than twice Neptune’s orbit. We can envision that giant ice blocks at this distance can have collided with one another, and now reveal their existence 10.5 light years out in space by the clouds of dust and other debris that was the result of the collisions. The clumpy nature of the observed disk, may indicate this is the case, but the bright clumps could also be areas where dust are captured by the gravity of one or more unseen planets. Models indicate that a 30 Earth-mass planet in a 40 AU orbit around ϵ Eridani, could do the trick (but again the dust to be captured need to have been created somewhere not too long time ago).

Vega is the brightest of the four stars, of spectral type A0V and only 25 light years away. One of the surprising things about the Vega debris disk is that its center is not on the star, but midway between the star and a bright clump of dust almost 100 AU from Vega. The last of the four bright stars is Formalhaut. It is in fact a binary star, with a dim K4V stellar component orbiting the bright A3V star at a distance of 55,000 AU. The ring itself was only successfully imaged in visual light 20 years later than β Pictoris, and it appears as a fairly narrow ring of $\approx$1.4 lunar masses of dust distributed smoothly between 130 and 160 AU from the star. In infrared and sub-millimeter wavelengths the tilt of the ring ($\approx 20^\circ$ from edge-on) makes the ring look rather like two clumps of material on either side of the star, but it is rather an observational effect related to the instrumental resolutions.

Another interesting disk is the one around the star HD 10647. Just like ϵ Eridani, also HD 10647 has an exoplanet discovered with the radial velocity technique. This time it is a 0.93 jupiter-mass planet orbiting at

	βPictoris	Formalhaut	Vega	AU Mic	ϵEridani
Stellar type	A5 V	A3 V	A0 V	M1 Ve	K2 V
Stellar mass [$M_\odot$]	1.75	1.75	2.0	0.5	0.8
Luminosity [$L_\odot$]	8.7	13	60	0.1	0.3
V magnitude	3.86	1.16	0.03	8.65	3.73
Distance [ly]	62.9	25.1	25.3	32.3	10.5
Stellar age [Myr]	8-100	100-300	300-400	8-100	500-1000
Disk radius [AU]	30-1500	100-150	80-800	50-210	50-80
Disk inclination [deg]	$>$80	70	5	$\sim$90	20
Dust mass [M_{lunar}]	$\sim$3	1.4	0.2-0.7	$\sim$1	0.07
Disk temperature [K]		40	80		35
Dust size [μm]	10	100	70		30

Table 1.1. Properties for five main sequence stars with disks. βPictoris and AU Mic belong to the same moving group, and are probably formed in the same cluster (together with 20 other known stars of the moving group). The disk radius list the estimated inner and outer sizes, respectively.

2.03 AU in a rather elliptic (e = 0.16) orbit. The disk itself is much further out than the planet, stretching from 75 to 130 AU from HD 10647.

The most solar-like star to have a debris disk, is the G2V star HD 107165, which is almost 100 light years away and posses a disk of almost a quarter of an earth-mass showing up at a distance between 60 and 185 AU from the star. The oldest star to show a disk is HD 53143, which is a dim, $\sim$ 1 Gyr old K1V star, 60 light years away. Table 1.1 summarize the properties of some of the slightly more than 20 nearby stars that are known to have associated debris disks.

2

The Formation of Solar Systems

— an analytical approach

2.1 The collapse phase: Infrared molecular cloud cores.

The collapse phase is the first stage in any theory wanting to explain the transformation from an interstellar cloud to a solar system. It includes the contraction from an interstellar cloud, the formation of a small proto-Sun in the center of the cloud, and the formation of a disk around the proto-Sun. We will see in this section that all that is required to qualitatively understand this formation phase is the virial theorem, the gravitational force, and the centrifugal force. These simple tools will allow us to estimate the timescale for the collapse, and to understand the reason for the collapse, for the concentration of the solar material in the center, and for the disk formation. In reality, of course the quantitative development is more complicated and requires a full numerical magneto-hydrodynamic simulation supported by observations,. However, we can reach a deep understanding of the basic processes, by investigating relatively simple analytical physics.

Consider first an interstellar cloud, and let it be described as N particles (atoms, molecules, dust-grains) at individual positions r_j relative to the center of mass, and with individual masses m_j, $j = 1, N$. The moment of inertia relative to the center of mass is then

$$\begin{aligned} I &= \sum_{j=1}^{N} m_j r_j^2 \Rightarrow \\ \ddot{I} &= 2\sum_{j=1}^{N} m_j \dot{r}_j^2 + 2\sum_{j=1}^{N} m_j \ddot{r}_j r_j \\ &= 4\sum_{j=1}^{N} \frac{1}{2} m_j v_j^2 + 2\sum_{j=1}^{N} F_j r_j \Leftrightarrow \\ \frac{1}{2}\ddot{I} &= 2K + \Omega \end{aligned} \tag{2.1}$$

where K is the kinetic energy of the system and Ω is the potential energy, defined as the sum of these quantities ($\frac{1}{2} m_j v_j^2$ and $F_j r_j$, respectively) for the N particles.

For a central force Ω is negative, and we see that if $|\Omega| > 2K$, $\ddot{I}$ is negative; i.e. the system will contract (or strictly have $\ddot{r}_j < 0$), while if $|\Omega| < 2K$, $\ddot{I}$ is positive and the system will expand. The system is in equilibrium if $\ddot{I} = 0$, i.e. if

$$2K + \Omega = 0 \tag{2.2}$$

Equation 2.2 is called the Virial Theorem. It states that for a system in equilibrium the time-average of the sum of the kinetic energies of the particles in the system equals the time average of their potential energy. This is a more general form of the case where thermal pressure balances the gravitational energy. We can in principle include any other force than gravitational force in the expression for the potential energy (since we have only used the definition that the potential energy is the energy required to move a particle of mass m_j the distance r_j under the influence of the force F_j in Eq. 2.1). Similarly, the kinetic energy can include turbulence, shock-waves, translational movements, and any other movements since we have only expressed it as $\frac{1}{2}m_j v^2$. In order to clearly see the physics behind the stability and collapse of the gas clouds that will form solar systems, we will in the following think of gravitational force and thermal energy only.

Under these assumptions the kinetic energy of a particle is

$$E_k^j = \frac{3}{2}kT \tag{2.3}$$

If the average mass of the particles is μ, and the total mass of the system is M, then there are $n = M/\mu$ particles in the system, and the total kinetic energy is therefore

$$K = \sum_{j=1}^{n} E_k^j = n\frac{3}{2}kT = \frac{3kTM}{2\mu} \tag{2.4}$$

Similarly the potential energy of the particle j is

$$E_p^j = -\frac{GM_j m_j}{r_j} \tag{2.5}$$

where the mass M_j inside r_j is assumed to be symmetrically distributed such that particle j feels it as being in the center (e.g. a sphere or an ellipsoid). This equation can be integrated to give

$$\Omega = \sum_{j=1}^{n} E_p^j = -\sum \frac{GM_j m_j}{r_j} = -\alpha\frac{GM^2}{R} \tag{2.6}$$

where M and R are the total mass and radius of the system (the gas cloud). α is a parameter which describes the distribution of the density as function of distance. For a homogeneous sphere α = 2/5. For material concentrated more toward the center it is smaller. We will often use the value 2/5 for the interstellar clouds, while we will use the solar value for the final stages of the system (the proto-Sun, the proto-planetary disk etc).

If the temperature is very high, we expect that the particles will fly apart due to their thermal motion (Eq. 2.4), whereas a high gravity (combination of density and mass of the system) will contract the system (Eq. 2.6). For any combination of two of the parameters density, temperature and mass, we can therefore calculate the critical value of the third parameter which just allows for a stable system that neither collapse nor expands. For a given combination of temperature and density, the critical mass of a system only under influence of gravity and thermal motion, is called the Jeans mass. We can calculate the Jeans mass by substituting Eq. 2.6 and 2.4 into Eq. 2.2, which gives us

$$\begin{aligned} 2K + \Omega &= 0 \Leftrightarrow \\ \frac{3kTM_J}{\mu} &= \alpha\frac{GM_J^2}{R} \Leftrightarrow \\ M_J &= \frac{3kTR}{\alpha\mu G} && (2.7) \\ &= \left(\frac{375}{4\pi}\right)^{1/2}\left(\frac{k}{\mu G}\right)^{3/2}\left(\frac{T^3}{\rho}\right)^{1/2} = 5.1\,10^{-13}\sqrt{\frac{T^3}{\mu^3\rho}} && (2.8) \end{aligned}$$

where we have used $M = \frac{4}{3}\pi R^3\rho$ (such that $R = (\left(\frac{3M}{4\pi\rho}\right)^{1/3}$) and α = 2/5 to derive the last line. In the derivation of the numerical constant (5.1 10^{-13}), μ is assumed to be in units of g per particle, and ρ in g per cm^3.

In a cold gas of 72% hydrogen by mass and 28% helium (or a solar composition gas), $\mu \approx 2.3\, m_u$ = 3.86 10^{-24} g per particle, and therefore
$M_J = 3.86\, 10^{-24}\, (T^3/\rho)^{1/2} \approx 30\mathrm{M}_\odot$ for T = 10 K and $\rho = 10^{-21}$ g/cm^3.
In other words, at the conditions of present day ISM (Inter-Stellar Medium), interstellar clouds of somewhat above 30M$_\odot$ are gravitationally unstable and will collapse (and fragment) if they are given a small push (e.g. from a galactic density wave). This is why young stars form in groups (open clusters or unbound stellar groups) of typically 100 stars. In the early phases of the evolution of our Galaxy, the density and temperature conditions were such that much larger interstellar clumps of typically 100,000 solar masses collapsed together to form the globular clusters of stars.

The fact that our Galaxy is filled with interstellar clouds immediately tells us that the clouds must fulfill the Jeans criteria (i.e., be stable), or at least that the forces that tend to collapse the clouds are in balance with the forces that tends to expand the clouds. Beside those forces that are included in deriving the Jeans criteria (gravity and thermal pressure), forces that prevent the clouds from collapsing, include the interstellar magnetic field, turbulence in the cloud-gas, radiative pressure, and rotation. These forces ensures that the collapsing clouds are a bit larger than the 30M$_\odot$ we derived above from the pure Jeans criterion. Forces that prevent the clouds from expanding include the pressure from the surrounding (dilute but hot) interstellar medium.

However, the fact that the universe is cold and consists mainly of hydrogen and helium, makes the clouds only marginally stable. This fact (together with the expansion of the universe) assures that the material in the present state of the universe will develop beyond the simplicity of hydrogen and helium and into heavier atoms and eventually planets. Had the universe been as hot as in the early phases where the gas was ionized, or had the gas consisted mainly of e.g. carbon, silicon and oxygen which rapidly forms dust at low temperatures, or had hydrogen molecules and helium not had the symmetry that makes them dipole inactive, then the interstellar clouds would have been largely opaque with dramatic consequences for the further development of the material. Such (hypothetical) clouds would be much more stable against collapse, with the consequence that stars and planetary systems not easily would form.

To understand why clouds of transparent gas collapse much easier than clouds of opaque gas, consider a cloud which at time t_0 has started contracting, i.e.

$$2K_0 + \Omega_0 = \frac{1}{2}\ddot{I}_0 < 0 \tag{2.9}$$

If the cloud is opaque, then to a first approximation potential energy released during the contraction will not escape to the surrounding space (adiabatic contraction), but instead be converted to kinetic energy (heat); $\Delta E = \Delta K + \Delta\Omega = 0$. At time t_1 shortly after t_0 the kinetic and potential energy of the system will therefore be

$$K_1 = K_0 + \Delta K; \quad \Omega_1 = \Omega_0 + \Delta\Omega; \quad \Delta K = |\Delta\Omega| \tag{2.10}$$

and therefore

$$2K_1 + \Omega_1 = \frac{1}{2}\ddot{I}_0 + \Delta K \tag{2.11}$$

Since ΔK is positive the contraction will stop (or almost stop), and the cloud will again find itself in equilibrium, and we would not form stars and planets of the cloud. Note that a cloud strictly in virialized equilibrium and completely unable to exchange energy with the surroundings cannot contract (the only solution to $\Delta 2K + \Delta\Omega$ = 0 and $\Delta K + \Delta\Omega$ = 0 is $\Delta K = \Delta\Omega$ = 0).

However, most clouds are (at least during the beginning of the contraction) quite transparent (i.e. optically thin), and therefore the created thermal energy during the contraction will quickly be radiated away (isothermal contraction), and in this case

$$K_1 = K_0; \quad \Omega_1 = \Omega_0 + \Delta\Omega \tag{2.12}$$

and therefore

$$2K_1 + \Omega_1 = \frac{1}{2}\ddot{I}_0 + \Delta\Omega \tag{2.13}$$

and since $\Delta\Omega$ is negative, the collapse will continue even more vividly until the material gets dense and opaque enough that the radiation is no longer immediately radiated away.

In practice some radiation is of course always radiated away, and therefore the system will always contract with some speed. This is an uttermost consequence of the fact that the universe is expanding. Since the universe contains a finite amount of energy, the energy content per volume space is constantly decreasing, and the second law of thermodynamics will therefore always demand a flow of energy from matter and into space. This flow has to be followed by a gravitational contraction (or the loss of some other energy source like rotation, re-crystalisation, or nuclear energy) according to the Virial theorem (in the general form of Eq. 2.1). Nuclear energy can for a limited period (however long it may be compared to our biological time-scales) stop (then we call the object a star) or slow down (then we call it a brown dwarf) the contraction by adding an energy input, but at some time the nuclear energy is used up, and the gravitation will immediately be back in action following the virial theorem. Jupiter is contracting 1 mm per year which gives energy enough to radiate two times more energy than it receives from the Sun.

The time scales for the contraction can be computed once we know the physical and chemical state of the object. Particularly simple is the isothermal case where no thermal energy is accumulated during the contraction. A small test particle of mass m at the "surface" of a homogeneous sphere of mass $M = M(r_0, t = 0)$ will have a potential energy $E_p(t = 0) = -GMm/r_0$ and a kinetic energy $E_k(t = 0) = 0$. After time dt when the sphere has contracted to $r_0 - dr$ (i.e. $M(r_0, t = 0) = M(r_0 - dr, t = dt)$, the potential energy of the test particle will be $E_p(t = dt) = -GMm/(r_0 - dr)$ and $E_k(t = dt) = \frac{1}{2}m(dr/dt)^2$. Therefore

$$-\frac{GMm}{r_0} = -\frac{GMm}{r_0 - dr} + \frac{1}{2}m\left(\frac{dr}{dt}\right)^2 \tag{2.14}$$

$$\frac{dr}{dt} = \left(2GM\left(\frac{1}{r_0 - dr} - \frac{1}{r_0}\right)\right)^{1/2} \tag{2.15}$$

which can be solved to give the time, $t_{\rm ff}$, it takes to contract from radius r_0 to $r = 0$

$$t_{\rm ff} = \left(\frac{3\pi}{32G}\right)^{1/2}\left(\frac{1}{\rho}\right)^{1/2} = \frac{2.1\,10^3}{\sqrt{\rho}}\text{sec} = \frac{6.7\,10^{-5}}{\sqrt{\rho}}\text{years} \tag{2.16}$$

where we have expressed the homogeneous sphere as $M = \frac{4}{3}\pi r_0^3 \rho(t = 0)$. Since it is assumed that nothing counteracts the contraction (no magnetic field, no pressure gradient, no nuclear energy production, no rotation, etc) this contraction is a free fall and $t_{\rm ff}$ is therefore called the free fall time-scale. Of course no real contraction will take place over this short time-scale, because the contraction always at least will be slowed down by some arising pressure gradient. However, the contraction will start like this when the density is low, and $t_{\rm ff}$ therefore gives an impression about the contraction from very large scale to a smaller size where the isothermal assumption is no longer good. I.e. the contraction timescale from the optically thin to the optically thick phase, according to Eq. 2.9 to 2.13.

We notice that the time-scale is shortest for dense material (because the gravitational attraction is largest when the particles are close to one another) and independent of the initial radius r_0 (because every volume contracts). A typical interstellar cloud has a density of 10^{-21} g/cm^3, and the free fall time-scale is therefore 2 million years according to Eq. 2.16. This means that if an interstellar cloud is hit by a galactic density-wave, or the shock waves associated with the ionization front in M42 or M16, or something else which is able to temporarily increase its density above the equilibrium value given by Jeans' criteria Eq. 2.7, then the cloud will collapse from its original size to a size where radiation no longer escapes freely, over a time-scale of $t \approx t_{\rm ff} \approx 2$ million years.

We also notice from Eq. 2.16 that if parts of the collapsing cloud happens to have at any time a larger density than other parts, this subsystem will collapse faster than the rest of the cloud. Therefore the cloud

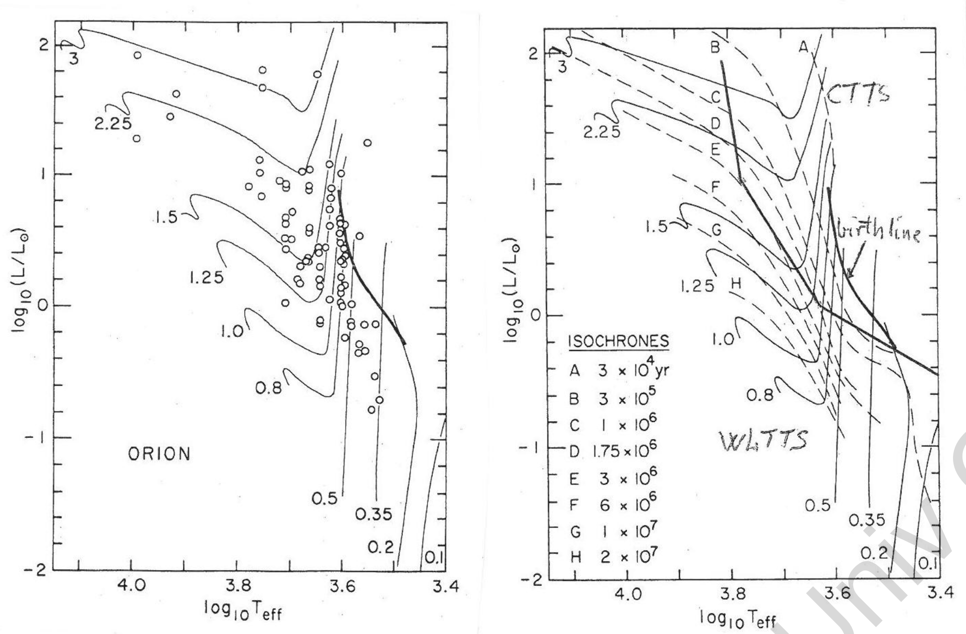

Figure 2.1. Left panel: the position in the HR diagram of young stellar-like objects in Orion (open circles) together with theoretical evolutionary tracks pre-main sequence objects of 3, 2.25, 1.5, ..., 0.1 solar masses. Since the YSO move downward in the diagram while contracting, the upper envelope in the diagram represent the place where YSO first gets visible through the surrounding cloud, and hence represent the "birth line". Right panel: Same as left panel, but with isochrones indicated (as dashed lines) of ages 30,000, 300,000, ..., 20 million years. Also shown is the division line between region of observed classical T Tauri stars (CTTS) and weak-line T Tauri stars (WLTTS).

will have a tendency to fragment during the collapse. Finally we notice from Eq. 2.16 that when the central regions become denser than the outer regions, the central regions will contract quicker. A free fall collapse is therefore an "inside-out" collapse, where the central parts collapse fastest and will drag the outer slower collapsing parts inward (through the disk). We will expect that once the proto-solar system starts forming from a collapsing interstellar cloud, a small dense inner region will quickly form with material from outside streaming into the system for some time after.

During the beginning of the collapse phase a small dense core – a protostar – is established in the center of the collapsing cloud (due to the inside-out collapse). Such a core might initially be just 1/100 of a solar mass. The cloud is still optically thick, and the protostar is not yet visibly (at least not in visual light). During the collapse a disk will form around the central core, because of the centrifugal forces and the conservation of angular momentum, and by the end of the period hydrostatic equilibrium will have established itself perpendicular to the plane of the disk. At this stage we will call the disk *the primitive solar nebula* (in the case of our solar system, or *the primitive stellar nebula* in the general case).

The position in the HR diagram where the object first time gets visible is called the (stellar) birth line, which for the Sun in some calculations is at R≈4-5$R_\odot$, T_{eff}≈4600 K, and L≈8$L_\odot$. Again the terminology is a bit confusing because of the definition of a star as an object where a static balance is reached between inward gravitational force and the outward force exerted by a stable hydrogen-to-helium nuclear fusion in the center. We will therefore also say that the star is born when it for the first time fulfill this criteria, which is at the line in the HR diagram called the zero-age-main-sequence. I therefore prefer the term birth line rather than stellar birth line for the place when a star-like object for the first time gets visible. It is the YSO (young star-like object) that for the first time gets visible here, and it would be even better (but not used) to call it the YSO birth line. Fig. 2.1 shows the position in the observed HR diagram where young stellar-like objects for the first time in their evolution get visible. When we see many more stars in the lower middle panel (IR light) of Fig. 1.3 than in the corresponding lower left panel (visual light), it is because gas clouds are optically thinner at longer wavelengths, so that we see a larger volume of the nebula and stars somewhat above the birth line in Fig 2.1 too.

Some calculations include so-called ambipolar diffusion in the collapse phase, which takes effects of the magnetic field carefully into consideration, and which will slow the contraction considerably, making the initial contraction last considerably longer than the 2 million years computed above. Other computations consider an initial velocity field which could shorten the collapse phase to as little as 100,000 years. Most likely it depends critically on the environment the cloud collapses in, such as possible nearby ionizing stars, the general magnetic field, exploding supernovas, surrounding turbulence in the molecular cloud, etc, and all these details may well affect which kind of planets, if any, eventually will be formed.

Models with a too long collapse phase, have problems with the meteoritic data. Radioactive dating of the CAI and chondrule inclusions in primitive meteorites, indicate that these must have formed quite short time after the solar nebula separated from the general interstellar medium, and it is hard to envision these to have formed earlier than during the finalization of the initial collapse phase. Even the free fall time scale of $t_{\rm ff}$ = 2 million years we calculated above is uncomfortably long seen in the light of these data. If the radioactive ^{26}Al ($\tau_{1/2}$=700 000 yrs) is to have played an important role in the initial melting of the planets, it imposes severe constraints on our model that 3 half-life times may have passed, alone from the separation of the nebula from the IMS to the end of the collapse phase. Even more severe is the measured abundance of the ^{41}K/^{40}Ca ratio in some CAI which seems to require unreasonably high abundances of radioactive ^{41}Ca (which decays to ^{41}K with a half life of only 100,000 years) in the collapsing nebula, if 2 million years (or more) were to pass from its production in a supernova to its inclusion in CAIs 20 halflives later. Did something very peculiar happen in our solar system in relation to these time scales, and was it important for the later developments? – we will return to these issues in the description of the meteorites.

2.2 Formation of the proto-Sun, and major dissipation of the nebula.

FU Orionis is the earliest, and quite short, stage of the low-mass stellar formation which is observable in the visual. It must be the first evolutionary stage following the collapse, in any theory for the formation of the solar system. The average net inflow is $\approx 10^{-5}$M$_\odot$/yr, and the life time of this stage is $\sim$50,000 to 100,000 years. $\sim$ 1M$_\odot$ flows through the disk during this stage. The bulk of the Sun is builds up here, but the proto-Sun at the end of the FU Orionis phase is still very big, and will shine bright for another $\sim$100 million years due to energy gained from a slow contraction. Jupiter and Saturn is still in this slow contraction phase (making them emit $\sim$3 times more energy than they receive from the Sun), because smaller gas-clumps contract slower. In this sense they are “miniature solar systems” in their FU Orionis phase, but with the gas disk in which their moons formed gone long time ago, and they will of course never become stars.

Two major questions to discuss for the FU Orionis phase of YSO are how the material moves through the disk to be included in the Sun, without being stopped by an increasing angular momentum, and how, in particular, the forming proto-Sun gets rid of the huge amount of angular momentum one should think it would build up during this phase. In other words:
(1) The planets orbit the Sun in stable orbits because of a balance between the gravitational and centrifugal forces, the nebula flattens out into a disk because of the effect of the centrifugal force, but then why should the disk material fall down onto the proto-sun when the planets don’t ?, and
(2) an ice-skater who performs her pirouettes is often used as an illustration in text books to illustrate the fundamental principle of conservation of angular momentum, but she increases her rotational velocity remarkably just by drawing her arms a factor of a few toward her rotational axis. When a rotating interstellar cloud collapses, it draws the bulk of its mass a factor of millions closer to its rotational axis – it should rotate incredibly fast, swirling around its axis during seconds only; when, where, and how did the Sun lose all this angular momentum ?

During the FU Orionis so-called high-state (lasting typically a few decades at a time) inflow is as high as 10^{-4}M$_\odot$/yr, and outflow through bipolar jets is 10^{-5}M$_\odot$/yr. A small core in hydrostatic equilibrium forms in the center of the nebula in the beginning of this phase. A quasi equilibrium exist in the disk, where the amount of accreted in-falling material approximately equals the material dissipated inward. The oldest solid inclusions we can identify in primitive meteorites, CAI’s and the oldest chondrules, are formed by the end of this phase.

Models where the dissipation of material inward through the disk is based on turbulent viscosity in the disk due to thermal convection, are called *alpha disks*. The friction caused by the turbulence will make the material lose energy, which makes it possible for it to move inward, but the efficiency with which this takes place is not known, and therefore the original viscous models were parameterised with the free parameter α. For the mechanism to be efficient one would typically need $\alpha \approx 0.1$. However, more advanced theoretical considerations later indicated that for the solar nebula conditions, α would be $\approx 10^{-4}$. Therefore it became necessary to seek a more efficient dissipation mechanism, and though the final word on this is far from having been said, the most widespread opinion seems to be a combination of one or more of the following three options: (1) disk-driven bipolar outflows, (2) some type of magnetic instability, and (3) spiral density waves. We will restrict the discussion here to the disk driven bipolar outflow, and just have in mind that additional mechanisms may play a role too. Recent observations are in agreement with seeing a density wave like pattern in some nearby debris disks.

When material flows in through the disk and approaches the proto-solar surface, part of it will ionized and therefore begin being affected by the magnetic field lines. The field lines rotates as a solid body with a velocity increasing away from the Sun. They can be thought of as solid lines (”robes”) swung around by the stellar rotation. From classical mechanics (and every-day experience) we know how such robes will stretch upward in a cone around the rotational axis. Qualitatively this is the physics behind the observed cones of out-flowing material in the FU Orionis objects. At a distance of 10-20 stellar radii, the velocity is about 300 km/s ($\equiv$ the escape velocity) and models predict that the ionization fraction has decreased to about 5%. Close to the disk plane, the gas pressure dominates over the magnetic pressure, such that the material will move as in a magnetic field free gas (see Eq. 2.24 and 2.25). At high disk latitudes the magnetic pressure dominates, because the gas pressure decreases according to the hydrostatic equilibrium. Therefore the (ionized) gas moves with the field lines at high latitudes, which gives it a velocity $v \propto R$. This means that the gas flows inward in the disk plane and is pressed outward at high latitudes through the cone, hereby forming a bipolar outflow, as we see it in for example Fig. 1.5.

Below, we will discuss how such neutral gas falling in under conservation of angular momentum, and being thrown back out as ionized material along the field lines with constant angular velocity, in principle is able to transport angular momentum away from the proto-sun, and hereby slowing down the solar rotation, in agreement with the observations that already T Tauri stars are slowly rotating, such that the angular momentum loss must take place during the FU Orionis phase (or earlier).

2.3 Angular momentum evolution in the standard theory

If the angular momentum was conserved during a simple collapse of the interstellar cloud that became the solar system, almost all of the angular momentum of the original cloud would today be in the Sun, and the Sun would rotate around its axis once every 3 seconds. Obviously, the Sun rotates almost (27-days/3-seconds$\approx$) a million times too slow, and something more complicated than a simple collapse with angular momentum conservation must have taken place. It is therefore a fundamental requirement to every theory for the solar system formation to explain where all the angular momentum went that should have been in the Sun. No theory is able to fully explain this, and this is why the issue is often called *the angular momentum problem*. Before outlining some indications of what the solution of the problem could be, and what it teaches us about the formation of solar systems, we will look at some basic observations that might hold the clue to the solution.

In general, the angular momentum L can be written as

$$L = I\omega \tag{2.17}$$

	R [AU]	P [yrs]	R^2/P	j[cm^{2}s^{-1}]
Sun	0.005	0.07	—	$7.2\,10^{14}$
Mercury	0.4	0.25	0.64	$2.8\,10^{19}$
Earth	1.0	1.00	1.00	$4.4\,10^{19}$
Jupiter	5.2	11.9	2.27	$1.0\,10^{20}$
Neptune	30.0	163.7	5.50	$2.4\,10^{20}$
Pluto	39.5	248.0	6.29	$2.8\,10^{20}$
Cloud core	20000.	$2.\,10^7$	20.	$3.5\,10^{20}$

Table 2.1. The present value of specific orbital angular momentum, j, for the planets, and the specific rotational angular momentum for the Sun and for a typical molecular cloud core. R is either the distance from the Sun (for the planets) or the radius of the object (for the Sun and the cloud core), and correspondingly P is either the orbital period or the rotational period.

where the moment of inertia, I, is

$$\begin{aligned} I &= \frac{2}{5}MR^2 \text{ for a homogeneous sphere of mass } M \text{ and radius } R \qquad (2.18)\\ I &= \alpha MR^2 \text{ for a non} - \text{homogeneous sphere of mass } M \text{ and radius } R \\ I &= MR^2 \text{ for a body of mass } M \text{ moving in a circular orbit of radius } R \end{aligned}$$

Numerical models for the Sun, give $\alpha \approx 0.055$.

We will also need the concept of *specific angular momentum*, j, which is defined as the angular momentum per mass unit (here gram) of the body,

$$j = L/M \qquad (2.19)$$

Observations show that a typical molecular cloud core which we believe to be on its way to form a solar system, has a rotation period of $P \approx 2\,10^7$ yrs, $M \approx 1\,M_\odot$, and $R \approx 0.1$ pc = $3\,10^{17}$ cm. If the cloud is approximately homogeneous, this means that its angular momentum, L_{mc}, and specific angular momentum, j_{mc}, are of the order

$$L_{mc} = 10^{54}\,g\,cm^2\,s^{-1} \quad \text{and} \quad j_{mc} = 5\,10^{20}\,cm^2\,s^{-1} \qquad (2.20)$$

We can also easily compute from the above equations, the solar angular momentum, L_{sun}, and specific angular momentum, $j_\odot$,

$$L_{sun} = \alpha\,2.6\,10^{49} \approx 1.4\,10^{48}\,g\,cm^2\,s^{-1} \quad \text{and} \quad j_\odot = 7.2\,10^{14}\,cm^2\,s^{-1} \qquad (2.21)$$

and the corresponding numbers for the orbital angular momentum of Jupiter (for all the planets, only a tiny fraction of their total angular momentum is in the rotation around their axis),

$$L_{Jup} = 1.9\,10^{50}\,g\,cm^2\,s^{-1} \quad \text{and} \quad j_{Jup} = 1.0\,10^{20}\,cm^2\,s^{-1} \qquad (2.22)$$

from which we see two important facts about the solar system:

- less than 1% of the solar system angular momentum resides in the Sun,
- the planets (Jupiter) have specific angular momentum comparable to the initial specific angular momentum of the collapsing molecular cloud.

We therefore conclude that the Sun lost almost all of its initial angular momentum, but it was not transfered to the nebula from which the planets formed (as often stated in text books on the subject).

We can get some more detailed feeling about the processes that must have taken place during the collapse, by looking at Table 2.1 where the specific angular momentums of selected planets, the Sun, and the initial (i.e., a typical) collapsing molecular cloud are given. We see that all the planets have slightly lower values

	R[Mkm]	P[days]	j[cm^{2}s^{-1}]		R[Mkm]	P[days]	j[cm^{2}s^{-1}]
Jupiter	0.071	0.414	~9 10^{14}	Saturn	0.060	0.444	~6 10^{14}
Io	0.42	1.77	7.25 10^{16}	Mimas	0.19	0.94	2.66 10^{16}
Europa	0.67	3.55	9.20 10^{16}	Dione	0.38	2.74	3.78 10^{16}
Ganymede	1.07	7.16	1.16 10^{17}	Rhea	0.53	4.52	4.47 10^{16}
Callisto	1.88	16.69	1.54 10^{17}	Titan	1.22	15.95	6.81 10^{16}

Table 2.2. Same as Table 2.1, but here for Jupiter and Saturn (with α = 0.1) and their moons.

of specific angular momentum than a typical molecular cloud, decreasing with a factor of 10 inward through the solar system, and that the Sun has a specific angular momentum approximately a factor 100,000 smaller than the typical molecular cloud from which it most likely have formed. We will first see how we can explain the decreasing near-interstellar specific angular momentum of the planets, and then dig into the huge drop down to the Solar value.

We notice in passing that the giant planets and their moons are often described as "miniature solar systems". We see from Tab. 2.2 that this is partly correct. The gradual decrease of the specific angular momentum inward through the system of moons indicate that they were formed in a disk around Jupiter/Saturn the same way as the planets around the Sun, with associated angular momentum moving outward as the material moved inward through the disk to form the moons. We also see, however, that although there is a drop in specific angular momentum from the moons to the planet, it is far smaller than the drop from the planets to the Sun, indicating that the mechanism that made the Sun loose its angular momentum (to be discussed below) was absent or not very active for Jupiter and Saturn, but interestingly bringing them to end with j approximately the same as the Sun.

2.3.1 The loss of solar angular momentum.

One way the Sun could have lost angular momentum without adding angular momentum to the disk, is through the coupling of ionized material moving out of the Sun with the solar magnetic field and leaving the solar system. Ionized solar wind material (mainly protons, electrons, and ionized helium) follows the field lines, at least in the vicinity of the Sun where the field is strong. The basic reason that the material will move with the field lines, is that the field $\boldsymbol{B}$ will give rise to a force $\boldsymbol{F}$ on a particle with charge q moving with a velocity $\boldsymbol{v}$ of the magnitude

$$\boldsymbol{F} = q\boldsymbol{v}X\boldsymbol{B} \tag{2.23}$$

It is seen that the field performs no force on material moving along the field lines, while a velocity composant perpendicular to the field will be "bend" according to the equation above. This will cause the particles to spiral around the field lines (i.e., "glue to the field lines") as long as the field is strong enough.

Below we will, however, rather think of this in terms of the pressure (= energy density) the magnetic field exerts, compared to the pressure the kinetic movement of the gas exerts. Since the magnetic field rotates as a solid body, the charged material will co-rotate with the Sun while moving outward, and hence gain angular momentum which is then lost from the Sun. As long as the magnetic pressure is stronger than the kinetic pressure, the material will stick to the magnetic field. Since the (dipole) field strength decrease with the cube of the distance from the Sun while the velocity of the material is approximately constant, there will be a distance where the magnetic field strength drops to a value where the magnetic pressure is below the kinetic pressure. At around this distance, the movement of the (ionized) material will change from being determined by the magnetic forces to being determined by the laws of mechanics. The magnetic pressure (or energy density) is given by

$$P_B = \frac{B^2}{2\mu_0} \tag{2.24}$$

where B is the field and μ_0 is the permability of free space. The kinetic gas pressure (also called the ram pressure) imposed by the solar wind of particles streaming out of the Sun, is the momentum transfer per second (which is the same as the energy density of the wind or the energy difference per cm of material streaming through each cm^2). It is given by

$$P_g = \Delta mv = (\rho v)v = nmv^2 \tag{2.25}$$

where Δm is the gas flowing through each cm^2 per second, which can also be expressed as the mass density ρ times v (v being the velocity away from the Sun, ii.e. the speed of the solar wind), or the particle density n (the number of particles per cm^3) times the mass m of each particle times v^2. The thermal gas pressure nkT is small compared to nmv^2 since in this case $mv^2/k \approx m_p(5\,10^5)^2/k \approx 3\,10^7$K.

Approximating the solar magnetic field B in the close vicinity of the Sun with a magnetic dipole moment D$_\odot$, gives the size of B as function of distance r from the Sun as

$$B(r) = D_\odot/r^3 \tag{2.26}$$

The non-dipole field term is small at small distances from the Sun. Now we can find the distance r_c where the material leaves the magnetic field. This will happen around the place where the kinetic pressure and the magnetic pressure are equal, given by

$$\frac{B(r_c)^2}{2\mu_0} = P_g \Leftrightarrow \frac{D_\odot^2}{2\mu_0 r_c^6} = nmv^2$$

Assuming that the solar mass loss is uniform in all directions out of the Sun, gives

$$nm = \frac{dM/dt}{4\pi r^2 v} \tag{2.27}$$

which finally leads to

$$\begin{aligned} r_c^6 &= \frac{D_0^2}{2\mu_0 v^2}\frac{4\pi r_c^2 v}{dM/dt} \Leftrightarrow \\ r_c &= \left(\frac{2\pi D_\odot^2}{\mu_0 v(dM/dt)}\right)^{1/4} \end{aligned} \tag{2.28}$$

With the present value of the mass loss dM/dt = 2 10^{9}kg/s, the dipole field D$_\odot$ = 8 10^{22} Tm^3, and the solar wind speed v = 5 10^5 m/s, we get r_c = 3.4R$_\odot$. This means that the angular velocity of the solar surface material continues with constant value of ω until r = 3.4R$_\odot$, after which point it continues by following the normal laws of mechanics, including conserving its angular momentum. Hence, the final gain in specific angular momentum of the material which flows out of the Sun along the field lines is

$$j_c/j_\odot = (r_c/R_\odot)^2 = 10 \tag{2.29}$$

If 1/10 of the solar mass was lost (disregarding any changes in the Sun due to this), all of the angular momentum would therefore have been lost. This indicates that magnetic braking can be a sufficiently powerful mechanism to take away significant amounts of angular momentum, and we can envision that significant amount of the initial nebula momentum has been lost if a sufficiently strong wind was active during early phases of the solar evolution. However, with the present rate of mass loss (dM/dt=2 10^{9}kg/s) the Sun would have lost only about 0.01 % of its mass until now, and hence about 0.1% of its original angular momentum. Therefore the initial mass loss rate and/or the magnetic field must have been much stronger in the past if magnetic braking should be able to bring the specific angular momentum down from $j_{initial}$ to the present value. There are observational evidences that the Sun passed through a phase of strong mass loss during its pre-main sequence phase, and there are theoretical arguments that the magnetic field at that time can have been about 1000 times stronger than today. However, it seems that this phase can only have lasted one or a few million years. Table 2.3 gives the fraction of the initial angular momentum remaining after 1Myr for different combinations of mass loss rate and magnetic field strength in a more detailed calculation than the one above.

$\dot{M}/\dot{M}_{today}$	dM/dt [g/s]	D[Tm^3]		
		8×10^{23}	8×10^{24}	8×10^{25}
10^4	2×10^{16}	0.9933	0.9350	0.5106
10^5	2×10^{17}	-	0.8083	0.1190
10^6	2×10^{18}	-	-	0.0011
D/D_{today}		10	100	1000

Table 2.3. The fraction of the initial angular momentum remaining after 10^6 years, with different combinations of magnetic dipole moment and rates of mass loss.

It is seen in Tab. 2.3 that combinations of a mass loss rate approximately a million times stronger than the present, a magnetic field approximately a factor 1000 times stronger than the present, and a wind lasting for a few million years, marginally could bring the solar angular momentum down from its initial value to its present by magnetic braking. These numbers are not largely inconsistent with the values observed for T Tauri stars – the phase the Sun is believed to have passed through during the early part of its pre-main sequence evolution.

We have seen above that if the theory of the solar system having formed from a collapsing gas cloud should be trustworthy, we must be able to find a mechanism that can dissipate angular momentum from the Sun. We have also seen that qualitatively magnetic braking in some form could be the angular momentum consuming agent. However, our rough calculations seems not to be quite in quantitative agreement with the amount of angular momentum loss that is necessary (table 2.2 required a loss of almost a factor $5\,10^5$ in j, while table 2.3 only gives a factor 10^3 in the best case), so some physics is still hidden from our understanding, and calls for further refinement of the theory. One possibility is that additional angular momentum is lost during other phases of the solar formation; for example while the hydrogen burning sets in, while the deuterium burning starts, or through magnetic torque while in-falling material is expelled through a short and strong bipolar outflow event in the early T Tauri phase.

The angular momentum as described above puts quite strong constraints on our model of the solar system formation. For example, Tab. 2.3 is not in agreement with postulating that the magnetic field (just) loops the material around the inner solar system, thereby heating it, as could otherwise be an attempting explanation for the chondrule formation, that we will discuss in the next chapter. The angular momentum has to be lost out of the system. On the other hand, the mechanism outlined above is in good agreement with envisioning that our solar system had a strong bipolar outflow (Herbig-Haro-like objects), just as we saw it around young stellar objects in for example the Taurus-Auriga association or in M20 in the previous chapter.

2.4 From molecules to dust

The terrestrial planets and the solid cores of the giant planets will have formed from solid dust particles that were present in the proto-planetary disk. To build a numerical model of the formation of the solar system, we therefore need to not only compute the collapse of a gas cloud under the influence of the gravitational and centrifugal forces, but also to be able to write some equations which quantify the amount of dust we expect to form under given physical conditions. The simplest, and probably to a large extend correct, way we can do this is to assume that time-scales are always long enough, and that the relevant chemical pathways exist, such that the atoms, molecules, and solid particles eventually will reach a chemical equilibrium. Under this assumption we can compute the relative abundance of all species in the nebula.

If we in addition know the optical properties of the most important of these species, we can thereafter compute the radiative transfer (i.e., the interaction between the radiation and the matter), which in turn will give us the gas pressure and temperature at a given place in the nebula, which was what we needed in order to compute the chemical equilibrium. By iteratively solving for the chemical equilibrium, the radiative transfer, and the physical state equations, we will therefore finally end up with a self-consistent model that tells us what the temperature, gas pressure, dust- and gas-composition, etc, is as function of position in the nebula.

We can then let this equilibrium model evolve with time by letting proto-planets be build up from collisions between the dust-grains that form in the model. We will then have a theoretical model of the solar system, and we will see below that such a model can explain the gross features of our own solar system (but it may be in severe conflict with some of the exo-planetary systems discovered in recent years). A more advanced model will include effects of dynamics (in- and out-flow of material), effects of the magnetic field, turbulent motions, and many other important features seen in real YSO (young stellar objects such as T Tauri stars). Fig. 2.6 shows a schematic model of a proto-planetary disk, and the results of a numerical simulation with in-fall in the model.

To illustrate how the chemical structure of the gas and the condensation through the disk can be calculated, we will here consider a gas consisting mainly of hydrogen, and at such a temperature that most of the hydrogen is in the form of H_2 molecules. This is a reasonable first approximation to the solar nebula. Since all other elements than molecular hydrogen in this approximation are considered trace elements, we can express the total number density, $N_{\rm atoms}$, of all atoms in the gas as

$$\begin{aligned} N_{atoms} = 2N_{H_2} + N_H + N_{He} + 3N_{H_2O} + 2N_{CO} + 5N_{CH_4} + ... \\ \approx 2N_{H_2} \approx N_{total,H} \end{aligned} \tag{2.30}$$

Remark that we here use N_H to symbolize the number density of neutral atomic hydrogen in the gas, and $N_{total,H}$ to symbolize the total number density of hydrogen atoms in the gas, independent of how they are chemically bound. While N_H therefore is the real number density of neutral atoms (for example per cm^3), then $N_{total,H}$ is a fictitious number, describing how many neutral hydrogen atoms we would have had (per cm^3) if all hydrogen atoms bound in molecules, solid grains, ions, etc had been in the form of free neutral atoms.

The total number $N_{total,i}$ of atoms i relative to the total number of hydrogen atoms, $N_{total,H}$, is known from solar (and stellar) abundance analysis (see Fig. 1.1), and is often expressed as

$$\alpha_i = N_{total,i}/N_{total,H} \tag{2.31}$$

We therefore have

$$N_{total,i} = \alpha_i N_{total,H} \approx \alpha_i 2N_{H_2} \approx 2\alpha_i \frac{P_g}{kT} \tag{2.32}$$

where we have used the assumption $P_g \approx P_{H_2}$ and the ideal gas equation

$$P_g = NkT \tag{2.33}$$

where N is the total number density of particles in the gas, k is Boltzman's constant, T is the gas temperature, and P_g is the gas pressure.

Similar to the fictitious number density defined above, we will define the fictitious pressures as the sum of the partial pressures all the atoms of a given kind (e.g. hydrogen) would have if all the atoms of the given kind were free and neutral.

$$P_{total,i} = N_{total,i}kT \tag{2.34}$$

If we now let the number of atoms i in molecule j be γ_{ij} (e.g. $i \sim$H, $j \sim H_2O \Rightarrow \gamma_{ij}$=2), we then can express the fictitious pressure of atom i as

$$P_{total,i} = \sum_j \gamma_{ij} P_j \tag{2.35}$$

where P_j is the partial pressure of molecule j. This can, however, also be expressed as

$$P_{total,i} = \alpha_i N_{total,H} kT \approx 2\alpha_i P_g \tag{2.36}$$

Combining the two last equations, give

$$2P_g \alpha_i = \sum_j \gamma_{ij} P_j \tag{2.37}$$

The complete set of equations of the form of Eq. 2.37 for all atoms i, is called the set of mass balance equations. It gives, however, only one equation per element. In general each element is part of several types of particles (e.g., the neutral atom, the one time ionized atom, several molecules, etc) and we therefore need also some other equation describing the relation between the different particles each element can be a part of (atoms, ions, molecules). This balance can be a very complicated function of dynamical processes in the gas (e.g., shocks, density waves, in-falling gas, radiative destruction etc.). In the simplest form of the system, all dynamical time scales are so long compared to the reaction time scales, that the gas has reached a static equilibrium. In that case we will talk about chemical equilibrium, and the needed additional set of equations is provided by the so-called Russell equations, or equations of chemical equilibrium.

$$\frac{P_A^m P_B^n}{P_{A_m B_n}} = K_{A_m B_n} \tag{2.38}$$

which relates the partial pressures P_A and P_B of the neutral atoms A and B to the partial pressure $P_{A_m B_n}$ of the corresponding molecule $A_m B_n$ in the chemical equilibrium

$$mA + nB \rightleftharpoons A_m B_n \tag{2.39}$$

$K_{A_m B_n}$ is called the chemical equilibrium constant of molecule $A_m B_n$, although it literally is not a constant but a function of both pressure and temperature, with approximately $K \propto 1/T$. The relation between the neutral and ionized forms of the atoms (and molecules) can be expressed in the same way or from the well known and more simple Saha equation.

The equilibrium constant can be calculated from the change in Gibbs free energy, ΔG_j^0, between the product (the right hand side of Eq. 2.39) and the reactants (the left hand side of the reaction) in the formation of molecule j,

$$K_j = exp(-\Delta G_j^0/RT) \tag{2.40}$$

where R is the molar gas constant (R = 1.987 cal K^{-1} mole^{-1}). The values of ΔG_j^0 can be computed from the molecular partition functions or they can be measured in the laboratory. A fundamental source of ΔG_j^0 is the extensive JANAF tables, which lists laboratory values of ΔG_j^0 for a huge number of molecules for temperatures up to a few thousand Kelvin.

Consider the following simple example to see how the mass conservation equation can be solved together with the chemical equilibrium equations to determine the partial pressures of molecules and atoms in a gas: Let this gas consist of only H, C, and O, which we assume can only be in neutral atomic form or as the molecules H_2, CO, and CH_4. We will further assume that α_O and α_C are so small (and the temperature such that $P_H \ll P_{H_2}$) that $P_g \approx P_{H_2}$. Then the system of mass conservation equations (crf Eq. 2.37) can be written

$$\begin{aligned} 2\alpha_O P_g &= P_{CO} + P_O \\ 2\alpha_C P_g &= P_{CO} + P_{CH_4} + P_C \\ 2P_g &= 2P_{H_2} + 4P_{CH_4} + P_H \end{aligned} \tag{2.41}$$

and the system of chemical equilibrium equations can be written

$$\begin{aligned} \frac{P_C P_H^4}{P_{CH_4}} &= K_{CH_4} = exp(-\Delta G_{CH_4}^0/RT) \\ \frac{P_C P_O}{P_{CO}} &= K_{CO} = exp(-\Delta G_{CO}^0/RT) \\ \frac{P_H^2}{P_{H_2}} &= K_{H_2} = exp(-\Delta G_{H_2}^0/RT) \end{aligned} \tag{2.42}$$

For a given α_O, α_C, P_g, T we therefore have 6 equations to determine the value of the 6 variables P_H, P_C, P_O, P_{H_2}, P_{CO}, and P$_{CH_4}$.

If we allowed an additional molecule to form in the example above, we would automatically have an additional chemical equilibrium equation to include in the system, and we therefore see that we would always have a system of n independent equations to solve for n independent variables.

Inclusion of dust particles in the above equilibrium description proceeds in the same way, with the minor difference that the formation of a dust grain will take place only when the partial pressures of the constituents are above the so-called saturated vapour pressure. For example, for the mineral corundum, Al_2O_3, we can write a reaction equation similar to the gas reaction equation 2.39, as

$$Al_2O_3(s) \rightleftharpoons 2Al(g) + 3O(g) \tag{2.43}$$

where (s) here refers to a solid state and (g) to a gaseous state. In analogue to the gas phase chemical equilibrium equations 2.38, we write also for the case involving solids

$$\frac{P_{Al}^2 P_O^3}{A(Al_2O_3)} = K_{Al_2O_3} \tag{2.44}$$

where now $A(Al_2O_3)$ is the so-called activity of the solid (corundum), which is unity for pure crystalline phase. Whenever a quantity on the left exceeds the equilibrium constant on the right, the mineral will condense. In the numerical computation, the condensed amount of Al and O that is condensed to solid Al_2O_3 then have to be subtracted from the gas-phase, and a new gas-phase equilibrium has to be calculated.

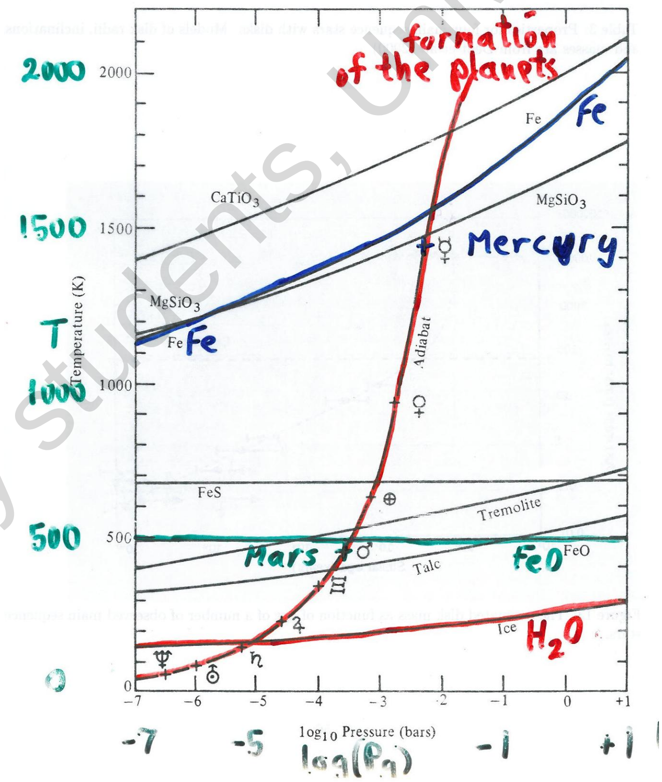

Figure 2.2. The condensation sequence and adiabat through the proto-solar nebula for the conditions where the planets could have formed. Minerals are named along the condensation curves. The position of the formation of the planets are marked by their astronomical symbols (Mercury, Venus, Earth$\sim\oplus$, Mars, the asteroids, Jupiter, Saturn, Uranus, and Neptune).

The results of such a computation is shown in Fig. 2.2 as function of temperature and pressure for a few representative condensates in the proto-solar nebula. In the figure is also shown an adiabat, which is believed to represent the temperature versus pressure conditions in the nebula at the time the dust condensed that later became the planets. The position of the present 8 planets on the adiabat are shown. We will return in the appendixes to why the temperature-pressure relation of the nebula must have been approximately an adiabat, but here just note that several qualitative characteristics of our solar system can be inferred from the figure (making the model trustworthy). At the position of Mercury, basically only Fe could condense out of the nebula. Hence we will expect Mercury to consist mainly of iron, in full agreement with observations (some theories assume a further increase in the relative iron abundance due to a late collision which stripped off large amounts of Mercury's mantel). We also see that iron condenses in the form of FeO at the distance of Mars (and outward) only. Finally, we see the *very important result* that H_2O ice can only condense from the nebula at approximately the distance of Jupiter and outward. For this reason water will be a (or even *the*) main material of which the solid cores of the outer planets and the bulk of their moons are made, and the terrestrial planets will have been created "dry" – i.e., free of water. In this way we understand why only the outer planets could grow large enough to create a nebula collapse, despite the fact that the gas density was much larger in the inner nebula region where the terrestrial planets formed. Water is one of the most common molecules in the universe. In the outer nebula regions where the conditions were right for water condensation, the amount of solid material which could form a planet core was much larger than in the inner region where water could not form solid particles ("you cannot form a snowball of water vapour, but only from ice grains and snowflakes").

It is a fundamental paradox in astrobiology that water can only condense in the outer region of proto-planetary nebulae, while life can only thrive in the inner part (as we will discuss in detail in a later chapter). Yet life is based on water. The reason for this paradox is based in the equations we derived above and those results were plotted in Fig. 2.2.

2.5 From dust to planetesimals.

In the appendixes I have described in more detail how the gravity of the Sun will drag the dust toward the mid-plane of the gas-nebula, how we can calculate that it will reach the plane within a fraction of the nebula life-time, and that the dust-grains will grow to chondrule size (i.e., mm to cm size) during their motion toward the plane; completely like a growing snowflake falling to the ground through the Earth's atmosphere. In my own oppinion, the fascinating thing about this is that it is possible from quite simple analytical calculations to predict that the first solid material will be chondrule sized clumps in the nebula midplane, formed within a fraction of the nebula lifetime (i.e. of the observed TTauri time scales). Why am I fascinated about that? Because we know that the material the Earth was formed from was exactly small solid clumps of chondrule sizes. How do we know that? Because the oldest and most primitive (in the meaning "unprocessed") material from the solar system formation, the carbonaceous chondrite meteorites, consist of 70% chondrules (plus 30% fine-grained dust). So we were able to calculate in a few pages the basic principles of why the solid planets formed the way they did: it just simply "snowed like Christmas evening" (but with "snowflakes" of solid rock and metals). A "small" detail that we cannot explain (and that differs substantially from the Christmas snow) is why these "snowflakes" were chock heated to more than 1000 degrees and quickly cooled off again, such that the grains in the end were not really snowflakes, but rather mm sized small spherical droplets that had melted and re-crystalized (i.e. became "glas-pearls") before they were incoorporated to become the solid planets. The bulk of the early Earth must have been an assembly of small glass pearls (i.e. chock heated and re-cooled chondrules). How exactly the small "snowflakes" became spherical "pearls", no-one knows (but there are hundreds of theories about it). To a first approximation one could be tempted to think of the rest of the planet building was as simple as snowflakes colliding to form "snowballs" – growing directly further to planet-sized bodies. This has, however, turned out to be non-trivial, because the snow would be swept away by non-gravitational forces before it could become even boulders, unless something more complex would take place.

Planets and moons in our solar system are almost solely affected by gravity, but smaller bodies are affected also by other forces; fx small dust grains are affected by radiative pressure and by the so-called Pointing-Robertson effect, described in more detail in the appendixes, and somewhat larger bodies, such as asteroids and meteorides, are affected by the so-called Yarkovsky effect (see Fig. 2.3). These effects will also have played a role during the formation of the planetesimals and protoplanets, but as long as the gas nebula was still present, friction and other drag forces between the gas and the solid particles (or even planets) will have been a dominating non-gravitational force.

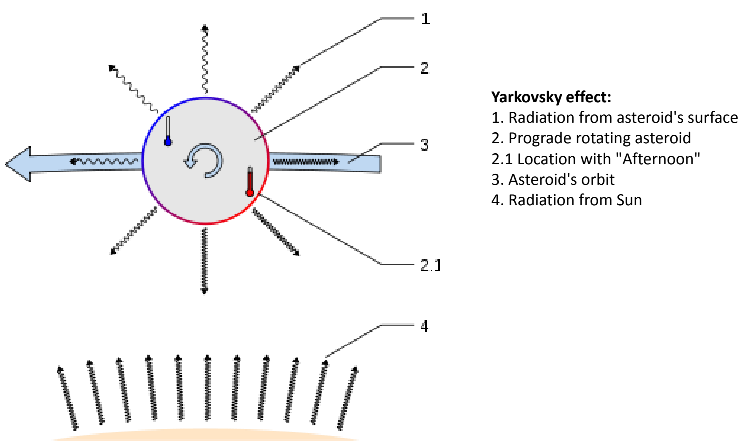

Figure 2.3. Rotating solid bodies will be heated on the dayside and cooling at the nightside. The higher temperature at "late afternoon" (red thermometer in the drawing) than just before dawn (blue thermometer) will result in a constant extra momentum on the afternoon side, which for prograde rotation will result in a push forward in the orbit and for retrograde rotation will result in a breaking in the orbital motion. This is called the Yarkovsky effect.

In principle every single gas molecule in the nebula can be thought of as being in Keplerian orbit around the protostar. However, constant collisions between the gas molecules makes them feel a pressure gradient (a force) directed toward lower pressures (i.e. outward),

$$F_{gas} = \frac{-GM\rho}{r^2} + \frac{\partial P_{gas}}{\partial r} \tag{2.45}$$

and therefore "the effective gravitational force", F_{gas}, felt by the gas molecules is smaller than the gravitational force felt by the larger solid particles. It is this reduced gravitational force that is balanced by the centrifugal force in the Kepler law for the gas ($F_{crintifugal} = \omega^2 r$), and therefore result in a reduced value of $\omega(r)$ for the gas compared to the $\omega(r)$ in Kepler's law for the solid particles. In other words, the solid particles will move faster around the Sun than the gas.

Very small dust grains will be pushed back and forth by collisions with gas-molecules, and therfore act the same way as the gas-nebula. We are used to see this with our own eyes, for example in a dust or smoke full room, where sunlight falling on the tiny particles reveal their random motion. They are pushed randomly in all directions by collisions with the air molecules, in so-called Brownian motion. Small enough dust therefore have no problem just following the gas in the nebula. But when the dust grains become cm-sized they are no longer easily draged around with the gas, and we start seeing the effects of the dust and gas being two different kind of objects. We can think of the outward push of the gas as a negative gravity (just like the air pushing a hot-air ballon upward from the Earth's surface). The gas (and tiny particles with it) therefore effectively feels a slightly smaller effective gravity from the Sun than the larger particles, and therfore it will move slower around the Sun than larger particles – chondrules or pebbles. This will be felt by the pebbles as a head-wind that breaks their motion. Eventually the larger particles will therefore loose momentum and fall toward the Sun if nothing stopped them in time. It is the simple friction against the "air" (i.e. gas), just as we feel it if we are on a bike with the wind against us.

When calculating the timescales for this inward movement of individual pebbles and larger grains, it turns out that they would end up in the Sun before they could grow to meter and km sizes and decouple completely from the gas. There are several different reasons for this inward move ("fall"), of which the head-wind is one of the most important, and there is for the moment no standard solution to how the dust eventually manage to grow from cm size to km size instead of being trapped into the protostar, but a promising theory for a solution involves what one could call "the migratory bird solution" or in more scientific terms is called "the streaming instability theory". Geese and other migratory birds that fly very far have developed the wellknown formation flying, just as we know it from groups of racing cyclists too. Both the cyclists and the geese form a group pattern that dimish the head-wind for each individual.

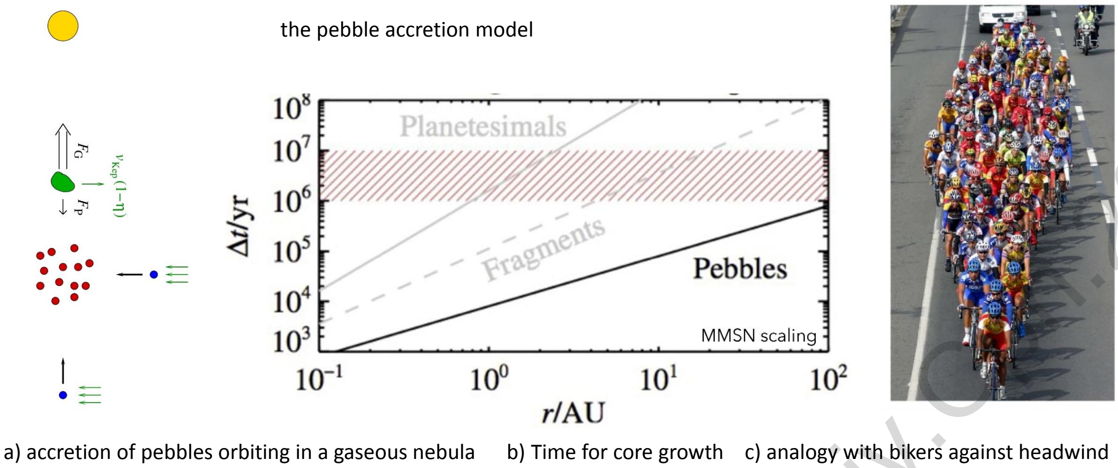

Figure 2.4. The streaming instability model predicts that pebbles (”chondrules”) in the protoplanetary nebulas will clump together due to the different velocity of orbiting gas and pebbles (panel a) in analogue with race bikers (panel c), hereby giving rise to a faster accretion toward planetesimals (i.e. objects of typically ca 100 km size) than other models, as shown for an example for the growth time to a 10 Earth-mass core in panel b.

Groups of cm-sized pebbles in orbit around the Sun, while the nebula gas still existed, will likewise have feelt a smaller head-wind than individual pebbles will. The groups of pebbles will therefore spiral slower toward the Sun than individual pebbles, and hence individual pebbles will eventually meet clumps of pebles during their inward drift. Thereby the clumps will grow larger and hence feel an even smaller head-wind from the gas, etc. In this way ”the migrating bird solution” will automatically result in more and more clumping of the pebbles, and numerical models show that the pebbles hereby can overcome the race in time, managing to grow fast enough to planetesimal size before they would otherwise had been draged into the protostar, as shown in panel b of Fig. 2.4.

Once the pebbles have grouped together like migrating birds, the low relative velocity between the grains will allow them to more easily stick together without destroying one another. It is tempting to note (again) here that the most ”primitive” (i.e. unprocessed) meteorites are exactly just aglomerations of small pebbles (chondrules) sticking together in a loose dusty clump, in close agreement with the streaming instability theory of pebble agglomeration. It is therefore reasonable to think that the first planetesimals were build up rather quickly by the streaming instability mechanism.

Let now as a next step in the building of planets, the density of free dust grains in the disk be ρ_0 [g/cm^3] and let a planetesimal be moving with a velocity v_{rel}[cm/sec] relative to the dust disk. Then the amount of dust, n_d, the planetesimal will pass through per second per cm^2 perpendicular to the direction of movement is

$$n_d = \rho_0 v_{rel} [\frac{\text{g}}{\text{cm}^2\text{s}}] \qquad (2.46)$$

If the radius of the planetesimal is r (= r(t)), the geometrical accreting area of the planetesimal is $A_{\text{geom}} = \pi\, r^2$. However the effective area, $A_{\text{eff}} = \pi\, l^2$, with which the dust feels the planetesimal moving through the disk, is larger because dust particles a bit outside the geometrical area is gravitationally attracted to the planetesimal too. The British astronomer Arthur Eddington (1882-1944) calculated A_{eff} by considering the conservation of energy and angular momentum of the small accreted particle relative to the accreting body. The result obtained in this way is widely used, and often called *the Eddington accretion mechanism*. In particular the Soviet astronomer Viktor Safronov (1917-1999) computed in the 1960's and 1970's sizes and time scales of Eddington accretion of solar system bodies, in what is now known as *the Safronov theory for planet formation*. Most work on planet formation today is an elaboration or modification of the work that was initiated by Safronov and Eddington, and we will therefore look a bit more in detail into the assumptions and the basic methods they introduced.

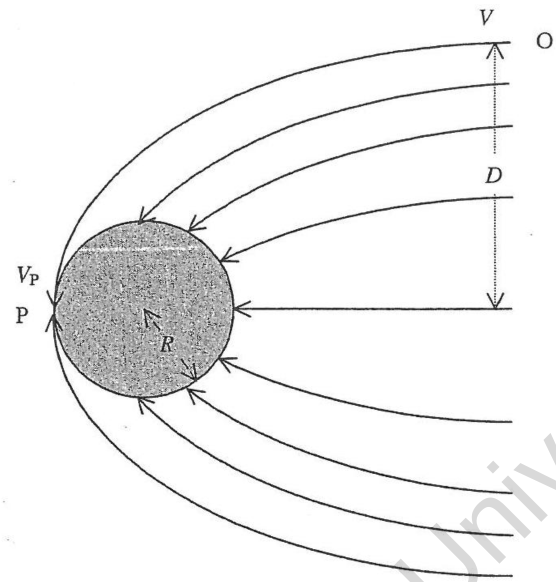

Figure 2.5. The geometry for a planetesimal accreting material due to its motion with relative velocity v_{rel} (toward the right in the figure) through a dust disk.

In Fig. 2.5 the situation is sketched for an accreting body moving through a medium of smaller particles with a relative velocity v_{rel}. Particles on the edge of the accreting area $A_{\rm eff}$ will hit the planetesimal on the backside of the planetesimal, with a velocity v_{tg} tangential to the planetesimal surface. Before the gravitational interaction with the planetesimal is felt by the dust grain, a mass unit of this dust will have an angular momentum $v_{rel}\, l$ relative to the planetesimal. After the (absorbing, inelastic) collision, it will have an angular momentum $v_{tg}\, r$. Conservation of angular momentum therefore gives

$$v_{tg}\, r = v_{rel}\, l \quad \Leftrightarrow \quad v_{tg} = \frac{l}{r} v_{rel} \tag{2.47}$$

Correspondingly, the conservation of energy gives

$$\frac{1}{2}v_{tg}^2 - \frac{Gm}{r} = \frac{1}{2}v_{rel}^2 + E_p(\infty) \tag{2.48}$$

which, with $E_p(\infty)$=0 and v_{tg} from Eq. 2.47, gives

$$l^2 = r^2(1 + \frac{2Gm}{rv_{rel}^2}) \tag{2.49}$$

or

$$A_{\rm eff} = A_{\rm geom}(1 + 2\theta) \tag{2.50}$$

with

$$\theta = \frac{Gm}{rv_{rel}^2} \; = \; E_{pot}/2E_{kin} \; = \; \frac{1}{2}\left(\frac{v_{esc}}{v_{rel}}\right)^2 \tag{2.51}$$

Qualitatively this equation says that the effective accreting area increases relative to the geometrical area with increasing density (m/r or v_{esc}) of the accreting body, and decreases with increasing relative velocity of the planetesimal and the dust. Since the ratio of the effective and geometrical area is the relative area the gravitational field is able to "focus" into the "target", the factor 1+2θ is sometimes called the gravitational focusing factor or the gravitational enhancement factor, while θ is known as the Safronov number.

The mass accreted by the planetesimal per time unit can now be expressed as

$$\frac{dm}{dt}[g/s] = \rho_0 v_{rel} \left[\frac{g}{cm^2 s}\right] \pi l^2 [cm^2] \tag{2.52}$$

If the density of the planetesimal is ρ_p, then its mass is $m = \frac{4}{3}\pi r^3 \rho_p$, and dm/dt can therefore also be written as

$$\frac{dm}{dt}[g/s] = 4\pi r^2 [cm^2] \rho_p [g/cm^3] \frac{dr}{dt}[cm/s] \tag{2.53}$$

where we have assumed that the density of the planetesimal is approximately constant during the accretion.

Setting the two expressions for dm/dt equal, one gets

$$\rho_0 v_{rel} \pi l^2 = 4\pi r^2 \rho_p \frac{dr}{dt} \tag{2.54}$$

which can be written

$$\frac{dr}{dt} = \frac{\rho_0 v_{rel}}{\rho_p} \frac{1+2\theta}{4} \tag{2.55}$$

For a virialized system

$$\begin{aligned} 2\theta &= \frac{E_{pot}}{E_{kin}} = 2 \\ &\Rightarrow \frac{1+2\theta}{4} \approx 1 \\ &\Rightarrow \frac{dr}{dt} \approx \frac{\rho_0 v_{rel}}{\rho_p} \end{aligned} \tag{2.56}$$

Eq. 2.55 expresses the not very surprising result that the growth rate of the radius of the planetesimal is proportional to the amount $\rho_0 v_{rel}$ of material it plows through per second per cm^2 divided with the density ρ_p of the material it is to be packed into (the planetesimal).

In the beginning when the planetesimal has only accreted a minor fraction of the total available amount of accretable material in the region of its orbit (i.e., ρ_0 = constant), the accretion rate will be constant according to our calculations above.

If the planetesimal manages to accrete an appreciable fraction of the available dust (i.e., ρ_0 decreasing), then of course the accretion rate must eventually go toward zero. However, in the beginning, or as long as the density of the accretable material is not substantially reduced from its original value ρ_0, dr/dt will remain constant, and we can then "integrate" Eq. 2.55 to give

$$r[cm] = \frac{\rho_0 v_{rel}}{\rho_p} [cm/s]\, t[s] \tag{2.57}$$

Numerical models as those described in connection with Fig. 2.2 and elsewhere in this chapter, give $\rho_0 \approx 10^{-10} - 10^{-9}$g/cm^3 at $r \approx$1 AU. If we set v_{rel} to a few hundred m/s (from numerical models $v_{rel} \approx 5 \times 10^{-3}v_{orb}$) and the density ρ_p of the solid planetesimals to a few g/cm^3, then we see from Eq. 2.57 that

$$r[cm] \approx 10^{-5} t[s] \Leftrightarrow t[years] \approx r[m] \tag{2.58}$$

i.e., as long as there are free dust available in the nebula for the pebble-agglomeration (the planetesimal) to accrete, then a planetesimal in Earth's distance from a solar-type star can accumulate of the order of a meter of solid material per year in Safronov-type accretion. Even in the outer parts of the nebula, where the gas and free dust densities are lower, pebble accretion followed by accumulation of surrounding free dust seems to be fast enough to account for the rapid formation of the $\sim$10M$_\oplus$ solid cores that are necessary for giant planet cores to manage to accrete the surrounding nebula gas quickly enough to become gas giants before the feeding nebula is gone. Hence, at the time of writing, the pebble accretion theory seems to be a promising solution of a long standing problem of understanding how gasplanets can have formed rapid enough (compared to the disk lifetimes) in our solar system as well as in exoplanetary systems.

2.5.1 When is the dust in the plane used up?

Obviously, the Eddington accretion approximation cannot be valid for large distances from an accreting object orbiting the Sun, because there must be a distance where the (proto)solar gravity will significantly disturb the accreted particles (the dust) compared to the simple picture in Fig. 2.5. It can be shown that the accretion will stop when there is no more material inside the so-called Hill sphere with radius R_H around the accreting mass (e.g. a planetesimal with mass M_{ps}), given by

$$R_H = a_\odot \left(\frac{M_{ps}}{3M_\odot}\right)^{1/3} \tag{2.59}$$

where $a_\odot$ is the distance from the planetesimal to the Sun. When the planetesimal moves around the proto Sun it will plow through a total mass of dust in the disk given by

$$\begin{aligned} M_{RH} &= 2\pi a_\odot \Sigma_p 2R_H \\ &= \frac{4\pi a_\odot^2 \Sigma_p \sqrt[3]{M_{ps}}}{\sqrt[3]{3M_\odot}} \\ &\propto r_{ps} \end{aligned} \tag{2.60}$$

It is seen that the mass which the growing planetesimal is able to accrete from, increases linearly with the radius of the planetesimal.

$$dM_{RH} \propto dr_{ps} \tag{2.61}$$

On the other hand the mass of the planetesimal when it accretes mass from the disk increase as

$$dM_{ps} = d(\frac{4}{3}\pi\rho_{ps}r_{ps}^3) = \frac{4}{3}\pi\rho_{ps}3r_{ps}^2dr_{ps} \propto r_{ps}^2dr_{ps} \tag{2.62}$$

Since the already accreted material increases proportional to $r_{ps}^2dr_{ps}$ and the available material inside the Hill sphere to be accreted increases only as $\mathrm{d}r_{ps}$, the accretable material will eventually run out. The maximum available material that can be accreted is given by

$$\begin{aligned} M_{RH}^{max} &= 4\pi a_\odot \Sigma_{ps} R_H^{max} \qquad \text{(from Eq. 2.60)} \\ \text{and } R_{RH}^{max} &= a_\odot \left(M_{RH}^{max}/3M_\odot\right)^{1/3} \qquad \text{(from Eq. 2.59)} \\ \Rightarrow M_{RH}^{max} &= \left(\frac{4\pi a_\odot^2 \Sigma_{ps}}{(3M_\odot)^{1/3}}\right)^{3/2} \\ &\approx 6\,10^{25} g \approx 0.01 M_\oplus \approx M_{moon} \end{aligned} \tag{2.63}$$

where the value in the last line is derived for the region of the Earth only (and for our proto-solar nebula only), by inserting the estimates for the nebula surface density Σ_{ps} ≈10g/cm^2 derived in the appendixes.

Older theories of the planet formation predicted that the planets were formed from accreting dust for a substantial amount of time. Figuratively, one can envision this as a winter day, where the snow (here more general dust) falls and falls and falls for a million years, until in the end so much cosmic dust has fallen on the proto-Earth that it has grown to the size it has today. However, Eq. 2.63 teach us that the formation has been a much more violent process. The accumulation of the lunar sized objects can to some extend be envisioned as the snowfall, although the size and material of the dust-grains and the speed with which they have hit the accumulating planetesimals would make it more correct to compare the process with sand-blasting than with snowfall (stone-metal grains of mm to cm size, which have hit the planetesimals with velocities typically around 500 km/h, and probably without the final breaking effect the Earth's atmosphere has on present day meteorites). Eq. 2.63 teach us that once the accumulation has produced lunar-sized objects, the dust is basically gone, and the interaction is no longer via "sweep-up" collisions. The planetesimals have emptied each their $2R_H$ broad track in the dust disk, and in the end there must have been at least 100 lunar mass objects in the Earth's present region. The distance between their orbits have been approximately two

Hill sphere radii – less than 3 times the present distance between the Earth and the Moon. The gravitational interaction between the many planetesimals have pumped their orbits to higher eccentricities. In this way they have started crossing one another's orbits, and over a much longer timescale than the accumulation phase, they will now occasionally have collided with one another and slowly build up the Earth. Numerical simulations show that the Earth can have build up in approximately 100 million years from collisions of 100 lunar-sized planetesimals. These collisions between celestial bodies the size of the Moon, or larger, have of course been much more violent processes than the old silent snowfall analogy. But not only this; if estimates, that we will soon return to, of the amount of short-lived radioactive elements in the early planetesimals are correct, then the lunar sized planetesimals will already have melted from the radioactive heat. Then the final collisions that formed the Earth will have been between hot clumps of melted stones and metals the mass of the Moon. Certainly a different event than a million years of silently falling Christmas snow!

2.6 From planetesimals to planets.

This final period of the life of the accretion disk is the time of the weak-line T Tauri (WTT) objects. They show no evidence of accretion from the disk down to an accuracy of $10^{-9}M_\odot$/yr. In the HR diagram the CTT (classical T Tauri) stars are generally above the WTT stars, because the PMS (Pre Main-Sequence) evolutionary tracks go from high luminosity of the large contracting cloud to lower luminosity of the smaller contracted star. Some WTT stars are, however, up into the CTT regime, indicating that for these stars the inflow has stopped early.

While the planetesimals of km-size (to perhaps Moon-size) have to build up during the CTT phase, these clumps can collide and build the final planets over the longer time scales of the WTT phase.

When the planetesimals have accreted the material in their local envelope, the accretion rate will slow down considerably, because growth is no longer determined by direct "plowing through" material, but is determined by the long-ranged gravitational force. Gravitational interactions between them is to be considered an elastic collision in a broad understanding of this concept. Elastic collisions will increase the average random velocity of the objects (planetesimals). When planetesimals collide (and build up one larger body), it is an inelastic collision, which decrease the average random velocity of the bodies. In other words, the gravitational interaction between the planetesimals will make the orbits more elliptical (increase the random velocity, or pump up the orbits as it is sometimes expressed), while inelastic collisions will circularize the orbits (decrease the random radial velocities). When the orbits of the planetesimals in a region are pumped up, they will cross one another, and therefore occasionally collide (inelastically). Originally the planetesimals therefore formed in circular orbits because of accretion in their local envelope. After the material in the local envelope fairly quickly was used up, the orbits were pumped up, which caused collisions, which finally circularized the accreting planets again, and separated the formed planets in non-interacting circular orbits.

Once one planetesimal has reached a size where its escape velocity, v_{esc}, is larger than the average random relative velocity of the planetesimals, $< v_{rel} >$, then this planetesimal will become the planet by accreting all the other planetesimals in the region.

If the largest planetesimal have radius and mass r_l and m_l, and the other planetesimals are build of the same material (i.e., have the same density), then the mass ratio and the radius ratio of a general planetesimal to the largest one, are related by

$$\frac{m}{m_l} = \left(\frac{r}{r_l}\right)^3 \tag{2.64}$$

Consider now the situation when the first body has grown to the size where its escape velocity, $v_{esc,l}$, is a small factor larger than $< v_{rel} >$, i.e.

$$\frac{v_{esc,l}}{< v_{rel} >} = \sqrt{2\theta} > 1 \tag{2.65}$$

Then Eq. 2.52 can by use of Eq. 2.64, 2.49, and 2.51 be written as

$$\frac{\dot{m}/m}{\dot{m}_l/m_l} = \frac{r_l}{r}\frac{1+2\theta(r/r_l)^2}{1+2\theta} \tag{2.66}$$

This equation is 1 for $r = \frac{1}{2\theta}\, r_l$ but is less than 1 for $r > \frac{1}{2\theta}\, r_l$. I.e., once the next-largest planetesimal has grown to the radius $r_{2l} = \frac{1}{2\theta}\, r_l$ it is forced to follow the same relative accretion rate as the largest one, and its mass will therefore always be the same ratio

$$\frac{m_{2l}}{m_l} = \frac{ri_{2l}^3}{r_l^3} = (\frac{1}{2\theta})^3 \tag{2.67}$$

of the largest planetesimal. If r_{2l} by accident happened to grow beyond this limit, its growth rate would decrease until the mass ratio given by Eq. 2.66 was again reached. For $\sqrt{2\theta} \approx 2$, the mass ratio between the largest and second largest planetesimal will be $m_l/m_{2l} = (2\theta)^3 \approx 100$. In other words, once the first planetesimal has reached a mass that gives it an escape velocity just twice the average relative velocity of the planetesimals, no other planetesimals can ever grow beyond 1/100 the mass of this planetesimal, and it is obviously now bound to win the game to be ”the planet of the region”.

In summary, the arguments above have made it plausible that the solar system can have formed from an interstellar cloud fragment of only a bit more than one solar mass, of which $1\mathrm{M}_\odot$ accumulated to form the Sun and about $10^{-2}\mathrm{M}_\odot$ formed a disk around the Sun from which $\approx 1\%$ of the mass accreted to form the planetesimals and other solid bodies in the solar system. The planets outside the snow-line accreted enough solid material that nearby hydrogen and helium masses could finally collapse onto the solid core a form the giant planets. Except for the few percent of the original proto-solar nebula that was eventually accumulated in the planets (mainly Jupiter), the rest might have been blown back into interstellar space and/or accumulated by the Sun. After this phase, planetesimals could still collide to build up the final planets.

While these ideas form a plausible working hypothesis, making it a quantitatively acceptable or well proven theory, we still need to create a numerical model which comes with specific quantitative predictions that can be tested, and we need to make good quantitative observations of other solar systems in various phases of their formation. This huge task has only begun, but recent years revolutionary discoveries of planets around other stars and proto-planetary disks in large numbers, have put more solid ground to base our models on, far beyond the previously single known example of our own solar system. We will return to discussing how the comparison with the known exoplanetary systems has improved the development of solar system formation models, in the chapter about exoplanets.

Appendixes

2.7 Angular momentum transport in a dissipative disk.

In the beginning, the collapse will be basically a free fall. The specific angular momentum will be conserved, and the angular velocity will increase correspondingly when the radius decrease. However, the centrifugal force, F_c (per mass unit), increases with increasing angular velocity,

$$F_c = \omega^2 R = j^2/R^3 \tag{2.68}$$

and for conserved specific angular momentum, F_c increases proportional to R^{-3}. Hence there will be a distance when F_c equals the gravitational force, F_g,

$$F_c = F_g \quad \Leftrightarrow \quad j^2/R^3 = GM/R^2 \quad \Leftrightarrow \quad R = j^2/GM \tag{2.69}$$

For M = $M_\odot$ and j = 3.5 10^{20} we get R = 62 AU. This is approximately twice the radius of Neptune's present orbit, or the speculated outer edge of the Kuiper belt. In other words, when the in-falling material in the collapsing cloud reach approximately the outer edge of the present solar system, conditions for Kepler orbits (F_c = F_g) are fulfilled.

All the planets (and the Sun) are inside this limit, and they have decreasing j the further inside they are (see Tab. 2.1). This might be a natural consequence of the dissipation of energy (gravitational torques, density waves, viscosity, radiative dissipation, etc). In circular orbits, we can write the kinetic energy, E_k, per mass unit as

$$E_k = \frac{1}{2}v^2 = \frac{1}{2}\left(\frac{j}{r}\right)^2 \tag{2.70}$$

and the total energy per mass unit can therefore be written as

$$E = \frac{1}{2}\left(\frac{j}{r}\right)^2 + \Phi \tag{2.71}$$

Imagine now two objects (molecules, planetesimals, etc) moving in circular orbits of radius r_1 and r_2, and with angular velocity ω_1 and ω_2, then

$$\begin{aligned} E_{tot} &= E_1 + E_2 \\ &= \frac{1}{2}\left(\frac{j_1}{r_1}\right)^2 + \Phi_1 + \frac{1}{2}\left(\frac{j_2}{r_2}\right)^2 + \Phi_2 \Rightarrow \end{aligned} \tag{2.72}$$

$$\begin{aligned} dE &= \frac{1}{2}r_1^{-2}2j_1 d_{j_1} + \frac{1}{2}j_1^2(-2)r_1^{-3}dr_1 + d\Phi_1 + \frac{1}{2}r_2^{-2}2j_2 d_{j_2} + \frac{1}{2}j_2^2(-2)r_2^{-3}dr_2 + d\Phi_2 \\ &= \frac{j_1}{r_1^2}dj_1 - \frac{j_1^2}{r_1^3}dr_1 + d\Phi_1 + \frac{j_2}{r_2^2}dj_2 - \frac{j_2^2}{r_2^3}dr_2 + d\Phi_2 \end{aligned} \tag{2.73}$$

but from the condition F_c = F_g we have

$$\begin{aligned} \frac{v^2}{r} &= \frac{-GM}{r^2} \Leftrightarrow \\ \left(\frac{j}{r}\right)^2 \frac{1}{r} &= \frac{d\Phi}{dr} \Rightarrow \\ \frac{j^2}{r^3}dr &= d\Phi \end{aligned} \tag{2.74}$$

which introduced into Eq. 2.73 gives

$$dE = \frac{j_1}{r_1^2}dj_1 + \frac{j_2}{r_2^2}dj_2 \tag{2.75}$$

If now during the passage (of two molecules, two planetesimals, etc) the total specific angular momentum $j_1 + j_2$ is conserved, then $dj_1 = -dj_2$, and

$$\begin{aligned} dE &= dj_1 \left(\frac{j_1}{r_1^2} - \frac{j_2}{r_2^2} \right) \\ &= dj_1(\omega_1 - \omega_2) \end{aligned} \tag{2.76}$$

If the system dissipates energy, then dE < 0 which according to Eq. 2.76 then implies that dj_1 and $(\omega_1 - \omega_2)$ have opposite signs. Hence, the object with the lowest ω will gain j (and the object with the highest ω will lose j). Generally ω decreases outward in the system (nebula), and the outer particle will therefore gain j, and the inner will lose j. Since

$$j = r\,v = \sqrt{GMr} \tag{2.77}$$

increasing value of j implies increasing r, so that the object with the lowest ω will move outward as a consequence of the passage, and the object with the highest ω will move inward. We therefore see that while the (inner) material will move inward, the angular momentum will move outward, and we understand that it is possible that the material that eventually forms the inner planets will have lower values of specific angular momentum than the outer planets. We can therefore qualitatively understand the trend of decreasing specific angular momentum inward through the solar system, as a consequence of the Keplerian orbits of dust-grains and planetesimals and gravitational interactions between them. While the inner objects pass the outer objects, material will move inward and angular momentum will move outward, as a natural consequence of the dissipation of energy.

2.8 A simple model for the disk structure:

Static disk, dust accretion, and formation of planetesimals; The classical T Tauri phase (CTT).

An important fraction of the formation of a solar system, takes place during the classical T Tauri (CTT) phase. The CTT objects show evidence of accretion, so some material must still flow from a disk and onto the star, but at a much smaller rate than during the FU Orionis phase. In the beginning of the classical T Tauri phase, the inflow rate onto the star is about 10^{-7}M$_\odot$/yr, and the bipolar outflow about 10^{-8}M$_\odot$/yr. It is unknown how long time mass continues to flow in through the disk from the surrounding molecular cloud, and it is likely to depend strongly on the specific environment in which the star is formed, as discussed in the chapter about the observations. During most of the CTT lifetime, the mass inflow rate is observed to be $\approx 10^{-8}$M$_\odot$/yr. By the end of the CTT phase the observed mass inflow rate has dropped to 10^{-9}M$_\odot$/yr, which defines the boundary to the weak line T Tauri (WTT) phase, where the inflow cannot any longer be traced with present observational capabilities. A mass flow rate of $\dot{M} < 10^{-9}$ M$_\odot$/yr can potentially be sustained for several million years. The gas disk could also have been fully accumulated onto the proto-Sun during the beginning of the weak-line T Tauri phase, in which case the nucleation of dust from molecules in the nebula must take place during the few million years of the CTT, together with the growth of the dust particles, the settling onto the mid-plane, and the collisional accumulation in the plane to bodies large enough to obey Kepler's laws rather than being affected by gas-drag, radiative pressure, and other forces acting on small grains.

The Sun, stars in the solar neighbourhood, and interstellar space have approximately 99% of their mass in the form of H and He, and only 1% of the mass is in the form of what we have most of around us on Earth (C, N, O, Fe, Si, etc). We therefore must assume that there has been at least $\sim$100 times more mass (mainly in the form of H and He) in the gas the Earth condensed from, than what finally became the Earth, and that the rest somehow disappeared later on. Similarly, there must have been at least 100 times more gas than there are mass of other elements than H and He in the other planets. Jupiter is by far the largest planet in our solar system. Of the 450M$_\oplus$ in the total planetary system, 317M$_\oplus$, or 70%, is in Jupiter. Even though the gas planets are composed mainly of hydrogen and helium, there is relatively much more heavy elements in them than in the Sun. So also they have selectively collected the heavy elements. Jupiter has a solid core of

$\sim$ 10M$_\oplus$ made of heavy elements and $\sim$ 300M$_\oplus$ hydrogen and helium, so a ratio of 1:30 of heavy elements to H+He compared to the solar 1:100. Saturn has approximately the same ratio, while Uranus and Neptune has a considerably higher ratio of heavy elements to H+He, maybe 1:3. In total we can sum up that there are of the order 40M$_\oplus$ heavy elements in the planetary system. There may have been another 1 to 10M$_\oplus$ in the form of $\approx 10^{12}$ comets in the early solar system which are now in the Oort cloud (and a bit in the Kuiper belt, and some have collided and were included in the Sun and the planets), and maybe 1M$_\oplus$ asteroids which we do not know where possible are today. In total there is therefore about 50M$_\oplus$ heavy elements in the solar system (plus 0.01M$_\odot$= 3000M$_\oplus$ in the Sun). The nebula this material condensed out of must therefore have been at least 50M$_\oplus$ $\times$ 100 = 5000M$_\oplus$ $\approx$ 0.01M$_\odot$. In other words the mass of the disk could have been as small as 1% of the mass of the Sun if the condensation of the heavy elements was 100% efficient. If only one tenth of the heavy elements in the nebula condensed into planets, the nebula must have been ten times larger, or 0.1M$_\odot$, etc, but we can conclude that the absolute minimum mass of the condensing disk must have been 0.01M$_\odot$.

Once we know the total mass of the solar nebula, we can estimate the mass distribution through the disk by virtually smearing out the masses of solid (i.e., non-H+He) mass of the planets half-way to the neighbouring planets. We express this best as the so-called surface density $\Sigma(r)$ (sometimes called column density) of the nebula, defined as the total mass in a column vertically through the nebula, measured for example in g/cm^2. This exercise show us that the total mass of the nebula to a good approximation is a simple polynomial of the the distance r from the Sun,

$$\Sigma(r) = \Sigma_0 \left(\frac{r}{r_0}\right)^n \tag{2.78}$$

where Σ_0 is the surface density at a reference distance r_0 (for example 1 AU).

By comparing Eq. 2.78 with the distribution of the solid planetary mass we fitted in order to derive it, we immediately notice that the solid planetary mass in the region of Mars and the asteroids falls markedly below the formula. An early formation of Jupiter would prevent the planetesimals in the Mars-to-asteroid region in assembling themselves into a large planet (just like Jupiter is forming gaps in the asteroid belt today, preventing asteroids in crossing from one asteroid zone to the other). The small planetesimals would suffer pertubations from Jupiter over long time scales and slowly be send out of the solar system or on collision course with the planets and the Sun. Therefore the fact that the mass of Mars and the asteroids are much smaller than predicted by Eq. 2.78, brings support to our solar nebula model with early Jupiter formation. Similarly, we notice that the observed mass of the Kuiper belt objects are markedly below the distribution given by Eq. 2.78. This could mean that many Kuiper belt objects are perturbed out of the system (by Neptune), but more likely indicate an incompleteness in our surveys to search for these remote objects, and therefore are likely to give us a hint about how many we have not yet identified.

Integration of Eq. 2.78 gives us the total mass of the nebula out to a chosen boundary radius R

$$M(R) = \int_0^R \Sigma(r) 2\pi r dr = \frac{2\pi\Sigma_0 r_0^{-n}}{n+2} R^{n+2} \tag{2.79}$$

If we estimate the total nebula mass, M(R), as the sum of the solid material in the existing planets, and we have determined n from a fit to the mass distribution, then also the zero-point, $\Sigma_0(r_0)$, is determined. This could conveniently be the density at 1 AU.

In praxis the determination of n and $M(R)$ of course has considerable uncertainty, caused by the difficulty to determine the core mass of the giant planets, to account for the mass that have been perturbed away (by Jupiter), etc. Representative values from the literature are (n, $M(R)$) = ($-$1.5, 0.01), ($-$1.5, 0.02), ($-$1.0, 0.1), which would predict the density at 1 AU to be Σ_0 = 1700 g/cm^2, 3200 g/cm^2, and 4250 g/cm^2, respectively.

All of these numbers (0.01, 0.02, and 0.1M$_\odot$) are in agreement with observational estimates of the T Tauri star disk masses. In Fig. 1.11 we saw the observed dust mass of proto-planetary disks to be $\sim$ 10 to 1000 M$_\oplus$. With an estimated total mass being 100 to 1000 times the dust mass, we end up with and observatiol range of total T Tauri disk mass of 1000 to 1 million M$_\oplus$, or 0.001 to 1 M$_\odot$, spanning well the 0.01 to 0.1M$_\odot$ range argued for by the simple model above. Varying $M(R)$ in Eq. 2.79 and the corresponding value of n, together with a parametrization of the relative amount of condensable material, forms the basis of calculating scaled solar system disk nebula models. Such models could be used to predict the expected number of different populations of exoplanets in the Galaxy, and then be further improved by comparing with the actually observed population, which for the moment is very biased by observational limitations, but could be expected to be better in the future. Also full spatial resolution in the observations of protoplanetary disks is crucial for developing more reliable models. All the technological development for these advances in science is fortunately well on the way, and we can expect great progress in the field in the years to come.

During the ∼50,000 yrs of the FU Orionis state $1\mathrm{M}_\odot$ flows through the disk (or $2\,10^{-5}\mathrm{M}_\odot$/yr flowing through a $0.01\mathrm{M}_\odot$disk, to eventually almost all of it ending up becoming the $1\mathrm{M}_\odot$ Sun). According to Fig. 2.1 a $1\mathrm{M}_\odot$ star will have approximately $10\mathrm{L}_\odot$ ($\sim 3\,\mathrm{R}_\odot$ if $\mathrm{T}_{\mathrm{eff}}$ is unchanged) when it ends its FU Orionis phase and becomes a CTT object. Therefore the collapsing gas cloud will have contracted from a very large radius ($\sim \infty$) to R = $3\,\mathrm{R}_\odot$ during the FU Ori phase. The corresponding change in potential energy is $\mathrm{E}_{\mathrm{pot}}$ = $\mathrm{G}(\mathrm{M}_\odot)^2/3\mathrm{R}_\odot$, of which half according to the virial theorem will have been converted into thermal energy, such that the accumulated thermal energy will be

$$\Delta E = \frac{GM_\odot^2}{6R_\odot} = 6\,10^{47}\mathrm{erg} = 6\,10^{40}\mathrm{J} \tag{2.80}$$

The system radiates a total of 50,000 years × $10\,\mathrm{L}_\odot$ = $6\,10^{46}$erg over the 50,000 years. So, there is ≈ 10 times more potential energy released than radiated away during the FU Orionis phase. This makes it reasonable to assume that the disk is heated primarily by conversion of potential energy, and not by radiation. Furthermore, we conclude that when only 1/10 of the available energy is radiated away to the surroundings, the assumption of an adiabatic collapse seems realistic, because an adiabatic process is a process that does not exchange energy with the surroundings.

As we will see in a moment, the adiabatic assumption allows us to calculate the temperature throughout the nebula as function of distance, provided we know the temperature at one specific place in the nebula. We therefore need to introduce one more empirical zero point from our own solar system, namely a temperature at one point in the nebula, but then also this scaling is brought over into the scaled solar models. In our own solar system there is one place where we know what the nebula temperature must have been. S-type asteroids are de-hydrated, while C-type asteroids formed at a distance in the nebula where it was cold enough for water to be bound into the condensing minerals in the form of hydrates. Today this division line is at 2.7 AU, and there are good reasons to believe that it has always been there (because Jupiter keeps the two populations separated from one another by a resonance gap). Depending on the exact gas pressure in the nebula, such a division between hydrated and de-hydrated minerals condensing of the nebula gas will be slightly below or above 170 K. We will therefore put our model nebula temperature to 170 K at 2.7 AU. This division line is most often called the snow line, although some authors will prefer to place the snow line a bit further out (such that it actually ' 'snowed" with regular snow and water-ice particles outside the snow line and not inside). The important issue is that the temperature was close to 170 K at 2.7 AU in the proto-solar nebula. The position is only slightly dependent on the stellar mass, since the heating of the nebula is not due to the (very mass dependent) stellar luminosity, but due to the (less mass dependent) conversion of potential energy during the infall. In our solar system, Jupiter formed outside the "snow line", while the terrestrial planets formed inside.

We are now ready to develop the rest of our simple protoplanetary nebula model. We will first see how the available condensable material relates to the pressure and density of the nebula disk, then we will estimate the temperature profile though a disk of a certain pressure profile, and finally we will look into which minerals (and hence planet types) can condense from a nebula of a given temperature-pressure relation.

We have already shown that a mass unit ρ in the disk is affected by a force F_z pointing toward the midplane of the disk, and that this force can be written as

$$F_z = -\omega^2 z\rho \tag{2.81}$$

where $\omega = \omega(r) = \sqrt{GM_\odot/r^3}$ is the Keplerian angular velocity. In order for the gas disk to be in hydrostatic equilibrium, this "downward" pointing force must be balanced by a corresponding pressure gradient (i.e., force per area) dP/dz pointing upward; i.e.

$$\frac{dp}{dz} = -\omega^2 z\rho \tag{2.82}$$

We now introduce the definition of the sound speed

$$v_s^2 = \frac{dP}{d\rho} \tag{2.83}$$

and the ideal gas equation

$$P = \rho RT/\mu \tag{2.84}$$

where R is the (molar) gas constant and μ is the molar weight. If we now further simplify the model by assuming that the temperature is constant in a given column perpendicular to the plane (i.e., vertically isothermal), then Eq. 2.84, 2.83 and 2.82 give us

$$v_s^2 = \frac{dP}{d\rho} = \frac{\rho R}{\mu}\frac{dT}{d\rho} + \frac{RT}{\mu} = \frac{RT}{\mu} = \text{constant} \tag{2.85}$$

(and $\frac{RT}{\mu} = \frac{P}{\rho}$ according to Eq. 2.84), so that

$$\frac{dP}{dz} = \frac{dP}{d\rho}\frac{d\rho}{dz} = v_s^2\frac{d\rho}{dz} \tag{2.86}$$

Introducing Eq. 2.86 into Eq. 2.82 and integrating, give

$$v_s^2 \int_{z=0}^{z=h} \frac{1}{\rho} d\rho = \omega^2 \int_{z=0}^{z=h} z dz \tag{2.87}$$

$$v_s^2 ln(\rho(h)/\rho(0)) = \omega^2 h^2/2 \tag{2.88}$$

$$\rho(h) = \rho(0) exp(-h^2/H^2) \tag{2.89}$$

where we have introduced the scale height $H = \sqrt{2}v_s/\omega$.

We are now ready to relate the physical variables P, ρ, and T to the surface density $\Sigma(r)$ (which we get empirically for our solar system by "smearing out the planets"), by integrating $\rho(z)$ from Eq. 2.89. Combining Eq. 2.89 and 2.84 and noticing that the gas and dust is symmetrically distributed around the midplane, and that $\int_0^\infty exp(-a\,z^2)dz = \frac{1}{2}\sqrt{\frac{\pi}{a}}$, gives us

$$\Sigma(r) = 2\rho(0)\int_o^\infty exp(-h^2/H^2)dz = \frac{\mu H(r)P_0(r)\sqrt{\pi}}{RT(r)} \tag{2.90}$$

$$P_0(r) = \frac{v_s^2(r)\Sigma(r)}{H(r)\sqrt{\pi}} = \frac{v_s(r)\omega(r)\Sigma(r)}{\sqrt{2\pi}} \tag{2.91}$$

In Eq. 2.91 and 2.90 we have explicitly marked the variables that are a function of the radial coordinate r of the nebula (in order to make the dependence clearer), but they of course have the same meaning as in the equations where the r-dependence is not explicitly marked. In Eq. 2.91 we now see that the gas pressure $P_0(r)$ in the midplane now is determined once we have determined $T(r)$ (since Σ is determined empirically, ω is a simple function of r, and $v_s(r) = \sqrt{RT(r)/\mu}$).

$T(r)$ can be estimated in several ways (and it is by no way obvious which one is "the correct one"), giving rise to a variety of models. A simple argument could be to notice that the nebula gas is not too far from being in an adiabatic state. This is most clearly seen by noticing that the luminosity seen from solar type stars during their disk phase, corresponds to a radiation of less than 10% of the potential energy released from the infall of the material. In other words, most of the energy available goes to heating up the disk and the protostar, and only a minor fraction of the energy is radiated away into the surroundings (the definition of an adiabatic process is that there is no energy exchange with the surroundings). Under these conditions the ratio of the temperature at two places in the gas, $T(r)/T(r_0)$, can be expressed as function of the pressure (or density) as

$$\frac{T(r)}{T(r_0)} = \left(\frac{P(r)}{P(r_0)}\right)^{(\gamma-1)/\gamma} = \left(\frac{\rho(r)}{\rho(r_0)}\right)^{(\gamma-1)} \tag{2.92}$$

where $\gamma = C_p/C_v$ is 7/5 for a diatomic gas and 5/3 for a monoatomic gas. Eq. 2.92 can by use of the equation 2.91 now be expressed as

$$\frac{T(r)}{T(r_0)} = \left(\frac{r}{r_0}\right)^{2(n-3)(\gamma-1)/(\gamma+1)} \tag{2.93}$$

showing that the slope (but not the absolute value) of the temperature through the disk can be estimated from these approximations. The most important relation for the determination of the composition of the planets, is to express T directly as function of P (Eq. 2.92), since this relation tells us which minerals can condense out where in the nebula (as shown in Fig. 2.2 from an adiabatic $T(P)$-relation). The great force of such relatively simple calculations, is that they actually tells us what the planets will be made from, and how big they can become. The fact that these results are in good qualitative and quantitative agreement with the observed (i.e., our own) solar system, is one of the main reasons that many people believe that this description by-and-large is a correct description about how solar systems form, in spite of its intimate empirical connection to our own solar system and in spite of some major problems with the theory to be discussed soon.

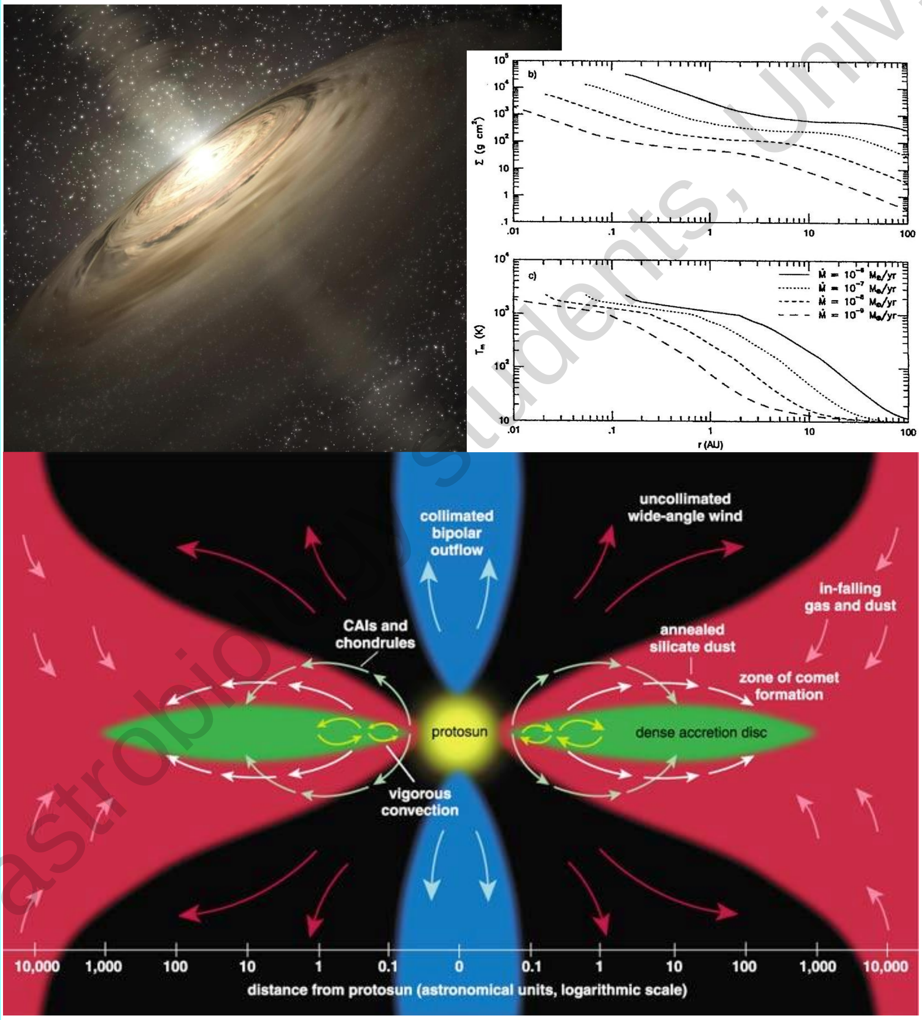

Figure 2.6. Artist impresion (upper left) and scematic illustration (lower panel) of the infalling and outflowing material and various regions in the diskforming protostellar epoche. Upper right panel gives four examples of model computations of the surface density and temperature structure of the disk.

2.9 Growth and settling of dust to the plane of the disk.

The condensation of dust in chemical equilibrium, as described above, is a well established process, which is valid to great accuracy provided chemical equilibrium prevails. Chemical equilibrium, however, requires that the reaction time scales are short compared to the destruction time scales. Naturally, this assumption may be invalidated either if the reaction time scales are too long (e.g., in the very dilute gas in the outer parts of the nebula) or if the destruction time scales are too short, i.e. if turbulent convection, shock waves, UV radiation, or other violent phenomena are dominant. In several regions (if not all) of the solar nebula, full chemical equilibrium was probably invalid, and molecule formation and dust condensation will have been a dynamical phenomenon described by rate coefficients and grain-gas-ion chemistry. The atoms may have been attached and evaporated many times before they were finally included into a mineral that was eventually incorporated into the formation of a planet. However, our understanding of the formation of solar systems is still in such an exploratory state that inclusion of equilibrium chemistry probably isn't the most severe approximation in the models.

When dust is formed, whether as a chemical equilibrium process or as a dynamical process, it will fall through the gas and settle toward the mid-plane of the disk – exactly as snowflakes fall through the Earth's atmosphere and settle at its surface. In fact the majority of the ”dust grains” in the solar nebula will precisely have been snowflakes from water ice, because H_2O was the most abundant condenseable molecule in the nebula (although only at solar distances beyond $\sim$ 5 AU), and the picture of a quiet winter evening where the snow is falling heavily, is far from just an analogy. Following the usual astrophysical terminology we will call any small solid particles for dust grains, whether being made of ice, metal, or rock, and we will approximate them with spheres, although a lot of the dust will probably have been fractal like a snowflake. In the beginning the dust-grains will be so small that we in daily life would have called them smoke, and since they travel quite a long distance and time through the gas before they settle at the mid-plane, we expect they have grown enough that we would probably rather had called them gravel (or hails) had they been on the Earth. In astronomy we will for simplicity call all of it dust.

The purpose of this paragraph is to estimate to approximately which size our scenario predicts the dust particles to grow, and how quickly we should expect them to settle onto the nebula mid-plane. These are quite important, but yet unsolved, questions in the understanding of the solar system formation, and substantial uncertainty about the nebula theory has concerned exactly this point during a few decades now. Some models estimate the timescales for the growth of the grains to be too long that they will ever manage to grow into forming planets. If our theory cannot predict that the dust formation and growth happens fast enough, all the solid material in the nebula would spiral into the Sun instead of forming planetesimals and planets, and we would have to abandon the theory. Maybe our modeling will eventually show that the right size grains are formed during the right time scales only if the conditions in the nebula are within narrow limits, and therefore only these, maybe rare, nebula will form planets of the right kind. Then we live in this solar system only because it had these rare conditions, and then we would predict that life is a very rare phenomenon in the universe – we still do not know!

2.9.1 The size of the dust grains

Envision now that a tiny dust grain has formed for example by the mechanism described in the previous paragraph. When the particle has reached a radius a_p, it will have a mass

$$m_p = \frac{4}{3}\pi a_p^3 \rho_p \tag{2.94}$$

where ρ_p is the density of the material it is made from.

The particle will be influenced by a force due to the proto-solar gravity (i.e., the symmetrically distributed mass inside the grain's radial distance from the nebula mass centre), pointing toward the nebula mass centre. When we put this mass to $\approx 1\mathrm{M}_\odot$ we think of the solar system formation. During the collapse phase and the FU Orionis phase, the proto-solar mass will have been very small, but most of the dust and planetesimal formation took place during the T Tauri phase, where we can approximate the solar system as a proto-solar central object of $\approx 1\mathrm{M}_\odot$, surrounded by a nebula of much smaller mass.

The radial component, F_r, of the gravitational force, $F_\odot$, will be approximately balanced by the centrifugal force, and the particle will therefore fall basically vertical through the nebula toward the mid-plane. The vertical component, F_z, of the force can be written as

$$\begin{aligned} F_z &= -\frac{GM_\odot}{r^2+z^2} m_p \frac{z}{\sqrt{r^2+z^2}} \\ &\approx -\frac{GM_\odot z m_p}{r^3} \end{aligned} \tag{2.95}$$

Kepler's 3$^{\mathrm{rd}}$ law in Newton's formulation can be written

$$\frac{T^2}{r^3} = \frac{4\pi^2}{G(M_\odot + m_p)} \tag{2.96}$$

which allow us to express the angular velocity as

$$\omega^2 = \left(\frac{2\pi}{T}\right)^2 \approx \frac{GM_\odot}{r^3} \tag{2.97}$$

which for $r^2 + z^2 \approx r^2$ and introduced into Eq. 2.95 gives

$$F_z = -\omega^2 z m_p \tag{2.98}$$

When the particle has attained a velocity $v_z = dz/dt$ vertical "down" toward the mid-plane, the viscosity of the gas will assert a drag force, F_{vz}, on the particle directed vertical upward, and given by

$$F_{vz} = \frac{4}{3}\pi a_p^2 \rho_{gas} v_z c_s \tag{2.99}$$

where ρ_{gas} is the density of the gas, and c_s is the local sound speed.

If the particle is moving not too fast, we can consider it in quasi-equilibrium, $F_{vz} = -F_z$, and write

$$\begin{aligned} \frac{4}{3}\pi a_p^2 \rho_{gas} v_z c_s &= -\omega^2 z m_p \Leftrightarrow \\ v_z &= -\frac{3 m_p \omega^2 z}{4\pi a_p^2 \rho_{gas} c_s} \\ &= -\frac{a_p \rho_p \omega^2 z}{\rho_{gas} c_s} \end{aligned} \tag{2.100}$$

where ρ_p is the density of the particle.

If the particle is spherical, it will meet the gas with an area πa_p^2. Let now the density of dust in the gas, ρ_{dust}, be constant in time and constant through the vertical path the particle travels, and let the probability for this dust to stick to the particle during collisions be P_{pd}. Then the growth $a_p(z) - a_p(z_0)$ of the particle-radius when it falls from initial height z_0 to the height z, can be found by integrating the growth rate da_p of the particle radius, which can be found from the rate of growth dm_p/dt per time unit of the mass of the particle:

$$\begin{aligned} \frac{dm_p}{dt} &= -\pi a_p^2 v_z \rho_{dust} P_{pd} \quad \Rightarrow \\ \frac{4}{3}\pi \rho_p d a_p^3 &= -\pi a_p^2 \rho_{dust} P_{pd} \frac{dz}{dt} dt \quad \Rightarrow \\ \frac{4}{3}\rho_p 3 a_p^2 \int_{z_0}^{z} da_p &= -a_p^2 \rho_{dust} P_{pd} \int_{z_0}^{z} dz \quad \Rightarrow \\ a_p(z) &= P_{pd} \frac{\rho_{dust}}{4\rho_p}(z_0 - z) + a_p(z_0) \end{aligned} \tag{2.101}$$

If the Earth is made up from dust accreted in the area between its present orbit and half way to respectively Venus and Mars, then there must have been at least $M_\oplus$ = $6\,10^{27}$g distributed over $A_{acr\oplus}$ = $\pi\left(\frac{1.5+1.0}{2} - \frac{0.7+1.0}{2}\right)$AU^2 = $5.9\,10^{26}$cm^2. The surface density (i.e. the total amount of mass in a cylinder of cross-section 1 stretching perpendicularly through the disk) of dust in the region of the Earth's orbit has therefore been

$$\begin{aligned} \Sigma_{dust} &= 2z_0\rho_{dust} \\ &= \frac{M_\oplus}{P_{acr}A_{acr\oplus}} \\ &= 10\mathrm{g\,cm}^{-2}/P_{acr} \end{aligned} \tag{2.102}$$

where P_{acr} is the total probability of the dust in the Earth's region to be incorporated into the Earth.

If the fraction of dust from the disk which is eventually included into the planets is determined by the efficiency of the falling particles to accrete the dust, then $P_{acr} \approx P_{pd}$. We will furthermore set $\rho_p \approx 1$ g/cm^3. Considering the approximate character of the computations, and that the first small particles probably are a bit fluffy, it is a reasonable estimate. A particle falling from $z = z_0$ (i.e., $a(z_0) \approx 0$) to the mid-plane ($z = 0$), will therefore have reached a radius $a(z)$ at the mid-plane which can be computed by combining Eq. 2.101 and Eq. 2.102,

$$\begin{aligned} a(z) &= \frac{\Sigma_{dust}}{8\rho_p}P_{pd} \\ &\approx \frac{10g\,cm^{-2}}{8g\,cm^{-3}}\frac{P_{pd}}{P_{acr}} \\ &\approx 1\,\mathrm{cm} \end{aligned} \tag{2.103}$$

Primitive chondritic meteorites consist on average of approximately 70% chondrules and 30% matrix. While the matrix seems to be agglomerated unprocessed dust from the nebula, the chondrules are small round particles from mm to cm size, which by an unknown process have been rather quickly heated and cooled again in the nebula. Seemingly 70% of the material that formed the terrestrial planets have been chondrules. It is fascinating to realize that the simple theory outlined above predicts that the bulk of the material that makes up the terrestrial planets should come from particles of exactly the chondrule size, based on a natural consequence of the dust accumulation that have taken place while these particles settled through the primitive solar nebula.

2.9.2 The timescale for settling of the dust.

We will now estimate the time it has taken, in our scenario, for the dust to fall from the height $z = z_0$ in the nebula to the mid-plane ($z = 0$). It is important for the credibility of the scenario we are investigating for the formation of the solar system, to argue that this time is short enough that it can take place within the nebula lifetime. Furthermore, particles live a dangerous life from the moment they decouple from the movement of the gas (i.e., when they become big enough that collisions with the gas molecules no longer affect their motion enough to make them feel the gas pressure gradient) to the time when they are big enough to not be dynamically affected by the momentum of the solar radiation or the gas drag force. During this intermediate stage in the dust size evolution, the particles are dragged toward the Sun, and will be lost for planet formation if the growth time is not short enough compared to these destructive time scales. For example, a 1 meter boulder will spiral from 1 AU to the Sun due to nebula drag force, on time scales of 100 to 1000 years.

From Eq. 2.100 we have an expression of the settling speed v_z. In order to calculate how long time it will take for the dust to reach from the outer parts of the nebula down to the mid-plane, we need also to estimate how far this is. If the nebula is in hydrostatic equilibrium, then

$$\frac{dP_{gas}}{dz} = -F_z \tag{2.104}$$

The vertical component of the solar gravitational force, F_z, on a volume of gas can be described in the same way as was done in Eq. 2.95−2.98, and we can therefore write Eq. 2.104 as

$$\frac{dP_{gas}}{dz} = -\omega^2 z\rho(z) \tag{2.105}$$

which can be integrated to give the gas pressure $P_{gas}(0)$ in the mid-plane

$$\begin{aligned} P_{gas}(0) = P_{gas}(0) - P_{gas}(h) &= \int_h^0 \frac{dP_{gas}}{dz} dz \\ &= -\int_h^0 \rho(z)\omega^2 z dz \end{aligned} \tag{2.106}$$

Approximating $\rho(z)$ with a constant ρ, allow us to perform the integration analytically, and obtain

$$\begin{aligned} P_{gas}(0) &\approx \rho\omega^2 \frac{1}{2} \left[z^2\right]_0^h + P_{gas}(h) \\ &\approx \frac{1}{2}\rho\omega^2 h^2 \;\Rightarrow \\ h &= \sqrt{\frac{2P_{gas}(0)}{\rho\omega^2}} \approx \sqrt{2}\frac{c_s}{\omega} \end{aligned} \tag{2.107}$$

Since $\omega = \omega(r)$ is known from Kepler's third law ($\omega(r) = \sqrt{GM/r^3}$) one can estimate $h = h(r)$ from Eq. 2.107 to give

$$h(r) \approx c_s/\omega \approx c_s\sqrt{r^3/GM} \tag{2.108}$$

which is (probably) the line that has been popular to draw in recent years to illustrate the form of the solar nebula, as opposed to the drawing which was popular previously where the opacity contour was more often drawn. Whereas the former show a nebula which becomes thicker outward, the latter shows a nebula which becomes thinner outward, looking more like two saucers in the way it is still popular to draw for example our Milky Way galaxy. Although it is rarely explained what the authors draw, it should be mentioned that nothing has changed in our overall view of what the nebula must have looked like, just what it is in fashion to draw.

In order to estimate how long time it takes a growing dust grain to fall from distance h to the mid-plane, we will approximate this time scale as $\tau = h/v_z$. By combining this approximation with Eq. 2.100, we therefore have

$$\frac{h}{\tau} = \frac{< a > \rho_p \omega^2 h}{\rho_{gas} c_s} \tag{2.109}$$

Introducing Eq. 2.107 into Eq. 2.109 gives

$$\begin{aligned} \frac{h}{\tau} &\approx \frac{< a > \rho_p \omega}{\rho_{gas}} \;\Rightarrow \\ \tau &\approx \frac{h\rho_{gas}}{< a > \rho_p \omega} \\ &\approx \frac{\Sigma_{dust}}{2\eta\rho_p < a > \omega} \end{aligned} \tag{2.110}$$

where $\eta = \rho_{dust}/\rho_{gas} \approx 10^{-2}$ is the ratio of dust to gas, and $< a >$ is the "average" radius of the particle ($< a > \approx$1mm because the particle falls slowest when it is smallest, and end with $a(0) \approx$1 cm in the mid-plane). Note that ρ_{dust} and ρ_p are two different things. The first describes the relative amount of dust in the gas, while the latter is the material density of the individual dust grains. Entering the numbers obtained above into Eq. 2.110 gives

$$\tau \approx \frac{10[\mathrm{g/cm^2}]}{2\,(10^{-2})\,1[\mathrm{g/cm^3}]\,1[\mathrm{mm}]\,2\pi[\mathrm{yrs^{-1}}]} \approx 1000 \text{ years} \tag{2.111}$$

We summarize from the above, that dust grains in the Earth's distance from the Sun can fall through the nebula from the outer region of the nebula and "down" to the mid-plane in about 1000 years, and accumulate enough material during their descent to grow to chondrule sized "gravel" of mm to cm size.

For collisions with relative velocities larger than $v_{rel} \approx$ 1km/s, the grains will typically fragment instead of growing, and only for $v_{rel} \leq$10m/s will the grains stick to one another. For v_{rel} in-between these values, the grains may bounce one another. Fractal structures ("snowflakes") will allow sticking for a larger v_{rel} range. For $h \approx r/10$ and $r \approx$1AU, "typical" velocities of the grains falling through the nebula toward the midplane, will be $v_z \approx h/\tau \approx 1.5\,10^{12}$cm/$10^3$yrs = 50 cm/s, so the dust grains in the Earth's region of the nebula will have no problem in growing during the drift toward the nebula mid-plane. When the dust grains reach the mid-plane, velocities will be a bit higher than the "typical" velocity, which may add to preventing the chondrule size to exceed $\sim$ cm size.

2.10 The mechanism behind the debris disks, the zodical light and the spread of cometary tails.

Once the planetary system is formed, it is of course prune to collisions between the objects, and such collisions can cause a renewed population of small grains. This time, however, the dust grains will sit in a gas-free environment – the original nebula will have disappeared. Therefore also the forces active on the grains are now not dominated by drag against the gas, but by a somewhat counter-intuitive force called the Poynting-Robertson drag. This force is responsible for cometary dust tails drifting inward once they have spread in space, and for debri disks like the Vega and β Pegasus disks.

When a photon is absorbed by a particle orbiting the Sun, it will transfer momentum to the particle, which will make the particle feel a bit smaller effective gravitational field than if there were no radiation field, and the particle would therefore, if nothing else happened, move outward. This is what we see as the first thing when dust grains are released from a cometary nucleus and the solar radiation pushes the dust away from the Sun and the comet, as the cometary (dust)tail. However, provided the particle is decoupled from the gas, such that the absorbed energy cannot be transfered to the gas via collisions, the absorbed energy must later be re-emitted as infrared photons (heat). This radiation will be isotropic seen from the reference of the particle, but since the particle has a systematic movement (an orbit) relative to the Sun, the momentum transfer in the forward direction of the orbit will be larger than in the backward direction. In total the particle will therefore lose orbital angular momentum due to the radiation, and therefore move inward. In summary, the solar radiation on a small particle will first push it outward, as intuition would tell us, but then drag it inward. This counter-intuitive effect is called the Poynting-Robertson effect, and is probably the most strange, but very effectfull, of the non-gravitational forces.

To understand this effect mathematically, consider a completely black, spherical particle of radius a, mass m, and density ρ, moving around the Sun in a circular orbit of radius R, and with orbital velocity $v = \omega R$. It will per time unit absorb the energy

$$E_{in} = \frac{L_{\odot}\pi a^2}{4\pi R^2} \tag{2.112}$$

where $L_{\odot}$ is the solar luminosity (i.e., the energy radiated from the whole surface of the Sun per time unit in all directions).

Half of this energy will be re-radiated forward in the orbital movement, and half of it backward. After one time unit, forward as well as the backward radiated energy will have moved the distance c in space, but since the particle will have moved the distance v in space, the volume in front of the particle which has been filled up by the emitted radiation is smaller than the one behind the particle. A rigorous integration of the asymmetry in the energy density (and hence pressure) due to the movement can be found in a number of papers in the scientific literature from the late 1970's and early 1980's, where also non-circular orbits, non-blackbody particles, etc, are considered. However, by approximating the integral with a cylinder in the forward direction and a cylinder in the backward direction, each of area πa^2, we will, for symmetry reasons, reach the same result (for the circular orbit, blackbody particles, etc. we are considering here). The volume in front of the particle filled up with radiation per time unit is then $\pi a^2(c - v)$, while the volume filled up

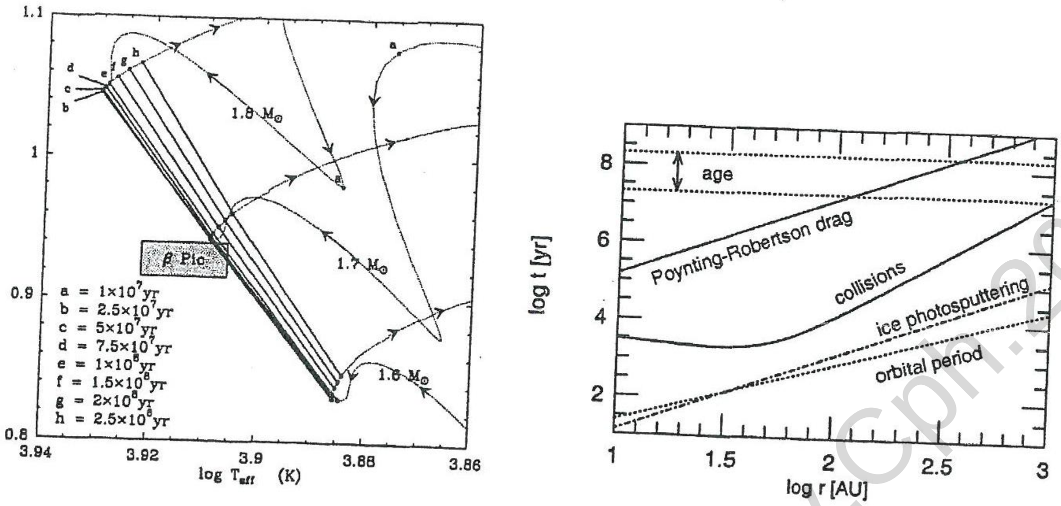

Figure 2.7. Left panel: Evolutionary tracks and ages of stars of 1.6, 1.7, and 1.8 $M_{\odot}$, and error-box for position of βPictoris in the HR-diagram. Right panel: Dust lifetime versus logarithm of distance (in AU) to the central star (βPictoris itself) for three different possible destruction processes. Also shown are the Keplerian orbital period and limits on the estimated age of the observed disk.
%vspace-1.0cm

behind the particle is $\pi a^2(c+v)$. The energy density E_f in front of the particle and the energy density E_b behind the particle will therefore be

$$\begin{aligned} E_f &= \frac{E_{in}}{2}\frac{1}{\pi a^2(c-v)} \\ E_b &= \frac{E_{in}}{2}\frac{1}{\pi a^2(c+v)} \end{aligned} \tag{2.113}$$

The energy density (e.g. in unity erg/cm^3) is the same as the force per area (e.g. in units erg/cm^3 = g cm^2 s^{-2} cm^{-1} cm^{-2} = g [cm/s^2] /cm^2). The resulting tangential force on the particle due to the isotropic radiation, is therefore

$$\begin{aligned} F_t &= \pi a^2(E_f - E_b) \\ &= \pi a^2 \frac{E_{in}}{2\pi a^2}\left(\frac{1}{c-v} - \frac{1}{c+v}\right) \\ &= \frac{E_{in}}{2}\left(\frac{c+v-(c-v)}{c^2-v^2}\right) \\ &\approx \frac{L_{\odot}\pi a^2}{8\pi R^2}\frac{2v}{c^2} \\ &= \frac{L_{\odot}}{4c^2}\frac{a^2\omega}{R} \end{aligned} \tag{2.114}$$

The particle angular momentum L and its derivative, are

$$L = mvR = m\omega R^2 \quad \Rightarrow \tag{2.115}$$

$$\begin{aligned} \frac{dL}{dt} &= mR\frac{dv}{dt} \\ &= -RF_t \\ &= -\frac{L_{\odot}a^2\omega}{4c^2} \end{aligned} \tag{2.116}$$

By use of Kepler's 3$^{\rm rd}$ law in the form $\omega = \sqrt{GM_\odot/R^3}$ one can also write the angular momentum and its derivative as

$$\begin{aligned} L &= m\sqrt{GM_\odot R} \Rightarrow \qquad (2.117) \\ \frac{dL}{dt} &= \frac{m}{2}\sqrt{\frac{GM_\odot}{R}}\frac{dR}{dt} \\ &= \frac{1}{2}mv\frac{dR}{dt} \qquad (2.118) \end{aligned}$$

Equating the two expressions for dL/dt above, allow us to find the time t it takes for the grain to spiral from distance R_1 to R_0 from the Sun.

$$\begin{aligned} \int_{R_0}^{R_1} Rdr &= -\frac{L_\odot a^2}{2c^2 m}\int_0^t dt \Rightarrow \\ t &= \frac{c^2}{L_\odot}\frac{m}{a^2}(R_1^2 - R_0^2) \qquad (2.119) \\ &= \frac{4\pi c^2}{3L_\odot}a\rho(R_1^2 - R_0^2) \end{aligned}$$

If we set R_0=0, we get the time it will take for a particle to spiral into the Sun from distance R_1 due to the Poynting-Robertson drag. With a in cm, ρ in g/cm^3 and R_1 in AU, Eq. 2.119 becomes

$$t_{[yrs]} = 7\,10^6 a_{[cm]}\rho_{[g/cm^3]}R^2_{[AU]} \qquad (2.120)$$

Meter-sized boulders are therefore basically unaffected by the Poynting-Robertson effect (on timescales of the solar system formation), but we see that a dust grain of a=1cm and ρ=3g/cm^3 will spiral into the Sun from the Earth's orbit in 20 million years, and a 1mm-sized grain of ρ=1 will take only 700,000 years to reach the Sun. We also see that our assumption above (that the grains that condense from the nebula will fall almost vertical down to the mid-plane) is justified; from the time they decouple from the gas until they reach the mid-plane as cm-size grains, they will have been dragged relatively little inward, and we found in Eq. 2.111 that they will grow to cm-size in $\approx$ 1000 years, at which time they will be safe from Poynting-Robertson drag over millions of years. They further have to grow from cm size to km size on timescales considerably less than 20 million years, in order that planetesimals can form before the cm sized chondrule-type building material has been dragged away.

On a clear night right after sunset, we can see a dim cone of light stretching away from the Sun in the sky along the plane of ecliptica. This is the zodiacal light; i.e. sunlight reflected on a disk of dust in the plane of our own solar system (see Fig. 2.8). The size of the zodiacal dust particles are typically 20 – 200 μm, and Eq. 2.120 therefore tells us that they will spiral into the Sun from Earth's orbit in 10^4 to 10^5 years (and from Saturn in less than 10 million years). It is therefore clear that the zodiacal dust disk cannot be left-over dust from the solar system formation, but it must come from a renewable source. We already stated this in last chapter, but now we have seen why. Small amounts of dust is continuously added to our solar system from evaporating comets on their journey around the Sun, probably from colliding asteroids, and maybe even from other sources too, such as the volcanic eruptions of Io, and they must be responsible for keeping this, the solar debris disk, alive.

It is also clear, from the same argument, that the dust-disk around Vega cannot be the disk in which planets formed, and even the famous dust-disk around the star β Pictoris can only marginally be related to a planetary system under formation. Figure 2.7 show the time scales for dust in the β Pictoris disk, calculated based on Eq. 2.119, and it is seen that at least the inner part of the β Pictoris disk would already have been dragged into the star if the disk wasn't a renewable phenomenon like our evaporating comets and zodiacal light, but millions of times stronger than our zodiacal light disk. Are they rather the sign of huge collisions in the outer parts of a Kuiper-like belt of icy objects, now delivering the building blocks to still boone dry rocky planets in the inner parts of the β Pictoris system, signaling the rise of biology around one more star in the Milky Way?

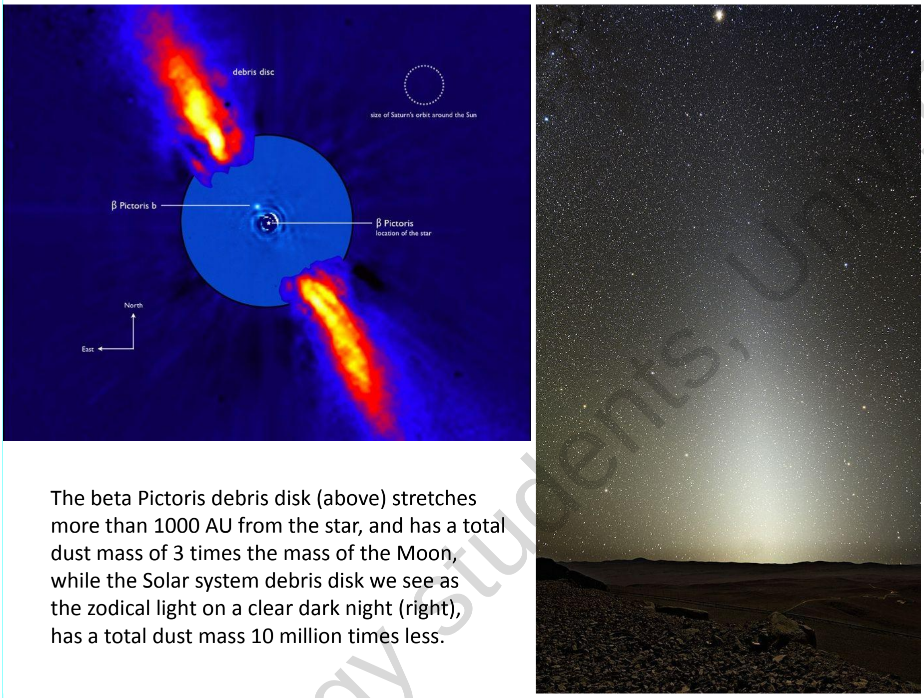

Figure 2.8. Right panel: The zodiacal light is reflected sunlight on a debris disk in our own solar system, caused by dust evaporated from comets and dust from colliding asteroids. It looks like a cone of "Milky Way" light stretching up from the Sun when it is a bit under the horizon, and it is seen best before sunrise during autumn or after sunset during spring. At dark places it can be seen stretching far up in the sky, and can in principle be traced as a band all around the sky along the ecliptic. Left panel: The debris disk around β Pictoris is a phenomenon similar to our zodiacal light, but at a larger orbital distance than our zodiacal light, and of a considerable larger total mass. The inset to scale in the middle of the occulting disk show a direct image of the jupiter-mass planet believed to cause the wrap in the disk.