

Chapter 10: The formation of the solar system

- radioisotopic dating
- formation of the solar system; exoplanets and young star systems
- sample problems

Radioisotopic dating

Many isotopes turn out to be radioactive, meaning that they are likely to *decay* by one of several mechanisms. The first step in the decay of ^{238}U is an *alpha* decay, meaning that the uranium atom spits out an alpha particle, which is basically a helium nucleus, leaving behind ^{234}Th . There are other possible decay paths that result in the emission of one or more nucleons – for instance, an unstable nucleus could emit a neutron (e.g., ^5He emitting a neutron and becoming ^4He) or it could *fission* into two or more chunks (that tend to include pieces that are larger than alpha particles). Decay by proton emission is less common, tending to occur in lab-generated isotopes. *Beta* decay refers to a decay process involving electrons. This would create a matter – anti-matter imbalance, so beta decays also involve neutrinos. Two of the last few steps in the decay of ^{238}U involve beta decay – ^{210}Pb beta decays to ^{210}Bi which beta decays to ^{210}Po which alpha decays to ^{206}Pb , which is stable. Lead, bismuth, and polonium are elements number 82, 83, and 84. In β^- decay one of the neutrons in the nucleus become a proton and we move up the periodic table. There are other possibilities involving electrons, such as positron (β^+) emission or electron capture. Alpha particles, which are more massive, tend to be less energetic than beta particles; the former can be stopped more easily, often by a piece of paper, whereas the latter are likely to take a few-millimeter-thick sheet of aluminum to stop.

Radioactive elements decay with a characteristic half-life, which describes the amount of time it takes for half of a sample to decay. Radioactive half-lives vary tremendously, from fractions of a second to billions of years. Relatively long-lived radioactive isotopes are useful for age dating the material in which they are incorporated. You are probably familiar with “carbon dating”: ^{14}C has a half-life of 5,700 years. That’s fine for dating human artifacts, lousy for dating 4.5 billion-year-old meteorites. Some more astronomically useful radioactive isotopes and their decay products are:

^{40}K decays into ^{40}Ar (or ^{40}Ca)	half-life 1.25 billion years
^{238}U decays into ^{206}Pb	half-life 4.5 billion years
^{87}Rb decays into ^{87}Sr	half-life 49.7 billion years

Let’s look at how this works. During each half-life, one half of the remaining amount in the sample decays. Start with 100% of something; after one half life there’s 50% of it left, after two half lives, 25% of the original is left, and so on.

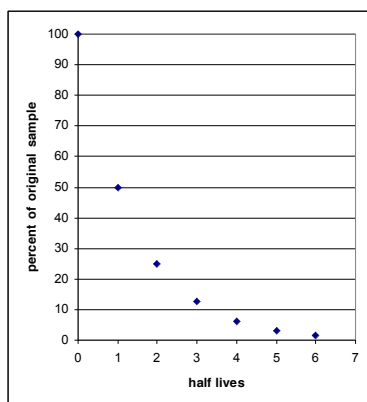


Figure 10.1: Radioisotopic decay

Here's the way the math works, both in base 2 and in base e .

$$\frac{n}{n_0} = 2^{-t/t_{1/2}}, \text{ where } t_{1/2} \text{ is the half-life.}$$

$$t = t_{1/2} \cdot 3.322 \cdot \log_{10} \left(\frac{n_0}{n} \right)$$

where we've used the following relations for log conversions:

$$\log_{10} x = \log_{10} 2 \cdot \log_2 x;$$

$$-\log \left(\frac{n}{n_0} \right) = +\log \left(\frac{n_0}{n} \right).$$

Calculus alert: $-dn = \lambda \cdot n \cdot dt$. Integration gives

$$\ln \left(\frac{n}{n_0} \right) = -\lambda t \text{ and, taking exponents,}$$

$$\frac{n}{n_0} = e^{-\lambda t}, \text{ where } \lambda \text{ is the decay constant, which is}$$

inversely proportional to the mean lifetime. λ and $t_{1/2}$ are related: $n = n_0 / 2$ means finding t such that $\ln 2 = -\lambda t$; in other words, $t_{1/2} = (\ln 2) / \lambda = 0.693 / \lambda$

Using this to get an age requires that you have some idea what the initial abundance of the decay product was in your sample. It would be handy, for instance, if there were initially no ^{87}Sr anywhere in the universe and the only ^{87}Sr in a sample came from the decay of ^{87}Rb . We aren't that lucky. If you have multiple samples, with different initial amounts of ^{87}Rb , though, you may be able to extrapolate back to $t = 0$. Let's look graphically at how this works:

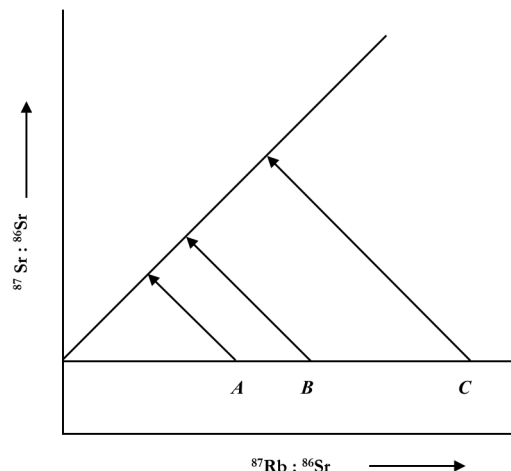


Figure 10.2:
Sample isochron

We have three samples, A , B , C . They all came from the early solar system and had the same initial value for the ratio of $^{87}\text{Sr} : ^{86}\text{Sr}$. It's possible that during the time when these elements were condensing and getting incorporated into rocks that they experienced some amount of chemical fractionation. Suppose that they were incorporated into objects large enough (large asteroids, say), and therefore hot enough, to have molten interiors. Different minerals crystallize out of a melt depending on the temperature, pressure, and composition of the melt. As some minerals crystallize out, the composition of the remaining melt, and therefore the next minerals to crystallize, is likely to change. In our example, this means that the samples are likely to have different initial values for the ratio of $^{87}\text{Rb} : ^{86}\text{Sr}$. Gravity can result in some mass fractionation also. We hope that our mineral samples all solidified with the same initial $^{87}\text{Sr} : ^{86}\text{Sr}$ ratio and different amounts of ^{87}Rb . The horizontal line in the sketch represents the values these ratios have for the three samples at time t_0 . As time goes by, ^{87}Rb atoms decay into ^{87}Sr . One ratio goes down as the other goes up; the arrows sloping up and to the left indicate the changes. The solid line running up to the right is called an *isochron*. It gives the values for the isotope ratios at time t_{present} . You can make this sort of plot for any

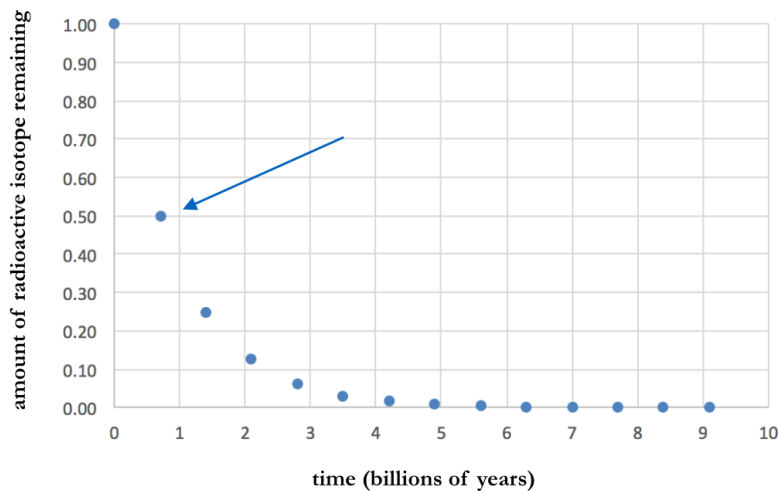
radiometric system. The slope of the line depends on how much time has elapsed and on how rapidly the particular radioactive isotope decays. The slope has a value

$$\frac{d(^{87}\text{Sr}/^{86}\text{Sr})}{d(^{87}\text{Rb}/^{86}\text{Sr})} = e^{(t_{\text{present}} - t_0)\lambda} - 1.$$

This is nifty for a couple of reasons: first, most obviously, it permits us to solve for the age of the samples; second, if any of the samples have been disturbed, i.e., have had their clocks reset by subsequent melting, fracturing, whatever, then the data points for the current isotope ratios won't all fall along a straight line, giving us a means to test whether the result is reliable. The line intercept gives us the initial value of ^{87}Sr ; using this method, we don't have to hope that it was initially zero.

Examples

1. Here is data on the decay curve for a particular isotope, both graphically and in tabular form:



x	y
0.0	1
0.7	0.5
1.4	0.25
2.1	0.125
2.8	0.0625
3.5	0.03125
4.2	0.015625
4.9	0.0078125
5.6	0.00390625
6.3	0.001953125
7.0	0.000976563
7.7	0.000488281
8.4	0.000244141
9.1	0.000122070

a) What is the half-life of this isotope?

The vertical axis gives the fraction of the isotope remaining; the added arrow points to the data point for 0.5; in the table, we can see that the age corresponding to a value of 0.5 is 0.7 billion years. Thus our isotope has a half-life of 0.7 billion years.

b) What percentage of the original isotope remains after 5 half-lives?

In the table we can see that at 5 half-lives 0.03125 of the original sample will remain. Does that make sense mathematically? $(\frac{1}{2})^5 = 0.03125$. Check.

2. Consider a one-step decay chain, where a radioisotopic isotope with a half-life of 1.3 billion years decays to one stable isotope. Assume that we have a sample that we have reason to believe initially had none of the decay product. The rock now contains $\frac{1}{3}$ rd as much of the stable product as it does of the radioactive parent isotope. How old is the rock?

First, it might help to translate the words into a table:

	initially	now
product	0	1
parent	4	3

Let's do this two ways, once in base 10 and once in base e .

$$t = t_{1/2} \cdot 3.32 \cdot \log_{10} \left(\frac{n_0}{n} \right) \rightarrow t = 1.3 \cdot 10^9 \text{ y} \cdot 3.32 \cdot \log \left(\frac{4}{3} \right) = 5.4 \cdot 10^8 \text{ y}.$$

$$\ln \left(\frac{n}{n_0} \right) = -\lambda t \rightarrow \ln \left(\frac{3}{4} \right) = - \left(\frac{0.693}{1.3 \cdot 10^9 \text{ y}} \right) \cdot t \rightarrow t = 5.4 \cdot 10^8 \text{ y}.$$

Some of the radioactive isotopes that are useful for dating old solar system rocks are not quite so easy to use as in this example because there are multiple decay paths. For instance, ^{40}K decays into either ^{40}Ca or ^{40}Ar . In that case we need to account for the branching ratio, i.e., the fraction of the decays that end up in each of the stable isotope products.

The oldest rocks on Earth are not going to be as old as the solar system because Earth has been too active. There are rocks older than 3 billion years in several places on Earth; some in northern Canada may be as much as 4.16 billion. Some of the oldest minerals are crystals of zircon, which often incorporate uranium atoms, that weather out of their original rocks and are subsequently incorporated into younger rocks. Researchers have dated some zircons from the Jack Hills of Western Australia at 4.4 billion years. Some of the most primitive chondrites are ~4.55 billion years old. Analysis of zircons in Apollo 14 lunar samples put the age of the Moon at 4.51 billion years. Earth to Moon and back? The composition of one piece of rock embedded in an Apollo 14 lunar sample, including bits of granite and quartz, suggests that it could be a terrestrial meteorite, i.e., a rock blasted off Earth during an impact that subsequently landed on the Moon and got incorporated into that Apollo 14 lunar rock sample. That rock fragment also dates to about 4 billion years, similar to those oldest Canadian rocks. Given that models suggest that the solar system took a few hundred million years to form, that 4.55 billion years is probably about as good an age as we can get for the age of the solar system.

Formation of the solar system

A compelling model for the formation of the solar system should be able to account for the various properties we've been noting over several chapters. For instance:

- Age. . .
 - the oldest meteorites are close to 4.6 billion years old;
 - the Sun and planets seem to have formed on similar timescales
- Angular momentum. . .
 - all the planets orbit the Sun in the same prograde direction in low-inclination orbits;
 - the Sun has ~99.9% of the solar system's mass but only 0.5% of the angular momentum
- Inclinations. . .
 - the Sun's equator is tilted ~7° to the mean plane of the planets' orbits;
 - with the exception of Jupiter and Mercury, the planets (not just Uranus and Venus) have noticeable obliquities;
 - many asteroids and TNOs have substantially inclined orbits
- Compositional differences. . .
 - meteorite inclusions vary in isotopic abundances and in formation temperature;
 - outer planets are rich in volatiles;
 - volatiles such as water have to get to the Earth and other objects inside the "snow line"
- Satellites. . .
 - the giant planets all have collections of regular satellites, like miniature solar systems;
 - giant planets also have many irregular satellites and ring systems;
 - the density of the Galilean satellites decreases with distance from Jupiter;
 - Earth's Moon orbits approximately in the plane of the ecliptic
- Minor planets; in particular, why do the asteroid and Kuiper belts exist

- Comets; in particular, how is the Oort cloud formed and maintained
- AND all the while keeping in mind the fact that there are, as of July 2026, over 8,000 known exoplanets, including ~1100 multi-planet systems (and several thousand more candidate planets awaiting confirmation). Many of these systems are very different from ours; explanations that successfully explain our solar system while predicting the “impossibility” of other systems known to exist should be viewed with suspicion!

Exoplanets. The division of solar system planets into small rocky objects near the Sun and giant planets beyond 4-ish AU seems to make sense. *If* the planets formed not too far from their current locations then it is reasonable for the colder outer planets to have a larger complement of volatiles. “Hot Jupiters” provide a challenge to that picture. Searches have turned up more than a few exoplanets with masses similar to Jupiter but orbit sizes less than one-tenth of an AU. Below is a plot showing the characteristics of a few of the known exoplanets (not including the very large number of detections from the Kepler mission) from the exoplanet.eu database. Out of ~8,000 known exoplanets, approximately 200 such objects are known, around stars similar to the Sun, that have masses greater than ~half the mass of Jupiter and orbit periods less than 10 days. To confuse matters further, some of these large planets are known to revolve in the opposite direction to the rotation of their host star.

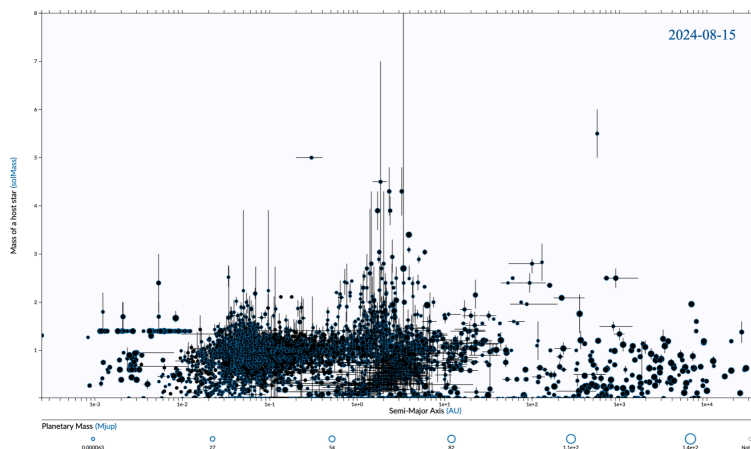


Figure 10.3: Characteristics of some known exoplanets. Data from the Extrasolar Planets Encyclopaedia <https://exoplanet.eu/plots/>

Exoplanets may be detected in several ways. The Kepler spacecraft, which has been the largest single source of exoplanet detections, observes transits. If a planet orbit is aligned so that from our point of view the planet will periodically transit across the disk of its host star, then, provided the planet is large enough, we will detect a dip in the star’s brightness during the transit. Planets need to transit multiple times to verify that we aren’t just seeing passing starspots. Kepler is followed by TESS, the Transiting Exoplanet Survey Satellite, which launched in April 2018 and began science observations in July 2018. The following figure illustrates what the light curve for a star might look like during a planet transit.

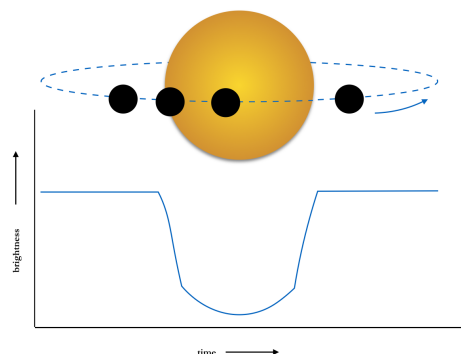


Figure 10.4: Example of an exoplanet detection due to a transit.

Note that the planet isn't likely to be contributing any light if we are observing in visible wavelengths. That means that we should be able to use the drop in the flux from the star to determine the size of the planet relative to the star. The flux we receive is proportional to the area of the stellar disk (allowing for limb darkening, stars looking brighter at disk center, which is why the bottom of the transit light curve isn't flat). Outside of the transit we receive light from an area proportional to R^2 ; during the transit we receive light from an area proportional to $R^2 - r^2$, where R is the star radius and r is the planet radius. Thus

$$\frac{flux_{\min}}{flux_{\text{outside transit}}} = \frac{R^2 - r^2}{R^2} = 1 - \frac{r^2}{R^2}.$$

The radial velocity method is used to detect planets that are large enough to tug noticeably on the host star; the resulting Doppler shift in the star's spectral lines reveals the presence of a companion. The resulting *small* Doppler shift; this method would also indicate the presence of an unseen companion black hole, which, being much more massive than a planet, would tend to produce much larger Doppler shifts in a star's spectrum. If we can estimate the inclination of the planet's orbit, measured relative to the plane of the sky, then the Doppler shift gives us the speed of the host star:

$$\frac{\Delta\lambda}{\lambda} = \frac{v_{\text{radial}}}{c}.$$

We know the period over which the spectral lines wobble back and forth so we can estimate the orbit size by using Kepler's third law and, initially at least, ignoring the mass of the planet:

$$P^2 = \frac{4\pi^2}{G(M + \sim 0)} a^3.$$

Use the orbit size and period to determine the planet's average speed. Conservation of momentum then requires that

$$m_{\text{planet}} = \frac{M_{\text{star}} v_{\text{star}}}{v_{\text{planet}}}.$$

If the mass of the planet were a significant fraction of the star's mass, we might want to iterate, using our estimation of the planet's mass to refine its orbit size, but often that won't be necessary. Proxima Centauri (or α Cen C) displays just such a Doppler wobble; astronomers at the European Southern Observatory announced the presence of Proxima Cen b in 2016. Proxima Cen is a low-mass star (type M6 V; spectral types are described in chapter 14), about 0.12 solar masses. Proxima Cen b has an orbit of approximately 0.05 AU and a mass of at least 1.27 Earth masses. In this case we don't know the orbital inclination precisely; we measure $v \sin i$ for the star and can estimate $m \sin i$, the minimum mass the planet must have.

Exoplanets don't tend to be very bright. Like planets here at home they will reflect the light of their host stars and emit themselves in the infrared. An exoplanet closer to its star and tidally locked (i.e., like our Moon, keeping one face toward its primary all the time) can have different enough temperatures from one side to the other that the difference (and thus also the tidal locking) can be detected by infrared observations. An exoplanet close to its host star might be more reflective, but get lost in the glare of its star when we try to image it directly. Coronagraphs, originally developed to block the solar photosphere so as to image the solar corona, can sometimes be used to block the starlight and reveal the exoplanet. If the planet is a bit farther out, especially if it is large and/or young and thus relatively warm, and we might be able to image it in the infrared. Twenty or so exoplanets have been imaged as of 2018.

Another twenty or so have been detected by gravitational microlensing. Recall that a consequence of general relativity is that objects with mass will warp the nearby fabric of spacetime. When a star passes in front of a more distant star, the light from the distant star will follow a curved path around the foreground lensing object. We will see the background star temporarily brighten during the days or weeks that the two are closely enough aligned on the sky. If the foreground star has a relatively massive planet, at a sufficient distance from its star that the planet's contribution to the lensing can be distinguished, then an extra bump in the light curve of the background star will indicate the presence of a planet. This method has the

disadvantage that repeat observations of the planet to confirm the detection are extremely unlikely because the stellar alignments are so infrequent. The following figure illustrates what the light curve during a microlensing event might look like.

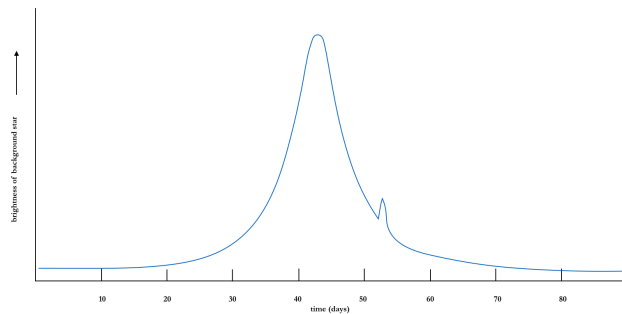


Figure 10.5: Example of an exoplanet detection during a microlensing event.

A few exoplanets have made their presence known by causing variations in the timing of some otherwise periodic event. For instance, a planet already detected by the transit method might show a variation in the periodicity of subsequent transits as it experiences slight gravitational accelerations from one or more additional planets around the same star. The first exoplanets discovered, in 1992, were detected because of changes they induce in the timing of the radio signals we receive from their host star, the millisecond pulsar PSR B1257+12. And no, we weren't expecting to find planets around pulsars, which are the neutron star remnants of supernova explosions!

It's reasonable to ask whether there are exoplanets that could support life, or at least life-as-we-know-it. Life on Earth requires liquid water, organic molecules, a source of energy, and has evolved on a planet where the temperature doesn't vary too wildly (and, relatedly, where we are not tidally locked on the Sun with only one side of our planet receiving sunlight) and where we are relatively protected by our magnetic field from high-energy particles from our star (unlike Proxima Cen b, hit by frequent flares). The term "habitable zone" has been used for many decades to refer to the region around a star where we expect liquid water could be stable on the surface of an exoplanet given the appropriate atmospheric pressure. As evidence has mounted for sub-surface oceans on several icy outer solar-system objects, it has become clear in recent years that the concept of a habitable zone needs to be updated. For those that transit, it's possible to determine the composition of an exoplanet atmosphere: outside of the transit we can record the spectrum of the host star; during the transit starlight will pass through the exoplanet atmosphere and this should result in additional absorption lines in the stellar spectrum (just as we see lines of molecular oxygen when we observe the solar spectrum from the ground). Life on an exoplanet might signal its presence by an atmospheric chemistry that was not in equilibrium. For instance, having both CH_4 and O_2 present in an atmosphere of Earth's temperature is not as likely as CO_2 and H_2O ; it requires some biological processes to maintain that disequilibrium. Some exoplanets get very close to their stars, within ~ 0.1 AU, and get very hot, thousands of degrees; we've detected atmospheres on a few of these planets, but they seem to be losing their atmospheres rapidly. Other, cooler planets, seem to have stable atmospheres. As of 2023 we've detected, not all for the same planet: water, sodium, carbon monoxide and carbon dioxide, methane, hydrogen and helium, and clouds of silicate grains. So far we haven't found any exoplanets that look as though they'd be desirable vacation destinations but it is worth continuing to look.

Young stars. The earliest stars, 13+ billion years ago, formed from hydrogen and helium, the only elements present in the early universe. Those early stars produced heavy elements, many of which were ejected at the end of the stars' lives, whence they became available to be incorporated into the next generation of stars. Clusters of stars today form from clouds of gas and dust several tens of light years across. That dust, some of which might be forming into planets, makes it tough to see what's happening because it scatters visible light. Observations in the near IR give us a better view.

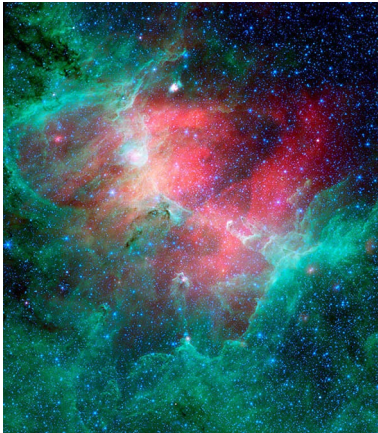


Figure 10.6: The Eagle nebula – M16

M16 is ~6,500 light years away; the colors are mapped as follows:

Blue – 4.5 μm

Green – 8.0 μm

Red – 24.0 μm

The field of view is 28.0 x 32.0 arcmin or 50-60 light years at the distance of M16.

NASA / JPL / N. Flagey & A. Noriega-Crespo; Spitzer Space Telescope

<http://www.spitzer.caltech.edu/images/1708-ssc2007-01a-Cosmic-Epic-Unfolds-in-Infrared-The-Eagle-Nebula>

The Eagle nebula, shown in the above infrared image, is an example of a location of active star formation. Some stars have already formed (blue in this image) and some have already blown up, blasting away material and leaving denser clumps and pillar of dust (green) within which new stars continue to form. The towers of dust on the right, looking faintly like a praying mantis, are sometimes called the “pillars of creation”. These columns of dust are about 5 light years tall. The following images are 2014 HST observations of these pillars taken in visible light, left, and near IR, right.



Figure 10.7a: Pillars of dust in the Eagle nebula.

O⁺, H-alpha, N⁺, and S⁺ visible wavelengths

<http://hubblesite.org/newscenter/archive/releases/2015/01/>



Figure 10.7b: Near IR image of dust pillars in the Eagle nebula

NASA / ESA / Hubble Heritage Team
– Hubble Space Telescope

Even in the infrared it's tough to see what's happening inside the densest of these dust clouds. A young star is likely to have a stellar wind that will blow away much of the surrounding dust. We do have observations of many young stars that have cleared away enough of the material in which they were embedded but are still surrounded by dust within which planets may be forming. To image such a system and see the disk requires masking the central star. Below are Hubble Space Telescope images of two of

these circumstellar disks, one edge-on and one face-on. AU Microscopium is ~12 million years old, about 0.3 solar masses; TW Hydrae is ~8 million years and also smaller than the Sun, ~0.8 solar masses. Both of these stars are variable and likely still accreting material.

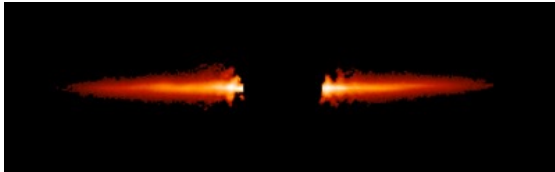


Figure 10.8: Dust / pebble disk around AU Mic; the disk is ~100 AU in diameter. Composite of many images combined into a monochromatic image. NASA / ESA / J.R. Graham, P. Kalas, and B. Matthews – HST image

<http://hubblesite.org/newscenter/archive/releases/2007/02/>

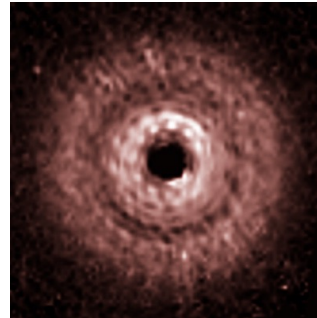


Fig 10.9: Particle disk with gap around TW Hydrae; the disk is ~320 AU in diameter. Composite of several images displayed as a monochromatic image.

NASA / ESA / STScI – PRC13-20a

<http://hubblesite.org/newscenter/archive/releases/2013/20/>

Solar nebula and protoplanetary disk. The following sketch is a cross-section of the forming solar system based on our current best models. In the center is the forming Sun. Stars form by gravitational collapse, meaning that in a nebula of dust and gas some clump of material becomes sufficiently overdense that it starts gravitationally attracting more dust and gas from many thousands of AU away. Infalling material feels a gravitational attraction toward the Sun but also toward the forming disk of material along the ecliptic. The early solar wind is constrained to blow up and out in an X direction but is inhibited along the plane of the ecliptic by the density of material. Planetesimals will accrete along the midplane of the disk.

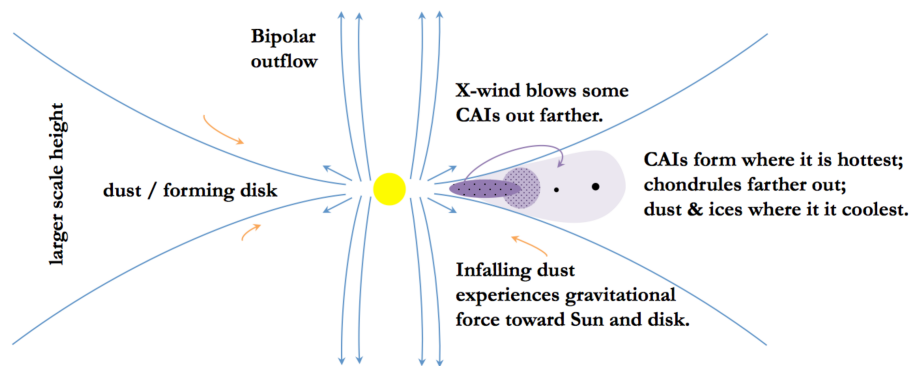


Figure 10.10: Forming solar system.

The initial cloud of dust and gas that led to the solar system was large, cool, low density, and rotating slightly. The nebula collapses toward the forming Sun and, because of conservation of angular momentum, it flattens and spins faster, leading to the development of the protoplanetary disk. The disk density increases to the point where condensation starts to happen. It's hot, perhaps 2,000 K in near the young Sun, so only the more refractory materials condense there. Some of these refractory condensates wind up in CAIs. Less refractory chondrules are likely to form farther out. It will also likely be a bit

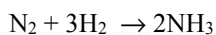
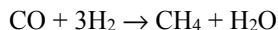
turbulent in the disk and the gas molecules and forming dust grains won't have been in nice neat Keplerian orbits around the forming Sun. Particles will have experienced viscous drag as they interact. Some material will get pulled along, speeding up, and migrate farther out in the disk. This process acts to carry angular momentum away from the material that's falling onto the Sun and increase the angular momentum in the outer regions of the disk.

Pause for a note about magnetic braking: Even if some angular momentum gets carried outward while the Sun is still forming, that isn't enough to explain why the Sun rotates so slowly. Observations of young solar-type stars indicate that most rotate much faster than our Sun and other older solar-type stars. Why do these stars slow down? A big reason is that as charged solar wind particles stream outward from the Sun they are forced by their interactions with the Sun's magnetic field to co-rotate with the solar surface. In other words, these particles can't just fall into longer-period Keplerian orbits. They are forced to go faster than a simple gravitational analysis would suggest and that means that the solar wind particles are carrying away angular momentum. One consequence is that the magnetic dynamo in a solar-type star gets less efficient as the star ages – remember that the dynamo involves an interaction between convection and rotation in an electrically conductive fluid. It basically converts rotational energy into magnetic energy. Less rotation, less dynamo action. That in turn means that the effectiveness of this braking mechanism decreases as the star ages. Young solar-type stars are more magnetically active (more starspots) and they slow quite rapidly; older solar-type stars are less magnetically active and the rate at which they slow has slowed down. Solar-type stars in the Pleiades, for example, which is about 2% solar age, rotate roughly ten times faster than stars the age of the Sun. Young stars are different from old stars in terms of their magnetic activity, but old stars are not very different from older stars, so by themselves the rotation rate and level of magnetic activity won't yield a precise age for the Sun. There are other indicators, beyond the scope of this chapter, that will tell us something about where a star is in its lifecycle. These are in accord with an age for the Sun of 4-5 billion years.

Condensation and coagulation. Back to the protoplanetary disk. Material builds up along the mid-plane of the disk. In-falling material thus feels a gravitational force toward the center of the solar nebula, where the Sun is forming, and vertically, toward the ecliptic. The areal density of material (kg / m^2) is lower farther out, meaning that the scale height of the disk will be larger out at, for instance, 30 or 50 AU than it is in at 5 – 10 AU.

In and near the disk the condensing dust grains are going to interact. Individual dust grains are too tiny to attract each other gravitationally. If the grains collided like little billiard balls they could fragment or simply bounce off each other. But remember that modern interplanetary dust grains, as we know from images, tend to be fluffy, which would tend to help them stick together. If the disk isn't too turbulent, then the grain collisions can be gentle enough that the particles grow in size fairly quickly; this sticking together process is called coagulation. How quickly the grains grow depends on how rapidly the dust settles toward the ecliptic and the density of material along the mid-plane gets high enough for particles to stand a good chance of bumping into each other. There are a range of models and estimates for how long it would take to grow pebbles several millimeters in size. Roughly speaking it should take a few thousand years in the 1 – 5-ish AU region and perhaps tens of thousands of years out around 30-ish AU, where the density of material is lower. The number of dust grains / m^2 doesn't decrease uniformly as distance increases from, say, 1 AU out to 30 AU. Recall that we expect there to be a "snow line" somewhere out near the asteroid belt. This isn't a hard and fast line, but rather the approximate location at which the temperatures in the protoplanetary disk are low enough for the first of the abundant volatiles – water, ammonia, methane – to start to condense. Interior to the snow line dust grains are metallic, silicates, hydrated silicates; beyond the snow line dust grains are all of these plus ices. Chondrules and CAIs give us insights into the nature of these early coagulating grains.

While we are talking ices and timescales let's revisit a concept about carbon and nitrogen disequilibrium chemistry (mentioned in the terrestrial interiors chapter). At low temperatures, given enough time, we expect the following reactions:



to occur. But, as chemists would say, these reactions have high activation energies, meaning that they are not going to proceed rapidly. The fact that we observe CO and N₂ ices on the surfaces and in the atmospheres of Pluto and Triton implies that conditions in the protoplanetary disk did not have time to come into chemical equilibrium before the planets formed. That's in accord with observations of disks around young stars, which seem to dissipate by the time the stars are ~10 million years old.

Accretion. The next size step, getting pebbles to accrete into kilometer-sized planetesimals, presents an issue for those modeling planet formation. If we still have quite a bit of free gas in the protoplanetary disk, that gas will have a noticeable pressure and there will be a pressure gradient in the radial direction because the density is lower further out. With a force outward partially offsetting the gravitational force inward, it is as if the gas molecules feel a somewhat lower gravity and thus they orbit the young Sun more slowly than particles. That fact means that particles feel a headwind as they catch up with the slower-moving gas. The gas headwind and the Poynting-Robertson effect due to the photons preferentially hitting on the leading edge of the particles combine to rob particles of angular momentum. Particles will spiral inward most quickly if they encounter an amount of gas approximately equal to their own mass each time they orbit the Sun. The worst affected particles are roughly a meter in size at roughly 1 AU – they should spiral in toward the Sun in a few *decades*. That's problematic; this is called the 1-meter barrier. It clearly must be possible for accretion through this size range to happen rapidly, or we wouldn't be here!

Densities and close-up images of asteroids and short-period comets may provide clues to how asteroid-sized particles accrete. These objects often look as if they are rubble piles. Recent (2015) work by Hal Levison and colleagues suggest that near 1 AU and again at 5 – 10-ish AU, i.e., beyond the snow line, the gas drag may make it possible to create large enough collections of cm-sized pebbles that they can collapse under their own gravity, effectively bypassing the 1-m barrier by growing almost instantly from cm-sized to km-sized planetesimals. Once a planetesimal is on the order of 10 km in diameter, it has enough gravity to attract smaller nearby objects by deflecting their trajectories and creating more frequent collisions. The more massive a planetesimal becomes, the more effective it is at becoming larger still, at least in the inner solar system where the density of available material is relatively high. The contact binary nature of the KBO Arrokoth (described in Chapter 4) supports the idea of pebble accretion; out beyond Neptune, objects such as this may simply not have encountered enough other planetesimals to grow to planet size.

Differentiation. As you've read before, by the time a planetesimal reaches perhaps 400 kilometers diameter it has enough gravity to pull itself into round (e.g., Mimas), and by about 1,500 km, achieve hydrostatic equilibrium (e.g., Rhea). With enough accretional and radioactive energy objects on the order of 2,500 km across are likely to differentiate, with the denser material sinking to the center to form a core. Triton, at 2,700 km, seems to be differentiated, whereas it's arguable that Callisto, at 4,800 km, isn't, so it's clear that there is no hard and fast size demarcation for differentiation.

In the inner solar system, the pool of many planetesimals shrinks to a much smaller pool of planetary embryos from which the terrestrial planets will ultimately form. As collisions become less frequent, and possibly more destructive, growth slows. Some models suggest that it might take 10 million years to grow several objects to about half the masses of the final terrestrial planets but as much as 100 million years to grow Mercury, Venus, Earth, and Mars to the masses we see today.

Solar system objects will form warm and cool over time. For instance, the amount of infrared energy Jupiter is radiating suggests that it is continuing to cool and shrink. As another example, deep inside the Earth the inner core solidified; current estimates suggest that that process took ~ 1 - 1.3 billion years.

Formation of the Moon. Our Moon is a bit of a puzzle. It should have formed at some point during that roughly 100 million year timeframe for terrestrial planet formation. Its composition is almost like Earth's but not quite: the Moon differs in that it seems to be depleted in iron and other siderophilic

elements and also in volatiles (e.g., Na and K) compared with the composition of Earth; on the other hand, it's amazingly similar in the ratios of some isotopes (e.g., of oxygen and titanium). The Moon's orbit is much closer to the plane of the ecliptic than to the Earth's equator, unlike other large moons such as the Galilean satellites of Jupiter. The current leading model for the formation of the Moon is the *giant impact hypothesis*, namely that a differentiated Mars-sized object (called Theia, for the Greek deity who was the mother of lunar deity Selene) impacts the young Earth, flinging a substantial amount of debris into Earth orbit where it accretes relatively rapidly (months? years?) into the Moon. The energy of the collision means lots of volatiles escape; the core of the impactor exits the stage, either sinking to merge with the core of the Earth or being flung away from the scene of the crash. Some models suggest that a portion of the debris formed a second, smaller trojan moon at one of the Lagrange points; that orbit eventually became unstable as tidal interactions cause the size of the Moon's orbit to increase, leading to a low-speed collision with the Moon in which its smaller sibling pancaked onto the Moon's far side, thereby explaining the fact that the Moon's farside crust is so much thicker than the near side. Several recent research projects — e.g., China's Chang'e 4 mission to the South Pole - Aitken Basin on the lunar far side and NASA's Lunar Reconnaissance Orbiter — have examined deep impacts on the Moon in order to better characterize the amounts of iron and titanium and mantle minerals such as olivine that might be hiding below the Moon's surface. Initial results suggest that the Moon may be less depleted in iron than previously thought. As with other large rocky objects, the Moon would initially have been warm inside. A combination of impacts and lava flows created the maria that we see today. Dating lunar samples and cratering density both provide insights into when the Moon cooled enough for that volcanic activity to cease. In 2020 the Chinese Chang'e 5 mission returned basalt samples from Oceanus Procellarum, known to be one of the youngest of the lunar maria. Initial results date these rocks to slightly less than 2 billion years, implying that the Moon was still able to produce lava that recently. This result, in turn, should help improve the surface chronologies that are based solely on crater counts for rocky objects that we haven't visited.

The giant impact model is not without issues. The origin of the impactor requires a bit of fine tuning to achieve lunar isotope ratios so similar to Earth's. Some researchers argue that many smaller impacts could generate enough debris to form the Moon primarily from terrestrial material. Further, a single impact of this magnitude should have melted the entire surface of the Earth, but geological evidence does not support the idea that the early Earth once had a magma ocean. Such an impact would also have imparted significant angular momentum to the Earth, resulting in the collision debris and the Earth's equator being in the same plane, requiring a bit more fine tuning to get the eventual Moon out of the plane of the Earth's equator. On the other hand, massive collisions clearly happen — several solar system objects sport the scars of impacts that must have been almost energetic enough to fracture them.

It is possible that a massive collision could explain the overdensity of Mercury. It is not odd for the uncompressed density of Mercury to be a bit higher than the uncompressed density of Earth since refractory elements tend on average to be denser and we expect the material from which Mercury formed to be more refractory since it's closer to the Sun and the solar nebula was hotter there. But Mercury's a bit too dense. One possibility is that, after Mercury differentiated, a collision fractured and ejected the outer mantle.

The oddly tilted Pluto system may also have started with a collision, although this far out in the solar system the proto-Pluto and its impactor would have hit slower, on the order of 1 km/sec. One, or rather four, pieces of evidence for this model are provided by Pluto's retinue of small moons. Styx, Nix, Kerberos, and Hydra are all low density and high albedo, as would be expected if they were made of fragments of an icy mantle of a proto-Pluto shattered by an impact. None of the four is tidally locked on Pluto and as far as we can tell so far, all four spin chaotically. We have only moderate- (Nix) and low-resolution images of these four — Styx and Kerberos were discovered while New Horizons was en route to Pluto — but we can tell that all four are elongated and clumpy looking, as if composed of pieces of debris swept up after a collision.

The giant planets. The major distinction between forming stars and forming terrestrial planets is that stars form by gravitational collapse and terrestrial planets form by accretion. What about the giant planets? Not shockingly, some models are built along the lines of gravitational collapse, some are built on accretion, and some are hybrids.

For giant planets, particularly Jupiter and Saturn with their extensive atmospheres of hydrogen and helium, accretion means core accretion. In other words, in these models the protoplanet accretes a core first. It's like the accretion of the terrestrial planets except that because both rocky material and ices are available and because the volume of space available to the forming planet is so much larger the core reaches a mass of perhaps 5 – 10 Earth masses. The core has to get large enough to start attracting, and retaining, its gaseous atmosphere before the young Sun's strong solar wind blows the remaining gas out of the protoplanetary disk. Accreting planetesimals and gas will release gravitational potential energy, some of which will heat the atmosphere and increase the outward pressure on gas molecules. To retain its atmosphere, the forming planet must, per the virial theorem, be able to radiate away half the released gravitational radiation or else the pressure will blow off any additional infalling gas. Core accretion takes longer farther out in the protoplanetary nebula because the density of material is lower; it would make sense that out too far, as in, beyond Neptune, it just isn't reasonable to form planets.

The other primary flavor of giant planet formation models is called disk instability. If there's enough mass in the protoplanetary disk and if enough variations in density develop, then some regions of the disk – smaller than the forming Sun but larger than terrestrial planets – should fragment and collapse gravitationally to form the giant planets. In this model, no core needs to form by accretion first. It's reasonable that farther out in the solar system, where the density is lower, disk instabilities don't arise and, again, we don't get planets beyond Neptune. An attraction of this model is that it might explain why the gas giants, in particular, with their retinues of satellites look a lot like miniature solar systems. A detraction is that the disk mass needs to be high, possibly higher than is realistic for our solar system (but quite possibly appropriate for other planetary systems).

Either model or a hybrid works reasonably well for explaining the giant planets' systems of moons. We expect there to be quite a bit of material, including growing planetesimals, around the forming giants, from which the regular satellites form. Irregular satellites, in distant or highly inclined orbits, are more likely to have been captured.

Migration and scattering. If the planets, particularly the outer planets, succeed in forming before the gas in the disk dissipates then the planets will interact with that gas, i.e., exchanging energy and angular momentum, and the result is that planets may migrate substantially – multiple AU in the case of the giant planets – from their original locations. Migration outward for Uranus and Neptune would mitigate one substantial problem, namely how to get planets this massive to form rapidly enough in regions where the density of available material is low. On the other hand, pebble accretion may be efficient enough that forming Uranus and Neptune in their current locations isn't a problem.

The forming giant planets don't just interact with the gas in the disk but also with planetesimals. Some, of course, they attract and accrete or capture as satellites or as Trojan asteroids. (Recall that Trojans have orbits around the Sun roughly 60 degrees ahead and behind a planet in its orbit around the Sun.) One piece of evidence in favor of a substantial migration for Jupiter (perhaps as much as 15 AU inward from its formation) is that there are ~1.5 times as many Trojans in the group ahead of Jupiter than in the trailing group; models of Jupiter's migration tend to pile up more objects in the leading camp. Other planetesimals will interact energetically enough to get scattered into different orbits. For example, a small fraction of Kuiper Belt objects have somewhat bluer surface colors than the majority, which may indicate that they formed in closer to the Neptune and were subsequently scattered into larger orbits. The outer planets may also scatter planetesimals inward, until they reach the orbit of Jupiter. Jupiter is massive enough to efficiently eject planetesimals from the inner solar system entirely. Giving energy to planetesimals would explain Jupiter's inward migration and provide a mechanism to populate the Oort cloud with $\sim 10^{12}$ few-km-sized nuclei of potential future long-period comets. Some current models suggest that Jupiter may have

migrated inward by about an AU and the other giant planets may have migrated outward, with Neptune moving several, and perhaps as much as 15, AU. Other models suggest that Jupiter and Saturn both migrated inward, bringing Jupiter in near where Mars subsequently formed, turned, and migrated back outward. Some models have shown that it is mathematically possible for Uranus and Neptune to have swapped places. Given the existence of the Oort cloud, some migration seems reasonable; how much migration is clearly not well constrained.

Migration could explain the abundance of extrasolar hot Jupiters: these planets may form farther out and migrate inward, stopping once the surrounding disk material has dissipated and the planets' orbit circularize. Scattering could explain exoplanets in extremely large orbits as well as "rogue" planets that seem to have escaped from their host stars entirely.

Recap. The solar nebula model works fairly well. In this general category of models, a nebula of dust and gas rotates and flattens, leading to the existence of a protoplanetary disk, within which some fraction of molecules condense and coagulate to form (more) dust grains, some of which then accrete into planetesimals; some of the planetesimals become planets which differentiate, some additionally capturing extensive atmospheres. Enough material is available around outer planets to explain their retinues of moons and rings, partly as objects that formed near the giant planets and partly as objects that are captured into less regular orbits. Planetesimals not incorporated into planets may stay in the asteroid belt, where Jupiter's gravity keeps things stirred up and prevents planet formation, or out in the Kuiper Belt, where the density is too low to form planets, or get ejected into the Oort cloud to await perturbation back down into the inner solar system by a passing star or giant molecular cloud. Because the planets form in a flattened rotating disk, their orbits all lie roughly in the ecliptic and most objects rotate in the same sense as well. Objects farther out, where the protoplanetary disk scale height was larger, are naturally expected to have more frequently inclined orbits. Enough collisions are expected to explain the obliquity of objects such as Venus. The expected temperatures and densities within the protoplanetary disk make it possible to understand the variation in volatile content in many objects that range in size from tiny inclusions in meteorites to the extended atmospheres of Jupiter and Saturn. Computational models can successfully get the Sun and planets to form on compatible timescales, before the Sun blows away the remaining gas in the solar nebula.

One thing that you might have noticed is missing from the above picture is the inclusion of magnetic fields. We know that they should be present, if only because the galaxy itself has a pervasive, low-level, magnetic field that would have been swept up in the protoplanetary disk. Today, the Sun, Earth, and giant planets all have magnetic dynamos that use the energy of convection and rotation to sustain a global magnetic field. Without the dynamo, any initial galactic magnetic fields concentrated in the forming Sun and planets would have rapidly (million-ish years?) dissipated. We are not absolutely sure how long it would have taken for the solar dynamo to establish itself, but it can't have been long. Observations of pre-main sequence solar-type stars (T Tauri stars, named after a variable in Taurus) seem to show strong magnetic fields, and there's no reason to think that the young Sun would have been any different. It's entirely possible that the solar nebula itself was able to sustain a magnetic dynamo. It's not clear whether the nascent planets would have had their own magnetic fields, or how strong, or exactly how being embedded in magnetic fields in the protoplanetary disk would have impacted their formation. To what extent could those early magnetic fields have pushed gas and dust around, either protecting growing clumps from being torn apart or inhibiting extra gas from piling on to the young giant planets? Increases in computing power are allowing theorists today to incorporate magnetic fields more fully into planetary formation models. This work may shed new light on the trends in planet sizes — lots of objects in the super Earth - Neptune mass range — we are seeing in the growing exoplanet database.

There are clearly a number of ad hoc features to the solar nebula model. For instance, it would be very difficult to prove that the giant planets migrated from the locations in which they formed, or that CAIs were blown outward and there incorporated into meteoroids along with the more volatile chondrules. It would be very difficult to demonstrate that the number and character of collisions in the early solar system really were adequate to explain the density of Mercury, the obliquity of Venus and of the Uranus system,

the formation of the Moon, and, possibly, the odd orbit of Triton. At many points the general solar nebula model is critically sensitive to timescales that must mesh just so, whether it's getting dust to settle to the ecliptic plane rapidly enough or getting over the 1-m barrier in moving from particles to planetesimals or forming Uranus and Neptune before the disk material dissipates. In other words, the model has its critics. That leads the model's proponents to demand more of their models; so far, none of the objections to the solar nebula model seem to be insurmountable.

Sample problems

1. After 7 half lives how much of a radioactive isotope will remain in a sample?
2. Suppose you read a report about a rock sample with a measured ratio of $^{238}\text{U} : ^{206}\text{Pb} = 3:10$. The report declares that the rock sample is from a meteorite that originated on the Moon and was retrieved from Antarctica. Do you believe the report? Why or why not? (^{238}U decays to ^{206}Pb with a half life of 4.5 billion years.)
3. Consider a "hot Jupiter" in a close orbit around a star with a mass similar to the Sun's. How close could the planet's orbit be before it would be inside the Roche limit of its host star?
4. Check that assertion that the Sun has less than 1% of the angular momentum in the solar system. You might want to put this into a spreadsheet to keep track of all the numbers.
 - a) Calculate the Sun's angular momentum; for the Sun it's the rotational angular momentum that matters. Recall that $L = I\omega$ and $I = CMR^2$ where the moment of inertia factor C varies by object, depending on how centrally concentrated the mass is. For the Sun $C \sim 0.07$. The Sun rotates differentially; for this problem use a solar rotation period of 26 days.
 - b) Check Jupiter's rotational angular momentum, using a moment of inertia factor = 0.254. This should convince you that planets' rotation probably isn't going to account for the majority of the solar system's angular momentum.
 - c) Next calculate Jupiter's orbital angular momentum, using $L = mvr$ for which the average (circular) orbital velocity will suffice. This clearly matters and perhaps other planets will also.
 - d) Calculate the orbital angular momentum for Saturn, Uranus, and Neptune.
 - e) Calculate the orbital angular momentum for the Earth to convince yourself that you probably don't need to bother with the other inner planets.
 - f) To be thorough calculate the orbital angular momentum for Ganymede to see if you need to consider planetary satellites.
 - g) Add up all the angular momenta you have calculated and determine what percentage is in the solar rotation.
5. Reading carefully? explain / define
 - a) accretion
 - b) differentiation (not the calculus variety)
 - c) giant impact model (for Moon)
 - d) chemical equilibrium (with respect to volatiles)
 - e) detection of exoplanets by transits
 - f) snow line
 - g) Oort cloud

Answers to selected questions are on the next page:

2. That “data” implies an age ~ 9.5 billion years, well older than the solar system.

3. Jupiter is somewhere between a fluid and a rigid body, so something between 0.012 AU and 0.006 AU (that latter being well within the Sun’s corona).

4. momenta (in $\text{kg m}^2/\text{s}$) are approximately

a) $1.9 \cdot 10^{41}$

- b) $4.2 \cdot 10^{38}$
- c) $1.94 \cdot 10^{43}$
- d) 7.9, 1.7, 2.5, all $\cdot 10^{42}$
- e) $2.7 \cdot 10^{40}$
- f) $1.7 \cdot 10^{36}$
- g) and the solar contribution accounts for ~0.6 percent