PX282 - Stars and the Solar System
Solar System Summary

L.J. Aitchison

May 16, 2023

Contents

1 Introduction

This is a summary and revision document for the Solar System section of the second year physics module PX282: Stars and the Solar System. It attempts to cover the main points of the different sections of this part of the module and delve into key points and the formation of the solar system, the various bodies within it and the mechanisms that occur between them.

Disclaimer: No gurantee can be made that this document is complete and withour error. Please contact me with any suggestions or corrections.

1.1 Fundamental Properties of Bodies

PX282: Stars and the Solar System

The fundamental properties of Solar System bodies

Component	Orbital separation a au	Orbital period $P_{\rm orb}$ yr	Mass M $M_{\rm E}$	Radius R $R_{\rm E}$	Density ρ $10^3\ {\rm kg\,m^{-3}}$	Spin period $P_{\rm spin}$ d	Day Length d	Spin obliquity to orbit degrees	Surface magnetic field B μT	Orbital eccentricity e	Orbital inclination[a] i degrees
Sun	∼0.005		333000	**109**	1.4	25.1					
Mercury	0.39	0.24	0.055	0.38	5.4	58.8	175.9	0.03	0.5	0.206	6.3
Venus	0.72	0.62	0.82	0.95	5.2	–244	116.8	177.4	∼ 0	0.007	2.2
Earth	1	1	1	**1**	5.5	0.997	1	23.4	25–65	0.017	1.6
Moon			0.012	0.27	3.3	27.3			0.01		
Mars	1.52	1.88	0.11	0.53	3.9	1.03	1.03	25.2	∼ 0	0.093	1.7
Ceres	2.8	4.6	0.00015	0.074	2.2	0.38	0.38	4	∼ 0	0.076	9.2
Jupiter	5.2	11.9	318	**11.0**	1.3	0.415	0.414	3.1	400–1300	0.049	0.3
Saturn	9.6	29	95	9.1	0.7	0.445	0.444	1.1	18–84	0.057	0.9
Uranus	19.2	84	14.5	4.0	1.3	–0.72	0.72	97.8	10–100	0.046	1.0
Neptune	30	165	17.2	3.9	1.6	0.673	0.671	28.3	10–90	0.011	0.7
Pluto	39	248	0.002	0.19	1.9	–6.40	6.39	122.5	?	0.248	15.6

[a] Inclination of orbit with respect to the *invariable plane*, representing the total angular momentum of the Solar System (which is dominated by the orbit of Jupiter).

Figure 1: Fundamental Properties of bodies in the Solar System

Definition of a Planet The IAU (International Astronomical Union) definition of a planet states that -
A planet is a celestial body that:

(i) Is in orbit around the Sun

(ii) Has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium shape

(iii) Has cleared the neighbourhood around its orbit

A planet must meet all three criteria. A **dwarf planet** is an object that fulfils only **a and b**. This distinction was included in order to account for the numerous smaller bodies that were being detected in the Kuiper belt. Other

objects orbiting the Sun which fulfil only the first criteria, (**a**), are known as **small Solar System bodies.**

There are certain issues with this definition. For example, criterion (**c**) attempts to distinguish between planets and dwarf planets, but borderline cases are clearly possible as the completeness of clearing is not specified (see Near Earth Objects for example.) Furthermore, the definition does not include **exoplanets** and also fails to produce and upper limit on the size or mass of an object that can be considered a planet.

1.2 Hydrostatic Equilibrium

This concept comes up throughout the course and is also included with in the previous definition of a planet. Hence, it is important to understand how to derive the expression and where it comes from. Gravity is balanced by pressure forces. Consider a cylinder within the planet. The force of gravity is given by

$$F_G = \frac{GM_r dm}{r^2}$$

where M_r is the mass of the planet enclosed within r and dm is the mass of the cylinder.
We also know that the pressure forces are given by

$$F_P = P(r)dA - P(r+dr)dA = -dPdA$$

Since $dm = \rho(r)drdA$, by balancing these two forces we get the hydrostatic equilibrium equation:

$$\frac{dP}{dr} = -\frac{GM_r\rho(r)}{r^2}$$

From this we can solve for $\rho(r)$ which is the interior structure of the planet. However, in order to do this we need to know the formula for the equation of state, $P(\rho)$.

2 Planetary Motion

2.1 Geocentric vs. Heliocentric

Up until the mid-1500s, the accepted view of the universe was a model that considered the Earth at the centre of with the Sun, Moon, Planets and stars all moving around in circles around it. This was the **geocentric** model. However, there were noticeable issues with this. For example, the retrograde motion of Mars was hard to explain. The Ptolemaic Model added **epicycles**, which is circular motion around average circular motion. This explains the retrograde motion, but the models became increasingly complex when trying to remain consistent with improving observations.

In 1543, Copernicus proposed a **heliocentric** model. This provided a simpler and more elegant explaination for retrograde motion, saying it would naturally occur as Earth overtook the other planets. Furthermore, this model gave the correct order of planets from the sun. However, this model was highly controversial and still required epicycles due to the assumption of perfectly circular orbits.

2.2 Kepler's Laws

Tycho Brache, a Danish astronomer, made precise measurements of planetary motion. In 1609, Johannes Kelper published his three laws that reproduce Brache's observations without epicycles.

Kepler's First and Second Laws The first two of Kepler's Laws are as follows:

(i) A planet orbits the Sun in an ellipse, with the sun at one focus of the ellipse.

(ii) A line connecting a planet to the Sun sweeps out equal areas in equal time intervals.

Let us discuss elliptical orbits in a bit more detail.

Elliptical Orbits Observe the following diagram

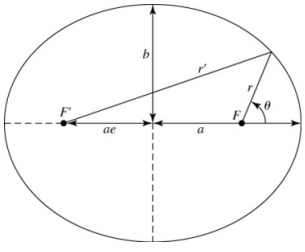


Figure 2: Elliptical Orbit

An ellipse is defined by the set of points $r + r' = 2a$ where a is a constant. On the diagram, a is the semi-major axis, b is the semi-minor axis and e is the

eccentricity (or ellipticity). Consider the point Q at the bottom of the ellipse:

$$\begin{aligned} r^2 &= b^2 + a^2e^2 \text{ with } r' = r = a \\ &\implies a^2 = b^2 + a^2e^2 \\ &\implies b^2 = a^2(1 - e^2) \end{aligned}$$

Considering the different eccentricities (all of which are permissable orbits):

(i) $e = 0 \implies$ Circle

(ii) $0 < e < 1 \implies$ Ellipse

(iii) $e = 1 \implies$ Parabola

(iv) $1 < e \implies$ Hyperbola

Kepler's Third Law Kepler's Third Law is as follows:

$$P^2 = ka^3$$

where P is the orbital period, a is the semi-major axis of an elliptical orbit and k is a constant. When the central mass dominates (i.e. a one body problem), $k = \frac{4\pi^2}{GM}$.
In Earth units, we get the simple formula:

$$P^2(\text{yr}) = a^3(\text{au})$$

which allows us to easily calculate periods or seperations for other bodies in the Solar System.
In 1687, Isaac Newton used Kepler's Laws to infer his law of universal gravitation. Here is a simplified derivation:

$$P^2 = ka^3$$

and, assuming a circular orbit,

$$P = \frac{2\pi r}{v}$$

Hence,

$$(\frac{2\pi r}{v})^2 = kr^3 \implies \frac{4\pi^2 r^2}{v^2} = kr^3$$

Therefore, we have

$$\frac{v^2}{r} = \frac{4\pi^2}{kr^2}$$

Now, using Newton's Second and Third Laws, we can obtain the result.

$$\begin{aligned} F = ma = m\frac{v^2}{r} &= \frac{4\pi^2 m}{kr^2} && \text{(NII)} \\ &= \frac{4\pi^2 M}{k'r^2} && \text{(NIII)} \\ &= \frac{4\pi^2 Mm}{k''r^2} \\ &= \frac{GMm}{r^2} \end{aligned}$$

where

$$G = \frac{4\pi^2}{k''} \text{ and } k'' = k'M = kM.$$

2.3 Two Body Problem

For the one-body problem, the mass of one object domiantes. The most massive object is considered the origin and the equation of motion is given by:

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{GM_\odot}{|\mathbf{r}|^2}\hat{\mathbf{r}}$$

where the left hand side is the acceleration of the planet (or object in orbit of the larger mass) and the right hand side is the acceleration due to gravity.
For the two-body problem, the mass of both objects is relevant. We place the origin at the centre of mass.

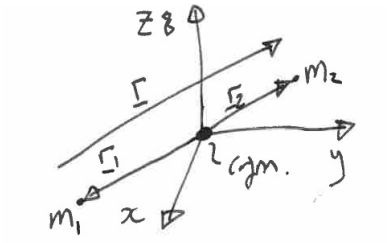


Figure 3: Diagram for the two body problem

Using the notation from the diagram, we can see that by definition:

$$\frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} = 0 \text{ and } \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1 \implies \mathbf{r}_1 = -\frac{m_2}{m_1 + m_2}\mathbf{r}, \quad \mathbf{r}_2 = -\frac{m_1}{m_1 + m_2}\mathbf{r}$$

This demonstrates that the two objects must have the same shape of orbit (as they are both propotional to $\mathbf{r}$), but with different amplitudes. Furthermore,

$$\begin{aligned} & \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_1 \\ \implies & \frac{d^2\mathbf{r}}{dt^2} = \frac{d^2\mathbf{r}_2}{dt^2} - \frac{d^2\mathbf{r}_1}{dt^2} \\ & = -\frac{Gm_2}{|\mathbf{r}|^2}\hat{\mathbf{r}} - \frac{Gm_1}{|\mathbf{r}|^2}\hat{\mathbf{r}} \\ & = \frac{G(m_1 + m_2)}{|\mathbf{r}|^2}\hat{\mathbf{r}} \end{aligned}$$

which is mathematically identical to the one body system.
And as an add-on, a general version of Kepler's Third Law is:

$$P^2 = \frac{4\pi^2}{G(m_1 + m_2)}(a_1 + a_2)^3$$

where a_1 and a_2 represent the semi-major axis of their respective objects.

2.4 Energy in Orbits

Here we will divert into a quick discussion about energy in orbits. Consider a planet orbiting a star. As the planet moves close to the star, it has a high velocity corresponding to a high kinetic energy. Conversely, when the planet is far away from the star, it has a comparatively low velocity and therefore a lower kinetic energy. The energy in the system oscillates between kinetic energy and potential energy.
Consider a circular orbit:

$$P^2 = (\frac{2\pi a}{v})^2 = \frac{4\pi^2 a^2}{v^2} = \frac{4\pi^2}{GM_\odot}a^3$$
$$\implies v^2 = \frac{GM_\odot}{a}$$
$$\implies E_K = \frac{m_p v^2}{2} = \frac{GM_\odot m_p}{a^2} = \frac{U}{2}$$

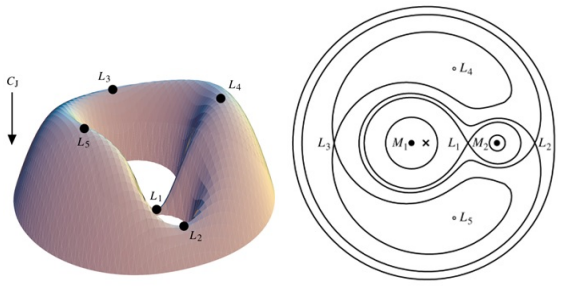
where U is the gravitational potential energy. A similar derivation can show that this applies to elliptical orbits as well. Therefore, we know from this that

$$\langle U \rangle = 2\langle E_K \rangle$$

Remarkably, this is true for **all** gravitationally bound systems in equilibrium (see the Virial Theorem for more information).

2.5 Three Body Problem

The three body problem famously has no analytic solution. Numerical calculations demonstrate that different solutions are found for small changes in initial conditions. Typically there is a complex interaction followed by one object (usually the lowest mass) being ejected and the remaining two forming a tight binary. However, analytic solutions are possible in certain special cases.
In the restricted three body problem, the third mass is considered to be negligible (e.g. The Sun, Jupiter and an asteroid). The two following figures demonstrate the **Roche geometry** of the gravitational potential and in the frame rotating with the two masses:



(a) 3D View (b) Top down view

Figure 4: Roche geometry

In (b) of (4) the different sections of the diagram correspond to contours of constant potentials. One can note the **Lagrangian points** on the diagram

where objects can have stable orbits with regard to other two larger masses. It is important to consider how big a region around the secondary object one can have where a stable orbit is preserved i.e. where the gravity of that object is going to dominate. This leads us to concept of the Hill sphere.

2.6 Hill Sphere and Derivation

The approximate sphere of influence around the second body in the (restricted) three-body problem is known as the **Hill sphere**. We can derive an approximate expression for the radius of the Hill sphere[1] by considering the position of the $L1$ point in the above diagram. Firstly, by balancing the following expressions

$$\underbrace{\frac{Gm_2}{R_H^2} - \frac{Gm_1}{(a-R_H)^2}}_{\text{Gravitational acceleration}} + \underbrace{\Omega^2(a-R_H)}_{\text{Centrifugal acceleration}} = 0$$

Using Kepler's third law, we can equate the angular frequency in the following way

$$\Omega^2 = \left(\frac{2\pi}{P}\right)^2 = \frac{Gm_1}{a^3}$$

Now we simplify the denominator as such

$$\begin{aligned}(a-R_H)^{-2} &= a^{-2}\left(1-\frac{R_H}{a}\right)^{-2} \\ &\approx a^{-2}\left(1+2\frac{R_H}{a}\right)\end{aligned}$$

where the Binomial expansion is used to appoximate in the second step. Now, we assume $m_1 \gg m_2$ and substituting the expressions above in

$$\begin{aligned}&\frac{Gm_2}{R_H^2} - \frac{Gm_1}{a^2}\left(1+2\frac{R_H}{a}\right) + \frac{Gm_1}{a^2}\left(1-\frac{R_H}{a}\right) \approx 0 \\ &\implies \frac{m_2}{R_H^2} - \frac{m_1}{a^2}\left(3\frac{R_H}{a}\right) \approx 0 \\ &\implies R_H \approx a\left(\frac{m_2}{3m_1}\right)^{\frac{1}{3}}\end{aligned}$$

and hence we have an expression for the *sphere of influence* around the smaller body in a two body system.

2.7 Orbital Perturbations and Resonances

There is an alternative case to the three-body problem and this is where the primary object in the system has a mass that dominates but the masses of the other objects are not negligible. One should note that this is really applicable to all of the planets as while the Sun dominates, the planets do not have zero mass so they influence one another. An interesting demonstration of this is the

[1] Note that actual area of domination around the body would most likely not be spherical. Furthermore, an object would have to be well in the area to maintain a stable orbit.

discovery of Neptune via perturbation of the orbit of Uranus.
An impertaive kind of orbital resonance are the **secular resonances** which arise from weak perturbations growing to significant net effects on timescales much longer than the orbital period. As the masses of the planets are much smaller in comparison to the main body (i.e. the Sun) this is modelled by *smearing* the mass of the planets around their orbits and then considering the gravitational interaction between the *rings of mass* instead.
Secular resonances are long-term variations of eccentricity and inclination of orbits due to relatively weak perturbations. These secular resonances are the cause of **Milankovitch cycles** of the Earth, which influence climate and ice ages.
There is another type of resonance known as **mean motion resonances**. These are resonances that occur for integer ratios of orbital periods. Usually causes perturbations to grow rapidly, driving orbits unstable. However, for some systems like the Galilean moons of Jupiter, these resonances are stabalized by tidal interactions.

2.8 Tides

In reality, objects within the Solar System are not point masses. **Tidal forces** arise for non-point masses due to differential gravity across an object. This results in the object becoming slightly elongated, exhibiting **tidal bulges**. This is not to be confused with *water tides* which are larger amplitude local oscillations that are driven by gravitational forcing. If we differentiate the force of gravity given by

$$F_G = \frac{GMm}{r^2}$$

we find that the differential gravity across a body is given by

$$dF_G = -\frac{2GMm}{r^3}dr$$

demonstrating that tidal effects are more sensitive to orbital separation in comparison with gravitational force. Consequently, tidal effects are more important for close-in objects.
We now wish to consider the net effect on the system from these tidal bulges. If the spin of the two objects were synchronized with their orbit then the tidal bulges would arise and face one another. This would result in no tidal interaction and is the end state of such systems. However, it is generally true that tidal bulges are not perfectly aligned to differential gravity due to rotation and friction. Take for example the Earth and the Moon. The Earth spins faster than the orbit of the Moon, so the tidal bulge *leads* the Moon's orbit. Differential gravity on leading and trailing bulges lead to *net torque* that transfers angular momentum from rotations of Earth to the orbit of the Moon. This is shown in the (5) where since force on bulge A is greater than bulge B, a net torque is creater on the Earth which slows its rotation. Friction is dissipative and the rotational energy is lost to internal heating. This transference only ceases once spin is *synchronized* to the orbit (which has already occurred for the Moon). For an alternative arrangement where the planet is slowly rotating, a moon orbiting said planet will spiral in.
As mentioned before, for a circular orbit once spin is synchronized to orbit the

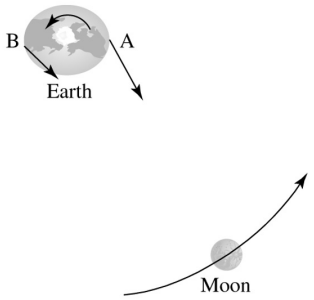


Figure 5: Tidal bulges

tidal interaction stops. However, for eccentric orbits (which will have variable velocity around the course of the orbit) the tidal bulges still lead and trail the rotation even *after* synchronization has ceased. As a result frictional dissipation and hence tidal heating continues. Energy is removed from the orbit but at a constant angular moment. Therfore, this continues until the orbit is *circularized* as circular orbits have lower energy than elliptic orbits. If eccentricity is mainted by gravitational perturbations (mean motion resonances for example) then the heating can continue indefinitely (this is what happens on the Galilean moons of Jupiter).

3 The Sun

3.1 Fundamental Properties

The Sun is a *mid-sized, middle-aged main sequence* (G2V) star. It accounts for 99.8% of mass in the solar system. The key properties are as follows

(i) Mass: $M_\odot = 2 \times 10^{30}\,\text{kg}$

(ii) Radius: $R_\odot = 7 \times 10^8\,\text{m} \approx 110 R_E$

(iii) Temperature: $T_{\text{eff}} = 5780\,\text{K}$

Kepler's third law can be used to derive the mass. To deduce the radius, we measure its angular size on the sky. This means we measure how much of the sky is covered by the sun and then by using trigonometry and the distance to the sun (which can be deduced by the length scale of the solar system) we can ascertain the radius. For the temperature, we need to analyse the spectra of the sun. From this diagram, one can see that the spectrum of the Sun is close to

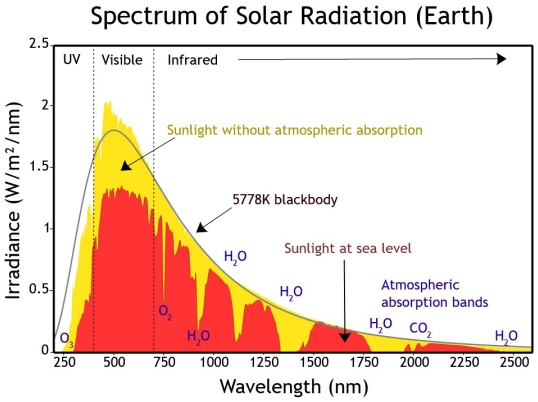


Figure 6: Solar Spectrum

that of a black body (denoted by the black line). Superimposed on this spectra are the dark absorption lines corresponding to atomic transitions in the Solar atmosphere. Furthermore, the red section corresponds to the spectrum of the Sun observed at the surface of Earth. The rough overall shape, but the *chunks* taken out of it correspond to absorption by molecules in the Earth's atmosphere. Moreover, the whole spectrum at the surface is surpressed into the blue which correspond to Rayleigh scattering in the Earth's atmosphere.
Another important property of the Sun is its **luminosity**. This can be calculated in the following way

$$L_\odot = A\sigma T_\odot^4 = 4\pi R^2 \sigma T_\odot^4 = 3.84 \times 10^{26}\,\text{W}$$

where σ corresponds to the Stefan-Boltzmann constant. A natural question to ask is how does the Sun maintain this luminosity? If we consider the Sun as a sphere with a shell of radius dr around it, we know from the potential energy of that shell is related as follows

$$dU = -\frac{GM_r dm}{r}$$

where, like before, M_r corresponds to the mass enclosed within the shell. Note that this result is negative because we are considering the *gravitational binding energy* or the energy that has been lost by the material as it collapsed to form the Sun. dm is the mass of the shell which is given as

$$dm = 4\pi r^2 \rho(r) dr$$

To do this properly, one would have to calculate the density for different parts of the Sun. This is not really necessary for our purposes, so we assume the Sun to have a uniform density. This allows us to easily calculate M_r as

$$M_r = \frac{4}{3}\pi r^3 \rho_\odot$$

where $\rho_\odot$ corresponds to the *meal Solar density*. Substituting this in and integrating we find the following

$$\begin{aligned} U &\approx -\frac{16}{3}\pi^2 G\rho_\odot \int_0^{R_\odot} r^4 dr \\ &= -\frac{16}{15}\pi^2 G\rho_\odot R_\odot^5 \\ &= \frac{3GM_\odot^2}{5R_\odot} \\ &= -2 \times 10^{41}\,\mathrm{J} \end{aligned}$$

All of this energy has to be have been released. If we consider a **thermal timescale** (also known as the **Kelvin-Helmholtz timescale**) this would simply be the ratio of the gravitaional potential energy to the luminosity as follows

$$-\frac{U_\odot}{L_\odot} \approx 1 \times 10^7 \text{ years}$$

Obviously if one were to do this properly one would have to consider the change in luminosity over time. However, for an order of magnitude calculation this is fine. Nevertheless, this estimate does not reach the age of the Sun we know ($\sim 4.6 \times 10^9$ years) from radionucleotide dating of meteorites that fall to Earth. Hence, another means must be at hand. It is now understood that the Sun is powered by *nuclear fusion*. The burning of hydrogen into helium is primarily what powers the sun and it is through a chain of reactionst that we get from one to the next. This end product is

$$4\,{}^1_1\mathrm{H} \longrightarrow {}^4_2\mathrm{He} + 2\,\mathrm{e}^+ + 2\,\nu_e + 2\,\gamma$$

where e^+ is a positon, ν_e is an electron neutrino and γ is a photon[2]. The *binding energy* of helium is 0.7% of the rest mass energy of the hydrogen nuclei. This leads us to a **nuclear timescale** as follows

$$\frac{0.0007 M_\odot c^2}{L_\odot} \approx 1 \times 10^{11} \text{ years}$$

which is more than enough to power the Sun for its known age.

[2]There is more detail on this in the Stars section of the course. See *proton-proton chain*.

3.2 Internal Structure

The interior of Stars is discussed in further detail in the Stars section of the course but is important here as the interior of the Sun influences the surface which in turn influences the planets. The first indication of the mechanisms of the Sun's inner structure come from analysing the surface. Solar granulation and sunspots demonstrate that the Sun is **convecting**. If one considers a parcel

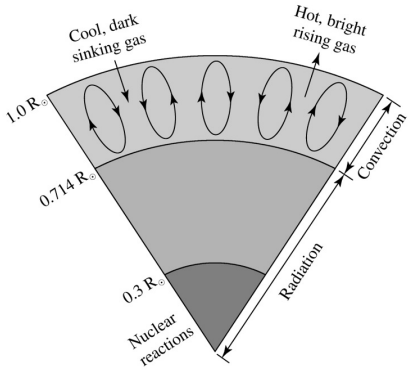


Figure 7: Interior structure of the Sun

of gas inside the sun which is hotter than the surrounding gas, then it will be buoyant and begin to float. As it floats, it enters a lower pressure area and hence begins to expand quickly and adiabatically. From this there are two possbilities. These are either that the temperature gradient is shallow meaning the gas loses buyoancy, stalls and falls back down. The other possibility is that the temperature gradient is steep and so the gas parcel remains buyoant and continues to rise. The region in which this can occur extends about 30% of the way into the Sun as demonstrated in (7). Beyond this, the Sun is more stable to convection and energy transport is **radiative**.
One should note that on the above diagram that the point at which energy transport moves from a radiative method to a convective method is quite well defined. This is ascertained through the use of **helioseismology**.

3.3 Sunspots and Magnetic Field

Sunspots are relatively dark regions on the surface of the Sun where the temperature of the photosphere is *much lower* $(3000-4500\,\mathrm{K})$ than the surrounding region. This, as well as the distorted granulation around a sunspot, indicates that convection is suppressed in these regions. A natural question is that if these regions are cooler then what is supporting them from collapsing?
Optical spectra taken across a sunspot reveal that **Zeeman splitting** of atomic absorption lines is occurring due to a magnetic field. The magnetic field lifts the degeneracy of atomic energy levels with different angular momentum states, splitting the individual lines of a spectrograph into multiple components. This allows us to locate magnetic fields from a far distance and hence demonstrates that sunspots have a strong magnetic field around them not present in surrounding regions. This answers two questions. The first is that the reason convection

is being surpressed is that the plasma is being confined by the magnetic field (due to the Lorentz force) and the second is that sunspots do not collapse as they are still in hydrostatic equilibrium with their surroundings due to thermal pressure being partly replaced by **magnetic pressure**.
If we spatially resole this Zeeman splitting, we get a real time view of the magnetic field of the Sun. From this we see that there is an opposite polarity on either side of the sunspot groups as well as a reversal of polarity either side of the equator. This further demonstrates that sunspot groups are locations of magnetic flux loops that are emerging through the surface. Hence, the Sun has a large scale dipole field together with higher order flux loops associated with sunspot groups. This field is thought to actively generated by a large scale magnetic dynamo, which requires a conducting fluid with differential rotation and convection as well as a small magnetic field to begin with. The general

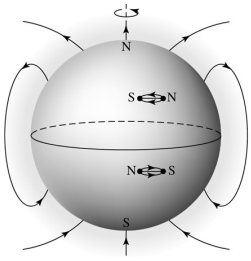


Figure 8: Large scale dipole field for the Sun

idea of how this dynamo works is that the weak initial field generates currents within the conducting medium, which is explaining strong and complex kinetic energy flows. This generates additional currents which in turn then enhance the magnetic field. There is a feedback loop between the magnetic field and the kinetic energy in the Sun which leads to an amplification of the strength of the magnetic field with the kinetic energy being converted into magnetic energy.
If we map the rotation of the Sun using sunspots we find that different latitudes rotate at different rates, with the equator rotating more rapidly than the poles. This is called **differential rotation**. Helioseismology demonstrates that this effect extends throughout the convective zone but that the radiative zones rotates as a solid body. The strong shearing flows at the tachocline (the boundary

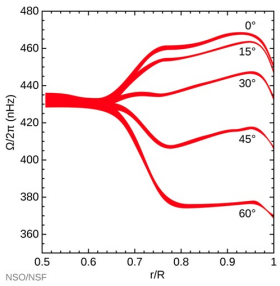


Figure 9: Graph showing change in differential rotation

between these zones) where plasma very close to itself that is moving at very

different frequencies, is thought to be the differential kinetic energy that drives the Solar dynamo.[3]

3.4 Solar Atmospheres

Before we discuss the upper atmosphere, we ask the natural question of what defines the quote-on-quote visible surface of the Sun? This is usualyl considered to be the **photosphere** which is where the solar atmosphere transitions from opaque (or optically thick) to transparent (or optically thin). If we consider the inside the Sun, there are many photons randomly scattering off atoms. It takes millions of years for a photon generated in the middle of the Sun to escape, but one can see that the distance these photons travel on average is going to increase as the density decreases the further you move from the core. Eventually, a scattering event will cause the photon to escape completely (which are the ones we see). This scattering is random as stated, but there is a probability of the scattering such that we can define a **mean free path**, which is how far on average a photon will travel before it scatters. This mean free path is given by

$$l = \frac{1}{\kappa\rho(r)}$$

We know the density decreases as we move out from the Sun. The constant κ is the **opacity**, which is how likely it is for a photon to be blocked by an electron or atom. This depends on composition as well as the wavelength. Therefore, we can see that photosphere is going to be the radius at which the mean free path becomes longer than the scale of the atmosphere. Within the photosphere, the mean free path is shorter than the scale so another scattering event will always occur. But as the density drops off, at some point the mean free path becomes longer than the scale of the atmosphere and therefore the photon is more likely to escape to infinity rather than undergo another scattering event.
As a bit of an aside, this matters for objects other than the Sun, so in general we consider a beam of light with intensity I. The change in intensity depeneds on the opacity of the material, the density of the material as well as the intensity of the beam at that point. Hence, we get the expression

$$dI = -\kappa\rho I dr$$

Integrating this gives

$$\int \frac{1}{I} dI = -\int \kappa\rho(r) dr = \tau$$

where τ is the **optical depth**. Peforming the integration and exponentiating gives the expression

$$I = I_0 e^{-\tau}$$

which shows that as we go through a medium, the intensity is going to go from an initial intensity I_0 to a new intensity I and will depend (go exponentially down) on the optical depth of the medium.
Returning to the Sun, in practice this means that we can see into the atmosphere to about one mean free path which is approximately equal to $\tau \approx 1$. Hence

$$\frac{I}{I_0} \approx \frac{1}{e} \approx 0.37$$

[3]Note that magnetic fields are also seen for fully convective stars.

So where $\tau < 1$ we can call the medium optically thin and where $\tau > 1$ we can the medium optically thick. This means there is an optical depth of about 1 between the observer and the photosphere.
The upper atmosphere of the Sun, divided into sections called the chromosphere and the corona, are heated to much higher temperatures than the photosphere, with temperatures in the corona exceeding 1×10^6 K. This results in a temperature inversion as shown in (10). There is obviously some kind of heating

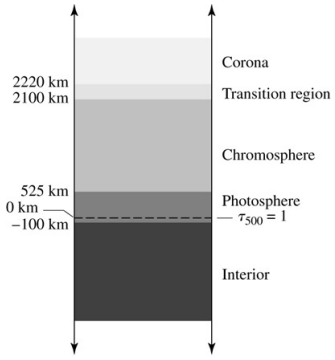


Figure 10: Temperature gradient of the Sun

mechanism, but the heating of this upper atmosphere is still an open question, but it is thought to be a combination sound waves, magnetic reconnection events (see below) as well as the influence of magnetic Alfven waves.

3.5 Solar Wind and Solar Flares

So far we have been consider the large scale magnetic field of the Sun to be a dipole, but in reality it has many more open field lines. This is a result of the competition between the magnetic field attempting to confine the heated corona and its thermal pressure. Coronal plasma can escape along these open field lines in a **solar wind**. This wind is also additionally accelerated by magnetic Alfven waves. Charges particles are driven out of the Sun at hundreds of kilometres per second. The interaction between this wind and the Solar magnetic field at large seperations provide a magnetic braking torque that is responsible for slowing the spin of the Sun. The complex magnetic field loops in the Solar corona can reconfigure suddenly to a simpler morphology. This is a process known as **magnetic reconnection**. Here, magnetic energy is converted to kinetic energy via particle acceleration and then to thermal energy via particle interactions. This results in an intense emission of X-rays and gamma rays on the scale of minutes. These are what we call **solar flares**. The most energetic of these reconnection events can cause coronal loops to burst open, ejecting hot plasma from the corona. These events are known as coronal mass ejections.

4 Terrestrial Planets

4.1 Spectrum of a Planet

We can discern a lot about a planet and its atmosphere by looking at its **spectra**. This split into separate components of **reflected sunlight** and **thermal emission** (absorbed and re-emitted Solar energy). Considering the reflected component first, we know that

$$L_{\text{reflected}} = \underbrace{\frac{L_\odot}{4\pi a^2}}_{\text{Solar flux at planet}} \cdot \underbrace{A_G}_{\text{Geometric albedo}} \cdot \underbrace{\pi R_p^2}_{\text{Cross-sectional area}}$$

The **geometric albedo** A_G describes what fraction of the light is reflected by a body (this will depend on the viewing angle). The spectrum of this reflected sunlight is very similar to the spectrum for the Sun as it is the same light but reflected.
Now, we consider the thermal component by equating heating and cooling[4]. First by considering the heat in we get

$$\begin{aligned} L_{\text{in}} &= \frac{L_\odot}{4\pi a^2}\pi R_p^2(1-A_B) \\ &= \frac{4\pi R_\odot^2 \sigma T_\odot^4}{4\pi a^2}\pi R_p^2(1-A_B) \\ &= \frac{R_\odot^2 \sigma T_\odot^4}{a^2}\pi R_p^2(1-A_B) \end{aligned}$$

where A_B corresponds to the **bond albedo** which is angle averaged. For the heat out we have the expression

$$L_{\text{out}} = 4\pi R_p^2 \sigma T_p^2$$

where in the first expression we have used the total surface area. This means we are assuming *uniform radiation* and efficient heat distribution from the day side to the night side (i.e. rapidly rotating planets or atmospheric thermal lag). Now, if we equate these expressions and rearrange we get the following expression

$$T_p = (1-A_B)^{\frac{1}{4}} \left(\frac{R_\odot}{2a}\right)^{\frac{1}{2}} T_\odot$$

which is the **equilibrium temperature**. Note that this expression does not depend on the size of the object.

4.2 Interiors

4.2.1 The Earth

Most of what we know about other planetary bodies in the Solar Systems comes from leveraging what we know at about the Earth. Hence, it is important to discuss in depth the structure and mechanisms of the Earth's interior. Earth is differentiated into distinct layers of different density which implies that the

[4] In this we assume no significant source of internal heating.

Earth was once molten and the heavier metals sank. The different layers of the Earth are the crust which is a wide range of silicate minerals, the mantle which is mainly a magnesium-iron silicate, the outer core which is a liquid comprised of iron and nickel and the inner core which is a solid iron and nickel object. These layers can be identified through seismology. Utilising **p-waves**, which

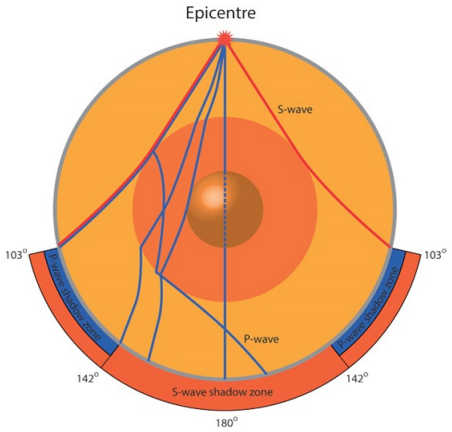


Figure 11: Diagram of movement of p-waves and s-waves

are compressional waves that propogate in solids and liquids, and **s-waves**, which are transverse and propogate only in solids, we can map out the interior of the Earth by analysing the *shadow zones* that result from the reflection, refraction and blocking of the waves. The liquid core reflects and refracts the p-waves but cannot transmit s-waves, whilst the solid inner core transmits s-waves excited by p-waves. These inner core s-waves are only detected indirectly via perturbations on p-waves.
The bulk Earth is composed primarily of only four elements, those being iron, oxygen, silicon and magnesium. It is strongly deficient in **volatile** elements with low melting or boiling points such as hydrogen, helium, carbon and nitrogen. This may seem quite surprising, as hydrogen and helium are by far the most

Element	Abundance (per 100,000)	Common molecules
H	92,700	H_2
(He)	7,200	
O	50	H_2O, silicates
(Ne)	20	
N	15	NH_3
C	8	CH_4
Si	2	Mg_2SiO_4
Mg	2	
Fe	1	Fe_3O_4

Figure 12: Cosmic abundance of elements (by number)

common elements in the universe as (12) shows. However, both of these are vulnerable to evaporation. Oxygen is common on Earth but only because it binds readily to form refactory compounds which have high melting and boiling points (such as silicates). The dominance of these refactory elements in the bulk composition shows that Earth formed in a warm environment where ices were

not present.
We now wish to consider the heating of the interior of the Earth. The fact that the inner core is liquid demonstrates that the interior is very hot. However, when the the measured temperature gradient of the Earth only gives a range of ages of the Earth of about $20 - 400$ million years. For the interior to remain hot after the accepted age of 4.6 billion years, an ongoing heat source must be required. Tidal heating is not sufficient for this and it is not understood that this heating comes from radioactive decay. One importance consequence of this heating is that Earth has a strong magnetic field which is believed to originate in a magnetic dynamo in the molten iron core.
This temperature gradient also drives convection in the mantle. The crust and the upper mantle, which together form the **lithosphere**, fractures into plates from the movement of convection cells in the more plastic deeper mantle, called the **asthenosphere**, and moves on top of these convection cells. This

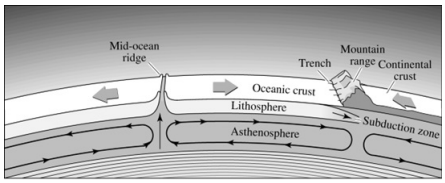


Figure 13: Movement of lithosphere

movement of plates results in new crust forming at **mid-ocean ridges** as these plates move apart as well as the **subduction** of ocean crust and the building of continents where the plates collide. Interestingly, the new crust formed at these mid-ocean ridges holds a fossil record of the magnetic polarity of Earth. This record demonstrates that the polarity of the magnetic field reverses on a half a million year timescale and provides strong evidence that the magnetic field is generated by a dynamo process. Plate tectonics also create chains of volcanic islands by moving crust across mantle plumes.

4.2.2 The Other Terrestrial Bodies

We expect the other massive rocky planets to be differentiated as all will be hot at formation and therefore molten. These planets are also heated by radioactive decay (with Mercury also being heated by tidal heating from the Sun). However, smaller bodies cool more efficiently due to the ratio of surface area to volume

$$\tau_{\text{cool}} \propto \frac{4\pi R^2}{\frac{4}{3}\pi R^3} \propto \frac{1}{R}$$

This means that smaller bodies are expected to form a thick lithosphere that does not fracture into plates and is therefore geologically inactive.
The Moon is the only other body with a seismic record (as a result of the placement of lunar seismometers on the Moon by the Apollo 11 astronauts) and so we know **moonquakes** occur. The shallow quakes are as a result of tidal strain and impacts of meteorites. The deeper quakes imply the existence of a thicker lithosphere on top of a plastic asthenosphere. The distinctive shadow of S-waves shows that there is a small liquid core, but the lack of a global magnetic

field shows that a dynamo does not operate. Nevertheless, rock samples suggest that a dynamo did operate on the young moon.
Seismology has been performed on Mars, starting from 2019, and demonstrates that a liquid core exists which is larger than expected. This suggests a low density of the planet due to the presence of sulphur and other light elements. The speed of the S-waves suggest a thick lithosphere as predicted. Moreover, the lack of a global magnetic field shows a dynamo is no longer operating but field frozen into rocks suggest that it did in the past.
We have less information about Venus but we know that it should have an interior very similar to Earth. The continuing geological activity shows that there is a very hot interior. However, no plate tectonics are seen (potentially due to the lack of liquid water) and there is no global magnetic field. Seismology is required for further information to be deduced.
Mercury is smaller than Mars and not geologically active. This implies that it should have a thick lithosphere. However, Mercury *does* exhibit a global magnetic field which suggests that a magnetic dynamo may be operating. This may be as a result of its anomalously high density which seems to imply an over-sized iron core combined with significant heating from Solar tides.
Certain properties of both the Moon and Mercury are thought to be as a result giant collisions in the final stage of the assembly of the terrestrial planets. The over-sized core of Mercury may be a result of it losing most of its mantle in a large scale collision. The Moon similarly is thought to have formed from the mantle of Earth following a collision with a Mars-sized planet (this explains the similarity between the Moon and the mantle of Earth).

4.3 Surfaces

With the exception of the Earth, the surface of the Moon is the one that has been studied in the most detail. The near-side surface is heavily cratered with impact craters and has darker, less-cratered **mare** which are giant impact basins that were resurfaced by basalt lava flows until about three billion years ago. The far-side is quite different, with much less coverage by mare. Here, the crust is thicker and less prone to penetration by impacts (most likely due to an uneven lithosphere). The mare face the Earth as the lava material that flooded the basins is denser than the crust material, so there was more mass in those mare. This means that as the Moon tidally synchronized with the Earth it entered an equilibrium where the heavier side was pointing towards Earth. Studying the surface of the Moon allows the history of impacts to be reconstructed. For example, overlapping impact craters must have occurred later and ghost craters must have been later flooded with lava.
The lunar surface is covered with a fine **regolith** that has been pulverised by large impacts as well as micro-meteorites.

4.3.1 Dating Samples

We date samples using radionucleotide dating. As the name implies, this makes use of radioactive decay which usually occurs in chains until a stable daughter product is reached. However, the half-life is usually dominated by a single rate-limiting step. We use this half-life to date the samples. Recall for radioactivity

that

$$N_{A,f} = N_{A,i}e^{-\lambda t} \quad \text{where } \lambda = \frac{\ln(2)}{\tau_{1/2}}$$

where $N_{A,f}$ is the final number of isotope A and $N_{A,i}$ is the initial number. Furthermore, $\tau_{1/2}$ is the half-life. If we then consider the decay to a daughter product we get the following

$$\begin{aligned} N_{A,f} + B_{B,f} &= N_{A,i} + N_{B,i} \\ &= N_{A,f}e^{\lambda t} + N_{B,i} \\ &\implies N_{B,f} = (e^{\lambda t} - 1)N_{A,f} + N_{B,i} \end{aligned}$$

In lucky cases (for example, lead in zircon crystal lattices), mineral samples can be assumed to form with negligible content of the daughter product and hence in the above expression $N_{B,i} = 0$. But, usually in practice we need to normalise with respect to an isotope of the same element with constant abundance in the sample, B'. Therefore we get the new expression

$$\frac{N_{B,f}}{N_{B'}} = (e^{\lambda t} - 1)\frac{N_{A,f}}{N_{B'}} + \frac{N_{B,i}}{N_{B'}}$$

where the final term is constant in the sample. This leads to a straight line relation and so given the graph one can read off the variables. To do this we need to make to assumptions, one of which is when the sample formed it was not biased towards which isotope it picked up. The second is that there is range of the daughter products and parent products (to make sure the x-axis on the straight line graph has a spread).

We now return to comparing the surfaces. The surface of Mercury is very ancient, around four billion years, which is shown from the heavy cratering of the surface. The frequency of impact craters on Mars shows that its surface is also very ancient. However, some resurfacing continued up until about two billion years ago. The presence of rift valleys and extinct volcanoes show past geological activity. Venus shows more impact craters but all are above three kilometres as the thick atmosphere shields the surface from smaller objects. The frequency of these impact craters shows that the entire planet was resurfaced about a billion years ago. As we have already discussed, large-scale lava flow and multitude of volcanoes demonstrates ongoing geological activity but there is no evidence of plate tectonics. With regards to Earth, impact craters *are* seen but they are rare. This is consistent with constant geological activity and water erosion causing constant resurfacing.

Volcanoes on Mars and Venus tend to be much larger as they are formed above mantel plumes which stay in one place due to the lack of plate tectonics. On Venus, the lithosphere is too thin to support the largest of these volcanoes resulting in sagging that creates large circular features called **coronae**.

The surface of Mars also shows evidence of abundant water flows in the past including river valleys, delta and sedimentary rocks.

4.4 Atmospheres

The atmospheres of Earth is heated significantly by the **greenhouse effect**, where optical light which is relatively transparent to the atmosphere is absorbed

by the planet and reradiated as a blackbody. The thermal emission of infrared wavelengths, of which the atmosphere is relatively opaque, results in the heat being trapped in the lower atmosphere which leads to higher surface temperatures. This opacity is dominated by water vapour (which is the strongest greenhouse gas) but the amount of water vapour responds to temperature. Carbon dioxide is the controlling species for this equilibrium as it determines the heat of the atmosphere as we will see.
There is no significant atmosphere on Mercury due to low surface gravities. Mars has an atmosphere mostly comprised of carbon dioxide, which may be a result of its heavier nature as well as its likely expulsion by volcanoes earlier in its history.
The atmospheric temperature profiles of Venus and the Earth are quite similar up until the higher pressures at the surface of each respective planet. To discuss this we need to consider the greenhouse effect more quantitatively. If one considers a single opaque layer in the atmosphere then the surface is heated twice by the reflection of light. This results in the expression

$$T_s^4 = 2T_p^4$$

by considering the luminosity relations. With N-optically thick layers, this becomes

$$T_s^4 = (1+N)T_p^4$$

and in terms of continuous, optical depth τ (recally that $I = I_0 e^{-\tau}$) we have

$$T_s^4 = (1+\frac{3}{4}\tau)T_p^4$$

With this, we return to our comparison. The climate of Earth is stabalised by the **carbon-silicate cycle** which is as such: carbon dioxide is removed from the atmosphere when it dissolves into water to become carbonic acid. This acid weathers silicate rocks, locking up carbon as calcium carbonate which is then deposited as ocean sediments (which is where most of Earth's carbon is). These ocean sediments are carried with the tectonic plates and eventually subducted beneath the continental plates at subduction zones. Volatile-induced melting then drives volcanism that returns the carbon back to the atmosphere, contributing to the greenhouse effect. Crucially, in this arrangement, there is a **negative feedback** loop as with higher temperatures, more rainfall occurs. This increases the draw down of carbon dioxide, reducing the greenhouse effect. The temperatures never become *too* low as the carbon dioxide is returned to the atmosphere by volcanic eruption.
However, there is a competing **positive feedback** loop since water vapour is a stronger greenhouse gas than carbon dioxide. Increasing temperatures result in more water being evaporated, increasing the water vapour of the atmosphere. This can lead to a **runaway greenhouse effect** wherein temperatures are continuously increased until all water is evaporated. Venus is thought to have undergone this process. This runaway greenhouse process requires water and strong Solar/geothermal heating. It is important to note that even if Earth had a thick carbon dioxide atmosphere, whilst it would still be too hot for life, it would not suffer a runaway greenhouse effect as the pressure of the atmosphere would keep water in liquid form. Ancient Earth also must have had a thicker carbon dioxide atmosphere to avoid becoming a snowball. Furthermore, note that a dry

Venus could not have undergone this process. The question remains as to where the water on Venus has gone. It is thought that the water vapour that rose was vulnerable to photolysis by ultraviolet light and so the hydrogen escaped from Venus due to X-ray heating of the particles and the solar wind sweeping it away (there are two means of this, **Jeans escape** and **hydrodynamic escape**). The evidence for this comes from the extreme deuterium to hydrogen ratio found on Venus, as deuterium is less vulnerable to atmospheric escape and so is preferentially held on to.
The escape rate of particles is energy limited, hence

$$\frac{L_X}{4\pi a^2}\pi R_p^2 \geqslant \frac{GM_p\dot{m}}{R_p}$$

where $\dot{m}$ is the mass loss rate. This rearranges to give

$$\dot{m} \leqslant \frac{L_X R_p^3}{4GM_p a^2}$$

Escape is also driven by Solar wind, the flux of kinetic energy is given by

$$f_{KE} = \frac{1}{2} n_H m_H v^3$$

4.5 Orbits and Rotation

We previously discussed secular resonances and how they affect eccentricity and obliquity. Orientation also varies due to N-body interactions, as seen in the advance of the perihelion of Mercury. Most of this advance is account for by N-body interactions, but about 43 arcseconds per century is not. It was thought this was due to an interior planet called Vulcan perturbing the orbit, but it was eventually realised that the advance was reproduced exactly by the theory of general relativity. Other tests of general relativity include the slight shift in the position of the stars when viewed close to the Sun during a solar eclipse as well as the time delay in radio signals to landers on Mars round superior conjunction.
It is important to discuss the difference in the **sidereal day** (the true rotation period with respect to the stars) and the **solar day** on Earth and Mars. This difference results from the fact that after one rotation they have advanced in their orbit, so need to rotate more to bring the Sun to the same altitude. The following expression gives a solar day

$$\frac{1}{P_{\text{Sol}}} = \frac{1}{P_{\text{Spin}}} - \frac{1}{P_{\text{Orbit}}} = \frac{1}{P_{\text{Sid}}} - \frac{1}{P_{\text{Orbit}}}$$

With regards to obliquity, the fact that Earth is rapidly rotating means it is slightly oblate. This means that the gravity of the Sun applies a slight torque driving precession of the axis of the Earth. The same things occurs for Mars. The weaker-time variable torques from the other planets drive variations in obliquity. This varies strongly for Mars on the timescale of millions of years, but not so much for Earth. This is due to the fact that the gravitational interaction of the Moon with the Earth stabalises its obliquity. The combination of obliquity and orbital variations define Milankovitch cycles, as stated in the earlier section, which influence long-term climate.

4.6 Moons

As we discussed previously, for the Earth, since rotation is faster than the orbit of the Moon the tidal bulges lead the Moon, thereby accelerating it. This increases the Moon's orbit and slows down the Earth. For Mars however, rotation is slower than the orbit of its moon Phobos. The tidal bulge therefore lags the moon in this case, decelerating Phobos and spinning up Mars. This leads to the in-spiral of Phobos.

Moons must be well within the Hill sphere of their parent planet to have a stable orbit, as discussed previously. However, if a moon approaches *too* cloes to its planet, it will overfill its own Hill sphere and be disrupted by tidal forces. This limit is known as the **Roche limit**. This limit can be approximated in the following manner. Recally that differential gravity is given by

$$dF_G = -\frac{2GM_m}{r^3}dr$$

The Roche limit is when this is *greater* than the self gravity of the Moon. Hence

$$\frac{GM_m m}{R_m^2} \leqslant \frac{2GM_p m}{a^3}R_m$$

This rearranges to give

$$a_R^3 = 2\frac{M_p}{M_m}R_m^3$$

Expressing the mass as volume of a sphere multiplied by density leads to the result as follows

$$\frac{M_p}{M_m} = \frac{\frac{4}{3}\pi R_p^3 \rho_p}{\frac{4}{3}\pi R_m^3 \rho_m} \implies a_R = 2^{\frac{1}{3}}\left(\frac{\rho_p}{\rho_m}\right)^{\frac{1}{3}} R_p$$

Note that if we didn't assume a sphercial Roche lobe, then the constant in front would be 2.456 instead.

5 Giant Planets

5.1 Introduction - Mass-Radius Relations

One can compare the the masses and radii of the various Solar System bodies with the mass-radius relations calculated for different compositions. This is shown in the following figure. The calculations for the bodies are based on the

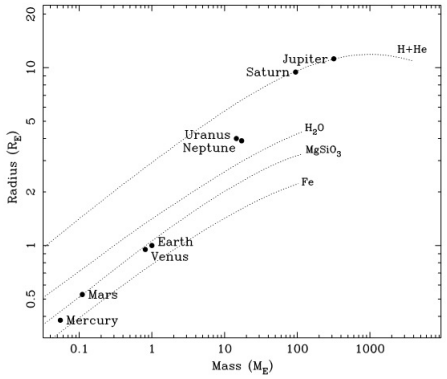


Figure 14: Mass-Radius Relations

assumptions of **hydrostatic equilibrium** and an **equation of state** (which relies on pressure). Looking at the data, one can immediately see that the terrestrial planets are consistent with having a rocky composition and an iron core. This also demonstrates that Mercury probably has a more dominant core. With regards to the giant planets, the data shows that Jupiter and Saturn are almost entirely comprised of hydrogen and helium. Uranus and Neptune on the other hand have a more intermediate composition, requiring the involvement of ice and rock. The density of these bodies also require gas so it is likely that other volatiles are included as well, for example: water, ammonia and methane. These elements are more cosmically abundant than the refactory elements and are able to form solids (or ices) at low temperatures. The dominance of ice is supported by the 10× over abundance of methane in their atmosphers in comparison to the gas giants.
We mentioned that to get these lines that we must consider hequations of state. The realistic equations for these are quite complex as well as being difficult to measure experimentally. In general, gas is more compressible than ice which in turn is more compressible than rock. However, at extreme pressures complex state changes begin to occur. Nevertheless, for hydrogen and a combination of hydrogen and helium the equation of state can be reasonably approximated by

$$P = K\rho^2$$

where $K = 2.7 \times 10^5\,\mathrm{Nm^4kg^{-2}}$. We will use this simplified equation of state in order to calculate the structure and radius of a planet. Assuming hydrostatic equilibrium, we get the familiar expression

$$\frac{dP}{dr} = -\frac{GM_r\rho(r)}{r^2}$$

We also assume mass conservation where

$$dM_r = 4\pi r^2 \rho dr$$

Finally, as discussed above we can assume the equation of state has the form

$$P = K\rho^2$$

Differentiating the equation of state gives

$$\frac{dP}{dr} = 2K\rho\frac{d\rho}{dr}$$

Then, by substituting this into the hydrostatic equilibrium equation, we can get

$$r^2\frac{d\rho}{dr} = -\frac{G}{2K}M_r$$

Differentiating again and substituting into the mass conservation equation gives the following second order ODE

$$\frac{d^2\rho}{dr^2} + \frac{2}{r}\frac{d\rho}{dr} + \left(\frac{2\pi G}{K}\right)\rho = 0$$

which has solution

$$\rho(r) = \rho_c\frac{\sin(kr)}{kr} \quad \text{where } k = \sqrt{\frac{2\pi G}{K}}$$

Here, ρ_c is the **central density**. The radius of the planet R will be where this expression has value 0. Hence

$$\rho(R) = 0 \implies \sin\left(\sqrt{\frac{2\pi G}{k}}R\right) = 0 \implies \sqrt{\frac{2\pi G}{k}}R = \pi$$

Ergo, the radius of the planet is given by

$$R = \sqrt{\frac{\pi K}{2G}}$$

which only depends on constants. This is true for the radii of gas giants with a similar mass to that of Jupiter, even when more realistic equations of state are used. This is since the increase of mass is balanced by the an increased compression of the interior. The radius of these bodies remains approximately constant for two orders of magnitude until the low mass stars are inflated by internal heating due to nuclear burning. Moreover, the radii of gas giant exo-planets seems to follow this flat mass-radius relation. This is with the exception of very close-in giants which are inflated by heating from their stars and a few over-dense bodies (most likely due to their unusually massive cores).
Now, if we wish to consider the central density ρ_c from above we can integrate the mass conservation equation

$$\begin{aligned} M = \int_0^R 4\pi r^2\rho(r)dr &= \int_0^R 4\pi r^2\rho_c\frac{\sin(kr)}{kr}dr \\ &= \frac{4\pi\rho_c}{k}\int_0^R r\sin(kr)dr \\ &\implies \rho_c = \frac{\pi M}{4R^3} \end{aligned}$$

Note that from this result you can also get the approximate central pressure of a gas giant planet by substituting the value back into the assumed equation of state.
Real equations of state are much more complex than our aprpoximation. They need to account for the mixture of elements, molecular chemistry, differentiation and phase changes. This phase diagram of hydrogen shows multiple phases.

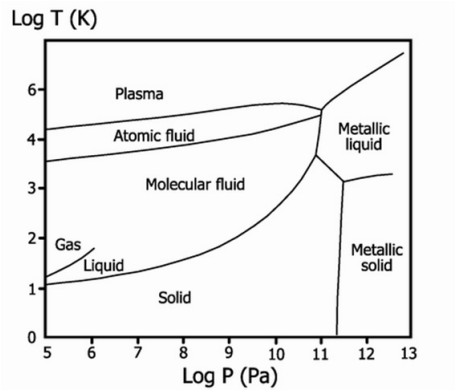


Figure 15: Phase diagram of hydrogen

If one considers the approximate temperatures of pressures in the interiors of the gas giants, a phase transition from molecular hydrogen to liquid metallic hydrogen can be shown to occur. At the extreme pressures, hydrogen electrons are no longer bound to individual protons and become mobile.[5]

5.2 Interior Structure (Gas Giants)

Due to the intense pressures and temperatures involved, gas giants have distinct layers with a gaseous, hydrogen dominated atmosphere above fluid molecular hydrogen envelopes which themselves are above a layer dominated by liquid metallic hydrogen. Below this there are believed to be cores comprised of ice,

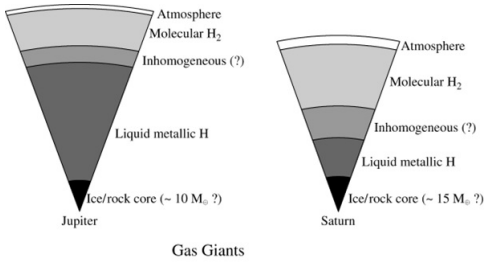


Figure 16: Interior of the gas giants

rock and/or iron about 10× the mass of the Earth. In the region in (16) marked *inhomogeneous* helium is no longer soluble in hydrogen and therefore is thought to form droplets and rain out, thereby driving ongoing differentiation and heating.

[5]This extreme state has been verified in laboratory experiments using shock waves.

We can actually go beyond this theoretical talk and make indirect measurements of the interior structures of the planet. We do this by studying the gravitational potential of the planet. Jupiter and Saturn rotate very rapidly for their size which in turn leads to a very high oblateness. This deviation from spherical symmetry means the gravitational field differs from a point source which allows the internal density structure to be determined.

5.3 Magnetic Field (Gas Giants)

The gas giants have conducting interiors of metallic hydrogen, as discussed before, as well as rapid rotation and convection. This seems to imply that Jupiter and Saturn should have a magnetic dynamo which we find is the case as both Jupiter and Saturn do have strong magnetic fields. Jupiter has a magnetic dipole moment of around 20000 times the Earth whilst Saturn has one of about 600 times that of the Earth. The magnetic dipole moment of Saturn is slightly lower as its region of metallic hydrogen is deeper and smaller.
Jupiter has also been a known source of radio emissions since 1955. These are in two components. That of very strong **cyclotron emissions** from low-energy orbiting magnetic field lines in polar regions and **synchrotron emissions** from high-energy (relativistic) electrons trapped in the radiation belts confined by a magnetic field.
The magnetosphere of Jupiter is the largest object in the solar system. This is from a combination of the facts that Jupiter has such strong magnetic dipole moment and that the solar wind is already spread much further out by the time it reaches Jupiter. We can estimate the extent of a general magnetosphere by equating magnetic energy density u_B with the kinetic energy density of the solar wind u_W. The magnetic energy density is given by

$$u_B = \frac{B^2}{2\mu_0}$$

where B is the magnetic field strength which is given by

$$B(r) = \frac{\mu_0}{2\pi}\frac{\mu}{r^3}$$

where μ is the magnetic dipole moment. The kinetic energy density is given by

$$u_W = \frac{1}{2}\rho_W v_W^2$$

Equating these implies the expression for the extent of the magnetosphere

$$\frac{1}{2\mu_0}\left(\frac{\mu_0}{2\pi}\frac{\mu}{r^3}\right)^2 = \frac{1}{2}\rho_W v_W^2 \implies R_m = \left(\frac{\mu_0}{\rho_w}\right)^{\frac{1}{6}}\left(\frac{\mu}{2\pi v_W}\right)^{\frac{1}{3}}$$

However, since the same solar wind impacts all the planets and we know about some magnetospheres more than others it can be more informative sometimes to consider how the radius scales between the different planets. We have the following proportionality relations for the two densities

$$u_B \propto \left(\frac{\mu}{r^3}\right)^2 \quad u_W \propto \frac{1}{a^2}$$

Therefore, by combining these we can get

$$\frac{\mu}{R_m^3} \propto \frac{1}{a} \implies R_m \propto (\mu a)^{\frac{1}{3}}$$

As the magnetosphere of a planet is defined by the interaction between the planetary magnetic field and the solar wind, it has a natural shape of being compressed on the side of the planet facing the Sun and with a long tail in the anti-Solar direction (which is shielded from the solar wind). This is a bit different for Uranus and Neptune as we will come to see.

5.4 Interior Structure (Ice Giants)

As eluded to previously, calculations with estimated abundances and realistic equations of state show that the interiors of these giant planets are dominated by ices (volatiles frozen at the time of formation) in liquid form where they make up about 90% of the mass. Below this, as shown in figure (17), there is most

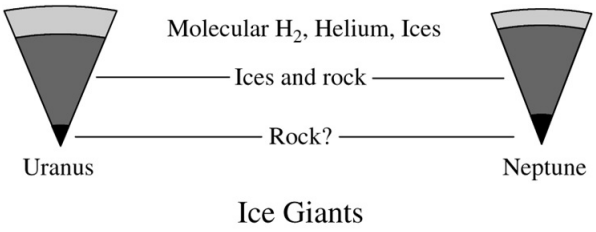


Figure 17: Interior of the Ice giants

likely rocky material that is partially differentiated. They also have envelopes of hydrogen and helium in liquid form (and gaseous atmospheres) that account for about two-thirds of their volume.
The ice giants are also rapidly rotating leading to oblateness. Both Uranus and Neptune have relatively high obliquities which are though to have been due to a giant collision during formation.

5.5 Magnetic Field (Ice Giants)

Despite being rapidly rotating, pressure in the gaseous envelope in Uranus and Neptune is not high enough for hydrogen to enter the metallic phase. Despite this, the ice giants have strong magnetic fields. This must arise from a magnetic dynamo acting in the fluid icy material. Here, pressure is sufficient to ionize the fluid ices which in turn provides the necessary electrically conducting medium. In addition to their high obliquities, both Uranus and Neptune also have unusually offset dipoles. This results in more complex *corkscrew* magnetospheres as shown in (18). For Neptune, the polar open field lines can at times point directly towards the Sun resulting in intense and extreme aurorae.

5.6 Internal Heat

The measured **effective temperatures** of Jupiter, Saturn and Neptune are significantly higher than their calculated **equilibrium temperatures**. Recally

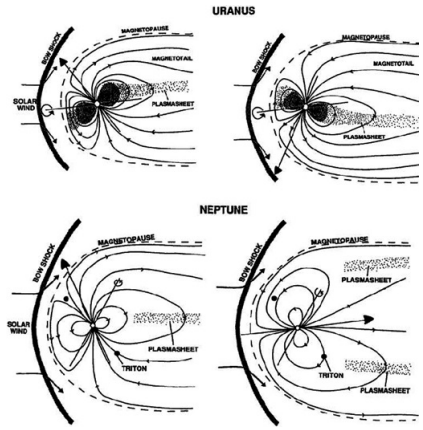


Figure 18: Magnetospheres of Uranus and Neptune

that the effective temperature of a body is the temperature of the blackbody that would have the same total luminosity as the thermal emission of the planet. The equilibrium temperature on the other hand are calculated by equating heating from the Sun with blackbody cooling. These high effective temperatures imply significant sources of internal heat. For Jupiter, this can be dominated by residual heat from gravitational collapse at formation. For Jupiter and Saturn, a deficiency of atmospheric helium suggest rainout of helium in the interior which would contribute to heating by ongoing differentiation. Neptune *may* be undergoing continuing differentiation. Uranus on the other hand is not well understood as there is not a discrepancy.

5.7 Atmospheres

5.7.1 Jupiter

The atmosphere of Jupiter has distinct bands of **bright zones** and **dark belts**. The bright zones are as a result of reflective clouds in the atmosphere. These clouds are thought to be comprised of ammonia in the top layer, ammonium hydrosulfide in the layer below that and then water. The distinctive colour of Jupiter is a result of unknown trace molecules created by ultraviolet-driven photochemistry. In the infrared the banding pattern is reversed since we can see deeper into the atmosphere at the cloud-free belt regions where the temperature is higher.
Jupiter has strong winds running in opposite directions in zones and belts. This is understood to be a global convection pattern that extends deep into the interior. The bright zones are where warm gas is rising which results in clouds condensing as said gas cools. Conversely, the dark belts are where cool material sinks and so the clouds evaporate as they sink. The pressure gradients created by the vertical convective motion drives North-South flows that are converted to East-West winds by the Coriolis force (since Jupiter is such a rapidly rotating planet).

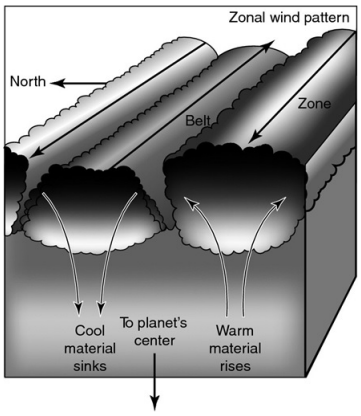


Figure 19: Wind pattern of clouds

5.7.2 Saturn

Saturn has a similar banding pattern to Jupiter. However, it is much more muted as clouds condense deeper into the atmosphere (which is cooler than Jupiter). Occasional large storms on Saturn can lift clouds to higher altitudes where they are more apparent. These powerful storms occur every few decades and could be seasonally related (the seasons are influenced and augmented by the shading of the planet by the rings) but it is difficult to be sure of this with our limited data. The north pole of Saturn has a huge hurricane located precisely at the pole with numerous smaller storms surrounding it.

5.7.3 Uranus

Uranus appears almost featureless but false-colour enhancements of the the planet bring out a faint banding structure. The atmosphere of Uranus has become much more active over the last twenty years.

5.7.4 Neptune

Neptune also has a pattern of clouds bands but this is even more muted than Saturn. These cloud bands are muted for the same reason as Saturn, since they are forming deeper in the cooler atmosphere. The overall colour of Neptune is blue as a result of the methane gas above the cloud layers strongly absorbing red light. Neptune also has transient high altitude reflective clouds of methane ice as well as dark spots that presumably related to storm systems. Seasonal variations may be visible as more reflective clouds appear in the Southern hemisphere as Neptune moves into Summer. Moreover, forcing by solar radiation may be driving storm systems.

5.8 Moons

Firstly, we consider the four main, large moons of Jupiter known as the **Galilean moons**. These were important in the debate of geocentric vs. heliocentric models as they demonstrated that not all astronomical objects circled the Earth. Jupiter is known to have at least 79 moons, but the others are much smaller,

mainly captured asteroids. The Galilean moons, in increasing order of distance from Jupiter, are **Io, Europa, Ganymede and Callisto**. As you move further out, the density of the moons drops with an increase in volatile materials in the composition as well as an increase in impact craters, corresponding to an increase in surface age as the lack of impact craters in the inner moons suggest ongoing resurfacing. The increase in volatile content may indicate the formation occurred in a circumplanetary disc with a temperature gradient, reminiscent of the protoplanetary disc in which the planets likely formed (see next section).
If we place the moons on the previous mass-radius diagram, we find that Io is similar in composition to the Moon. The gravitational moment also demonstrates a differentiated rocky interior. The moon density *drops* with orbital seperation as stated before. Ganymede specifically has a much lower density which suggest that the internal composition is dominated by water (this reflects the formation beyond the snowline of both the Sun and Jupiter).[6] With regards to the surface of Io, we see intense volcanism (up to nine volcanoes erupting at any one time). The surface is therefore quite young, with no impact craters, from this continuous resurfacing by volcanism. Recall that when we discussed

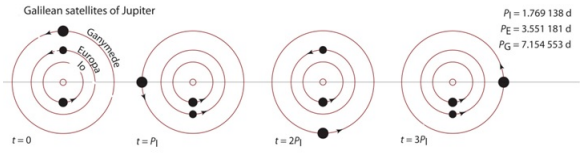


Figure 20: Mean motion resonances of the Galilean Moons

mean motion resonances, we used the Galilean moons as an example of how despite the resonances exciting the eccentric orbits, these resonances are damped by the ongoing tidal interaction with Jupiter. This also results in internal heating of the planets, especially for the closest moon Io where the volcanism is driven by this heating. Tidal heating also occurs on the less dense Europa. This results in what is thought to be a sub-surface liquid water ocean below a thin ice crust. The surface of Europa is also regularly resurfaced by liquid water. Ganymede is the largest and most massive moon in the Solar System, being bigger than Mercury. It has a thicker water crust, with less cracking and more impact craters. Despite this, resurfacing still occurs. The linear features suggest that in the past there may have been a sub-surface ocean as well as water-based volcanism and geology has led to resurfacing (but at a slower rate when compared to Europa).
We now turn our attention to Saturn, which has at least 82 moons, including 7 that are approximately spherical. However, the total mass of the moon system ($\sim 96\%$) is dominated by **Titan**. Titan is the second largest moon in the Solar System with a mass and radius similar to Ganymede. Remarkably, Titan has a thick atmosphere dominated by nitrogen with a surface pressure of 1.5 atmospheres. The atmosphere of Titan also includes methane as well as being shrouded by a thick high-altitude photochemical haze of hydrocarbons. Moreover, at the stated surface pressure and temperature of 93 K, methane can condense into liquid form. Evidence was found for methane rain and surface flows, much like water on Earth. Images from the surface showed rounded boul-

[6]Note that Ganymede is larger than Mercury but less massive.

ders and pebbles of rock-hard water ice. Another moon of Saturn that is of particular note is **Enceladus**, which is relatively small and orbits fairly close to Saturn. It has a tidally-heated liquid water ocean beneath an icy crust with water geysers breaking through the surface.
Uranus has at least 27 moons, of which 5 are approximately spherical. The moons appear to be composed of rock and ice and have a range of cratering and resurfacing. There also may be sub-surface oceans on the largest of these. It is interesting to note that the orbits are all aligned with the 98 degree spin obliquity of Uranus, suggesting formation from debris of the same giant impact. The smallest of these moons, **Miranda**, stands out with extraordinary surface features, called **coronae**, the origin of which is not understood. The coronae have fewer impact craters, showing that they are younger than the rest of the surface. It has been suggested that the moon was fractured by a giant impact, or that slightly smaller impacts led to melting and liquid water flooding the surface.
Neptune has at least 14 moons, but only one is approximately spherical. That moon is **Triton**. It has a retrograde orbit, suggesting that it may have been a captured Kuiper Belt object that potentially destabalised the original moon system. Geyser-like jets have also been detected and recent resurfacing that has been suggested as being a result of water volcanoes driven by tidal heating.

5.9 Rings

The giant planets all have rings systems, of which the one of Saturn is the most famous. The rings of Saturn are composed of icy particles predominantly in the centimetre to metre range. The gaps correspond to unstable orbits which correspond to strong resonances. The rings are very thin, maybe only tens of metres. This is probably maintained by collision of particles, as if they start to leave the plane they are still in a Keplerian orbit which crosses the plane of the rings, increasing likelihood of collisions which keep the objects in plane.
The main rings visible from Earth (A-C) extend out from 1.2 to 2.3 Saturn radii. Additional rings were observed by Voyager 2, including the E ring, which extends out to 8 Saturn radii. The main rings are likely to have originated from the disruption of an object that crossed its Roche limit. This corresponds with the extent of the main rings, which matches the Roche limit for an object of density around $1 \times 10^3\,\mathrm{kgm}^{-3}$. The main Cassini division corresponds to a $2:1$ orbital resonance with the moon Mimas. The E-ring is likely to be material emitted by Enceladus.
Other objects have been observed breaking apart after crossing the Roche limit of a body, for example, the comet Shoemaker Levy 9 disrupted into at least 21 fragments in 1992 while passing close to Jupiter and crossing its Roche limit. In 1994 the fragments collided with Jupiter.
The ring systems of the other three giant planets are much less prominent and consist of dust, most likely emitted from moons.

6 Formation of the Solar System

6.1 Protoplanetary Disc

The nebula hypothesis is the theory that the Sun and the planets formed from the gravitational collapse of a local over-density interstellar medium. The collapsing nebula formed a disc around the the rapidly rotating proto-Sun. The reasons for this is that for a cloud of gas to collapse under gravity it must lose two things, those being energy - since as it collapses the gravitational potential energy gets converted to kinetic energy which it must lost in order to collapse into a small body - and angular momentum - since there will be some angular momentum in the overall gas cloud which could stall the cloud as it attempts to collapse down. The energy is fairly to transfer, since as the gas collapses it will heat up and radiate the energy into space. The angular momentum on the other hand is harder to get rid of, which leads to the natural situation where the mass is concentrated at the center and a disc around the mass which contains most of the angular momentum of the original dust cloud.
This resulting disc is thought to be the origin of the planets and smaller bodies of the Solar System. This hypothesis is supported by the fact that the planets all orbit roughly in the same plane in approximately circular orbits, aligned with the spin of the Sun. Protoplanetary discs are examples of **accretion discs**. Material is in an approximately Keplerian orbit around the central mass but *viscosity* between neighbouring annuli allows material to lose angular momentum and gradually spiral in or gain angular momentum and moves outward. As material moves towards the central object, it falls deeper into the gravitational potential well and gravitational potential energy is released which heats the disc. If one assume heat is radiated locally as blackbody emission, a steady-state accretion disc will have a temperature profile of

$$T \propto r^{-\frac{3}{4}}$$

One can ascertain a lower limit on the mass of the protoplanetary disc by

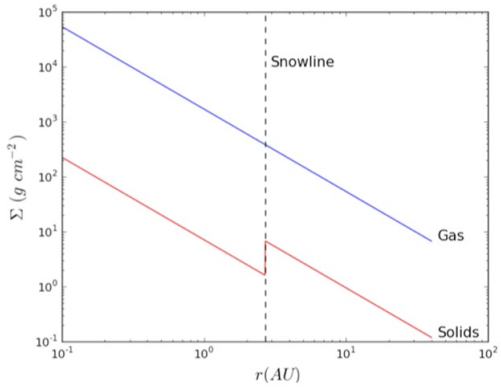


Figure 21: Surface density plotted against separation

smearing out the current mass of the planets into annuli and *topping up* the mass to cosmic abundance. This is known as the **minimum mass solar nebula**

which leads to a surface density (collapsing the disc with a density into a plane) profile of the following power law distribution

$$\Sigma(r) \propto r^{-\frac{3}{2}}$$

which gives us an idea of how the mass might have been distributed in the protoplanetary disc. One should note that whilst density is dropping off with radius, the area of the disc is increasing by the radius squared. Hence, the total mass of the disc is dominated further out. The mass is dominated by gas but one can see that once the **snowline** is crossed, the disc becomes cool enough for ices to form (specifically water can freeze). This means there is much more solid material available, as the material that makes up ices is much more abundant in the cosmos than the material that makes up the other solids. By integrating under this power law, one can find a lower bound of the total mass of the disc which turns out to be

$$M_{\text{disc}} \geqslant 0.01 M_{\odot}$$

Protoplanetary discs have been directly imaged around young stars. The proportion of these young stars with an excess of infrared emission from a protoplanetary disc drops rapidly with age. This demonstrates that protoplanetary discs only survive for a few million years and that planet formation must proceed on this timescale.
Protoplanetary discs should not be confused with **debris discs**, which are around much older stars but are less massive and consist only of dust, which is believed to have been replenished by collisions between asteroids.

6.2 Core Accretion Scenario

The core accretion scenario is the leading model of planet formation. Terrestrial planets form predominantly within the snowline where rocky materials can form solids. Beyond the snowline gas giants form much more massive cores that include ices and rocky material. This core forms relatively quickly and eventually reaches a critical mass where runaway gas accretion occurs, leading to the total mass increasing rapidly until the local supply of gas is exhausted. Ice giants

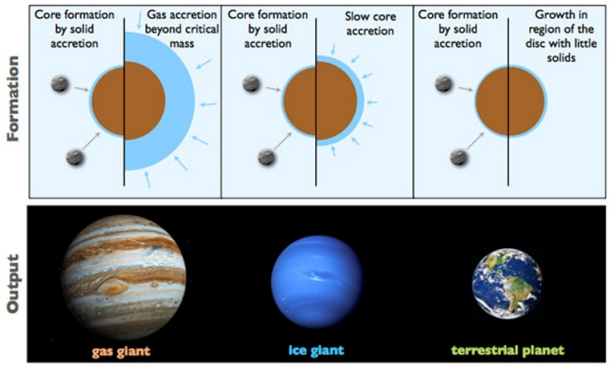


Figure 22: Core accretion scenario

also form massive cores beyond the snowline but fail to reach the critical mass

before the gas disc is photoevaporated by the young star.
This process can be broken down into approximately seven steps as follows:

(i) Gravitational collapse of over-dense region in interstellar medium to form proto-star with protoplanetary disc.

(ii) Condensation of sub-micron-sized dust particles within disc.

(iii) Growth of solid particles by pair-wise collisions, from dust to 10 km sized planetesimals, initially assisted by electrostatic forces.

(iv) Further growth by pair-wise collision from planetesimals to protoplanets with mass around $0.1M_E$, assisted by mutual gravity. Growth to isolation mass, where Hill sphere of protoplanet has swept out an annulus in disc.

(v) Larger cores form beyond snowline in disc, where temperature low enough for ices to form.

(vi) Gas from the disc is accreted by these larger cores to form the giant planets. Limited by photo-evaporation of the disc by the young Sun.

(vii) The terrestrial planets assemble on a longer timescale by orbital perturbations leading to giant collisions.

One big challenge of this process is how dust can grow to macroscopic scales. Laboratory experiments demonstrate that growth of solids via pair-wise collisions of dust particles does occur and is initially aided by electrostatic forces, as stated above. Experiments also show collisional growth can build macroscopic *dust cakes* which are larger but quite fragile. Icy particles show much more efficient growth of particles due to their sticky nature. Ice coatings on silicate particles also aid growth for solids.
Another challenge is how the dynamics of these solid particles is affected by the gas drag. The orbit within the protoplanetary discs of gas is given by

$$\frac{v^2}{r} = \frac{GM_\odot}{r^2} + \frac{1}{\rho}\frac{dP}{dr}$$

where the first two terms are from standard Keplerian orbits and the additional support against gravity comes from the gas feeling its own pressure. This means the gas does not have to orbit as fast for there to be a balance. This means in practice that the velocity of the gas is lower than the Keplerian velocity, approximately such that $v_{\text{gas}} \approx 0.995 v_{\text{kep}}$. This leads to a headwind that the solid objects feel of about $v_{\text{wind}} \approx 100\,\text{ms}^{-1}$. This results in a significant drag on the solid objects. This results in the solid objects being pulled into the midplane, increasing the chances of collisions and hence growth. However, this drag also cause the particle to gradually spiral in to the proto-Sun, potentially losing them. In order to put some numbers to how long this might take, we consider the drag force given by

$$F_D = -\frac{1}{2} C_D \pi R_p^2 \rho_{\text{gas}} (v_{\text{kep}} - v_{\text{gas}})^2$$

where R_p is the radius of the solid particle. The characteristic timescale to alter the orbit of these solid particles will be given by

$$\tau_D \sim \frac{M_p v_{\text{kep}}}{|F_D|}$$

The mass of the particle will be given by

$$M_p = \frac{4}{3}\pi R_p \rho_p$$

So the timescale is hence given by

$$\tau_D \approx \frac{8\rho_p R_p v_{\text{kep}}}{3C_D\rho_{\text{gas}}}(v_{\text{kep}} - 0.995v_{\text{kep}})^2 = \frac{8\rho_p R_p}{3C_D\rho_{\text{gas}}0.005^2 v_{\text{kep}}}$$

If we consider the values of 1 astronomical unit, a Keplerian velocity of $3 \times 10^4\,\text{ms}^{-1}$ and hence a headwind of $100\,\text{ms}^{-1}$ then for a planet we get a timescale of about $\tau_D \sim 1 \times 10^8$ years, longer than the lifetime of the gas disc. On the other extreme we get that for a dust particle the timescale is mere seconds. However, the dust is carried by the gas so moves as such. The issue comes for mid-sized objects of about 1 metre, where the timescale is that of about a 1000 years. This means objects of this size tind to spiral in and be removed from the disc. This is a major challenge for this model. Growth at this scale needs to be rapid and it is thought that this can be aided by **instablities in disc dynamics**. This results in a non-uniform pressure profile meaning that whilst the pressure gradient is overall negative, which leads to the inward drift of metre-scale objects. However, the density waves and other structures can lead to pressure maxima and local outward drift. This in essence traps particles in

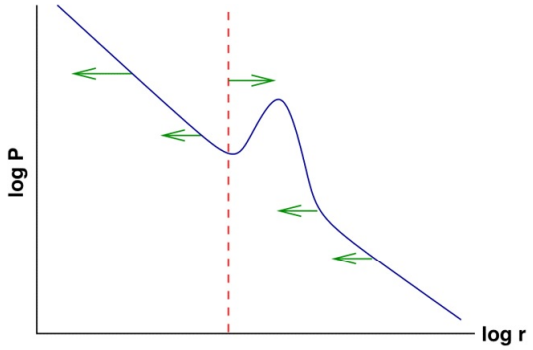


Figure 23: Non-uniform pressure gradient

local disc structures, increasing the rate of particle growth. A similar effect can be found in **streaming instability**, where particles shelter one another from the headwind. This results in slower drift of particles and increased clumping, thereby aiding growth beyond this problematic zone. Finally, we wish to consider the more massive cores form beyond the snowline and how this will depend on seperation. From the surface density relation found earlier, we can find that the mass of solids in an annulus of thickness dr, denoted dm, has the following proportionality

$$dm \propto 2\pi r dr r^{-\frac{3}{2}} \propto r^{-\frac{1}{2}} dr$$

From this, if we consider the spacing of planets dr to be increase approximately proportional to distance r, we can get the following result

$$M_{\text{core}} \propto r^{\frac{1}{2}}$$

This is not the end however, as beyond the snowline the amount of available solid material is approximately 4× higher than at the separation of Earth due to the relatively high abundances of hydrogen and oxygen (as well as nitrogen and carbon). Therfore, the mass of the core of a planet located at a distance d and beyond the snowline is given by

$$M_{\text{core}}(d) \approx 4 \cdot M_{\text{core}}(1\text{au}) \cdot d^{\frac{1}{2}} = 4 \cdot M_E \cdot d^{\frac{1}{2}}$$

6.3 Disc Evaporation

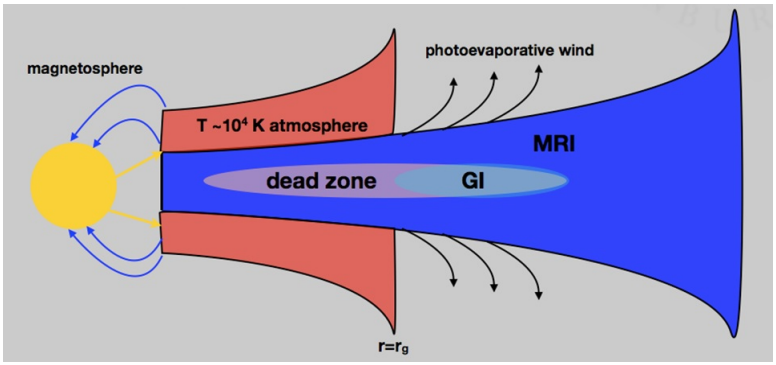


Figure 24: Diagram of disc evaporation

The gas disc is dispersed by X-ray driven photoevaporation once the Sun turns on as an active star.[7] We can find a radius for which the disc is evaporated off after it by equating thermal and kinetic energy

$$\frac{3}{2}kT = \frac{1}{2}m_p\bar{v}^2$$

where T is set by the radiation temperature of the chromosphere, which dominates the UV photon production ($\sim 1 \times 10^5\,\text{K}$) and $\bar{v}$ is the mean velocity of the gas. We then consider the escape velocity from the disc as

$$v_{\text{esc}} = \sqrt{\frac{2GM_\odot}{r}}$$

and equate it to the mean velocity, meaning the thermal velocity is equal to the escape velocity. Rearranging for r gives

$$r = \frac{2GM_\odot m_p}{3kT}$$

which is approximately equal to 10au for the values we consider. This means that for radii beyond that the heating is putting particles above the escape velocity, evaporating the outer parts of the disc off and hence shutting off the supply of new gas to the planets.

[7] This is a similar process to atmospheric escape from planets.

6.4 Assembly of the Terrestrial Planets

When the gas disc dispersed, the inner Solar System is thought to have included many rocky protoplanets that had each grown to its isolation mass. While the orbits of the planets are stable to mutual perturbations today, this was probably not the case for the protoplanets where the growth of eccentricity had been previously damped by the gas disc. Secular resonances will have grown eccentricities until orbits crossed and protoplanets collided, eventually forming the four terrestrial planets and Earth's Moon that we see today.

7 Habitability and Extra-Terrestrial Life

7.1 Requirements for Life

Life is surprisingly hard to define, as attempts are often too narrow and fail to include some living things, for instance mules, which cannot reproduce, or viruses, which do not have a metabolism. Other attempts are too broad and include things that are not alive, such as crystals, which can self-replicate, or fire, which replicates and can be considered to have a metabolism.
We can look for more robust requirements by analysing extremophiles, which are able to survive in extreme conditions such as ionising radiation, extreme pH values or incredibly high pressures. From this, we can infer some fundamental requirements of life. These first of these is thought to be a generally available energy source, for instance Solar radiation for photosynthesis, as life requires non-equilibrium chemistry. The next requirement may be carbon, which is needed for complex organic chemistry. However, this is unlikely to be a limiting factor as it is the sixth most abundant element in terms of cosmic abundance. Silicon-based life is often discussed, since it has the same valency as carbon, but silicon forms strong bonds with oxygen and gets tied up in silicate rocks. Finally, an abundant solvent which provides a benign environment for organic chemistry is probably required. This is usually assumed to be liquid water, as it is required for all life on Earth. However, liquid water is rare in the Solar System which may point to it being the key limiting factor for extra-terrestrial life. Fossil and chemical records for life is found in the oldest sedimentary rocks on Earth, approximately 3.8 billion years ago. This suggests that life began as soon as conditions were favourable, which supports the possbility of being able to find life elsewhere in the Solar System. Recall that the Sun was 25% less luminous as this point in the history of the Solar System, which makes it surprising that conditions were so favourable for life at this point. This is the faint young Sun paradox, which suggests that there was a stronger greenhouse effect on Earth at this time.

7.2 Environments for Life

We finally wish to consider how the requirements of life may be found in other bodies of the Solar System. Mars has long been the focus of efforts to detect extra-terrestrial life in the Solar System. It is the only other object in the Solar System that can plausibly have liquid water at the surface. Moreover, it has an atmosphere and is located in the **habitable zone** of the Sun, where surface temperatures allow liquid water, at least for some time.

7.2.1 Habitable Zone

The habitable zone is the distance from a star where an Earth-like planet could have liquid water at the surface. Earth and Mars are located within this zone. Notice that the Earth is on the inner edge of this zone. As the Sun continues to brighten, the Earth is projected to become uninhabitable in 1-2 billion years. Mars is only *borderline* habitable due to a low atmospheric pressure due to atmospheric escape. At this atmospheric pressure, Mars is very close to the triple point of water, below which liquid water is not possible. However, atmospheric

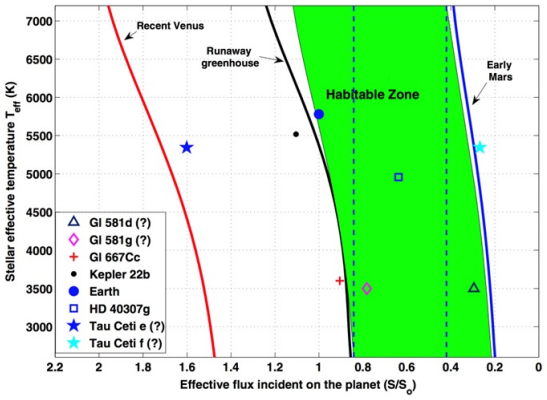


Figure 25: Diagram of the scale of the habitable zone

pressure is slightly higher than this in the deepest canyons and craters. However, there is plenty of evidence for extensive flows of liquid water on Mars in the past. This includes river valleys and deltas, as well as geological and chemical evidence from sedimentary rock formation and hydrated minerals. However, impact craters demonstrate that these are ancient and date from the very early history of Mars. Like Venus, Mars has a strongly enhanced deuterium to hydrogen ratio which implies that there was significant loss of water to photolysis and atmospheric escape. Despite this, water ice and seasonal variations are readily detected. This suggests that there is at least the occasional transient precense of liquid water at the surface during episodes of melting.
Studies with ground-penetrating radar have found evidence for sub-surface lakes of liquid water about one kilometer below the South polar icecaps of Mars. Detected as a high reflectivity layer, similar to sub-glacial lakes in Antarctica. However, it is likely that this water is hyper-saline and therefore may not be a suitable habitat for life. The icecaps on top provide the confinement of liquid water, even with no atmospheric pressure.

7.2.2 A Warning on Ambiguity

However, one must be cautious when looking for evidence of liquid water on other bodies as the evidence for it can be ambiguous. On Mars for example, recurring slope linae (RSL) are dark streaks of downward slopes that seem to vary with the seasons. Some publications suggested that these were seasonal water flows and that hydrated salts had been detected. However, more recent research suggests that since they are only found on steep slopes that they are due to the flow of dry sand.
Another cautionary tale is the case of martian meteorite Alan Hills 84001 (ALH84001), which is by far the oldest of the 130 or so known (at about 3.6 billion years old). There were claims of microbial microfossils and organic magnetite materials. These claims proved highly controversial, with alternative explainations being proposed such as abiotic mechanisms.

7.2.3 Other Bodies

On Venus, habitability would not usually be considered likely due to the extreme high surface temperature and loss of water. However, in a 2020 study phosphine was claimed to have been detected in the atmosphere of Venus. This is difficult to explain by abiotic chemistry alone as phosphorous would be oxidized in the atmosphere. Whereas phosphine is a known product of some bacteria on Earth. Moreover, liquid water droplets *are* available on Venus, in the form of concentrated sulphuric acid in the middle and lower cloud levels where the temperature is lower and the atmospheric pressure is similar to that of Earth. Here, we find a potential habitat for microbes to be able to survive. Nevertheless, this claimed detection still remains controversial.

As discussed previously, the moons of the giant planets are rich in water and heated internally by tides. Some of these moons, for example Europa, have very young surfaces with few impact craters, indicating ongoing resurfacing of the icy surface. This young icy surface is covered in an intricate network of cracks, showing ongoing geological activity. These features are thought to have arisen from tectonic activity, in a similar manner to the tectonic plates found on Earth. Considering the interior of Europa, recall that it fairly volatile poor in comparison with Ganymede and Callisto but still has plenty of water to have a global sub-surface ocean. Significantly, this sub-surface ocean is also shallow enough to remain liquid down to the rocky interior with hydrothermal vents likely to cycle nutrients into the ocean. Whilst Ganymede and Callisto are more volatile rich, with deeper water layers, they have weaker tidal heating and older surfaces. It is unclear whether they have sub-surface oceans, although Ganymede does exhbit features of tectonic activity like Europa. If these bodies do have deep oceans, then the ocean floor may be solid ice due to the extreme pressure. Moreover, there would be a lack of hydrothermal vents.

The dwarf planets are also water rich as well as showing evidence for geological activity. They may also have sub-surface oceans, although they have less tidal heating than the icy moons discussed previously.

The surface of Enceladus is mainly very young and exhbits similar tectonic activity to Europa. Water geysers were detected breaking through the icy surface by the Cassini probe. Mass spectrometry was able to be performed on these plumes which confirmed they were mainly dominated by water with hydrocarbons and molecular hydrogen being present. These indicate hydrothermal vents and demonstrating a potential energy source for life.

Finally, note that the hydrocarbon lakes detected on Titan are the only standing surface liquid in the Solar System, besides water on Earth. These hydrocarbons could conceivably replace water as the solvent needed for life to evolve.