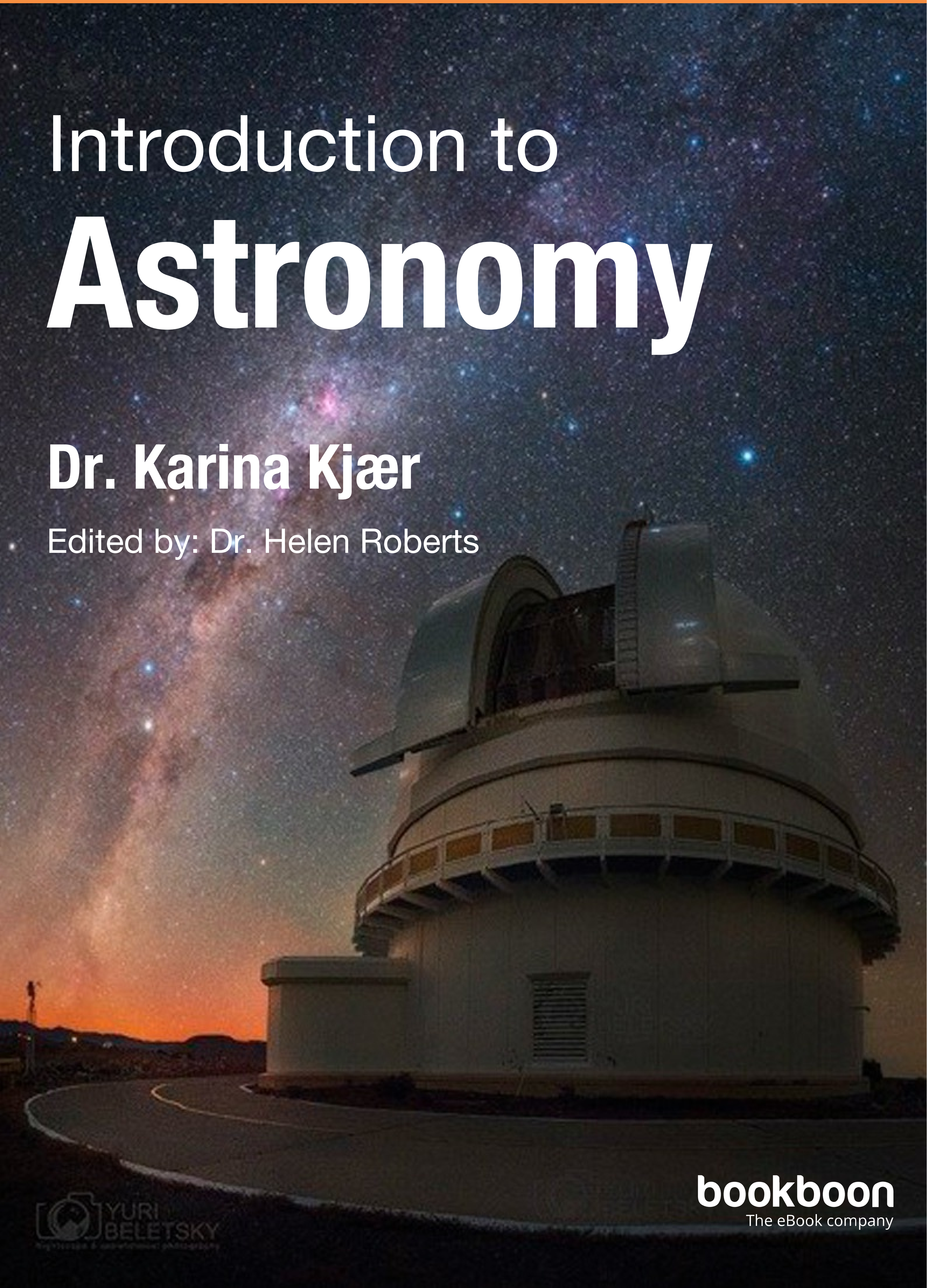
Introduction to Astronomy

Dr. Karina Kjær

Edited by: Dr. Helen Roberts



bookboon
The eBook company

DR. KARINA KJÆR

INTRODUCTION TO ASTRONOMY

Introduction to Astronomy
2nd edition

ISBN 978-87-403-2912-4
Peer review by Dr. Helen Roberts co-founder and COO of Redfern Electronics (redfernelectronics.co.uk).
Cover photo by Dr. Yuri Beletsky: A pre-dawn view from Carnegie Las Campanas observatory located in the South of the Atacama desert in Chile.

CONTENTS

1 THE EARTH-MOON SYSTEM

Your goals for this chapter are to learn about:

- The seasons on Earth.
- The phases of the Moon.
- Solar and Lunar eclipses.
- Basic celestial mechanics and orbits.

1.1 THE NIGHT SKY

If you go outside on a cloudless night and look up you will see countless stars. I urge you to leave the light of the city, seek out a dark place without street lamps and look up. It is truly amazing. As you look up into the darkness your eyes get more and more sensitive to the light and soon you can make out a white trail of stars across the night sky. That is our galaxy the Milky Way, named so by the ancient Greeks.

If you are fortunate enough to look up at the night sky from the South- ern hemisphere, you are looking towards the centre of the Milky Way, where the density of stars is at its greatest. If you are somewhere really dark, you may even be able to see the dark dust clouds running along its band of stars. Looking at the Milky Way from the Northern hemisphere is still beautiful, albeit the view is away from the centre and the density of stars much less.

As the Earth makes its way around the Sun some stars remain in view throughout the year, and some we associate with the seasons. This is be- cause the rotation axis of Earth, which to a first approximation goes through the North and South pole, points towards the same place in the sky. This does not change as Earth revolves around the Sun, because the stars are very far away compared to the relatively small distance Earth moves in half a year, when the Earth is on the opposite side of the Sun. The stars that are always in view from the Northern hemisphere, are those which the rotation axis points towards, when we imagine the axis to be extended all the way to the stars. The same is true on the Southern hemisphere, where the stars always in view are in the area, to which the rotation axis points when extended the other way.

Another way to understand this is to imagine the solar system to be confined to a very large flat surface (like a very large sheet of paper). The Sun sits in the middle and the planets orbit around it in circles drawn on this paper. The rotation axis of Earth points out of the paper. All the stars above (and below) the paper will remain in view throughout the year.

The stars that change with season are those who sits on the paper, when it is extended far beyond the solar system. The stars are actually in view of the Earth always, but we cannot see them during the day. In other words, the stars associated with winter will be in the sky during the day in summer. And conversely, the stars associated with summer are in the sky during the day in winter. The stars associated with the constellations of the zodiac signs are all associated with seasons.

In Mathematics and Physics a large flat surface described above is called a plane, and most of the major planets in the Solar System orbit the Sun in nearly the same plane. Viewed from Earth the Sun traces out a path across the sky. This path is called the Ecliptic, and it is actually showing us the plane of Earth's orbit. Because the planets in the Solar system all orbit in planes close to the ecliptic, they appear on our night sky close to the path the Sun traced out during the day.

1.2 THE EARTH'S SEASONS

The Earth is the third planet from the Sun in our solar system. It spins around its own axis every 24 hours, and it orbits the Sun once every 365.25 days thus making an Earth year. Earth's seasonal changes are caused by the tilt of Earth's rotational axis with respect to its orbit around the Sun.

Due to conservation of angular momentum an unperturbed planetary system will have all the planets going the same way around the star, and all planets spinning the same way on their rotational axes. However, planets can deviate from spinning perpendicular to their orbital planes if they have been part of a merger or suffered impacts from larger bodies. Earth's rotational axis is not perpendicular to its orbital plane, it is tilted by 23.5^{o} away from the perpendicular, see Figure 1.1.

Remember that the rotation axis always points to the same spot in the sky as the Earth makes its way around the Sun. This means that some- times the Northern hemisphere will be tilted towards the Sun (summer in North), and sometimes it will be pointing away from the Sun (winter in North). When the Northern hemisphere is experiencing summer the Southern hemisphere is experiencing winter, and conversely when it is winter in the Northern hemisphere it is summer in the Southern hemisphere.

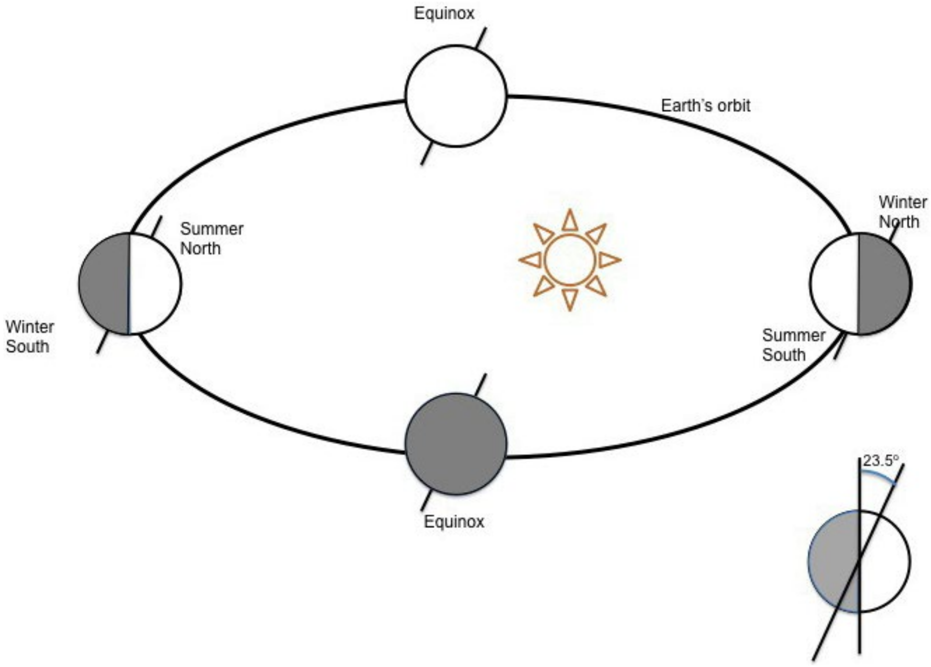


Figure 1.1: Earth's seasons are due to a tilt of the rotation axis of 23.5°.

The spot that the rotation axis points towards actually changes very slowly over time, which is called precession. It is as if the pointing of the rotation axis traces out a small circle over 26,000 years.

Summer is the hottest season, not only because the days are longer, but also because the position of the Sun is higher in the sky, i.e. more towards zenith -the point directly above your head. That means the shadows are smaller and the ground receives more direct sunlight.

Another important factor in how much sunlight you get is your latitude. Latitude is your distance from Equator as measured in angular degrees, 0° is at the equator and +90° is the North pole (–90° is the South pole). At either of the poles summer and winter are very different. Above the arctic circle at latitude +66°, the Sun never sets in summer, and in winter it never rises. However, even in summer the Sun will never get close to zenith, and consequently the shadows will never become small and the ground will have less direct sunlight. Furthermore, the sunlight will have to travel through a lot of atmosphere to hit the ground and thus be weaker than what is received at the equator. If you live at the equator the maximum height of the Sun is close to zenith all year around. The days and nights are almost equally long and you do not experience the seasons as a change in sunlight, but rather as the occurrence of weather, e.g. rain season, drought and so on.

1.2.1 EQUINOX AND SOLSTICE

The four seasons are highlighted by an Earth-Sun system event. The summer solstice marks the longest day in the year for a given hemisphere, in Northern Europe it is also often called midsummer. When summer solstice happens in one hemisphere the other will have winter solstice. Winter solstice is the shortest day in the year.

Equinox is when the night and day are equally long. That happens twice a year, once in spring and once in autumn. It is the whole planet that has equinox at the same time, as the rotation axis of Earth is pointing neither towards the Sun or away from it. The spring equinox is also called the Vernal Equinox, and while it marks the onset of spring in the Northern hemisphere it marks the beginning of autumn in the Southern hemisphere. Similarly the autumn equinox marks the moment when autumn starts in the Northern hemisphere and spring in the Southern hemisphere. At equinox the Sun rises directly in the east and it sets directly in the west.

1.3 PHASES OF THE MOON

The Moon shines because it reflects sunlight. Just as Earth experiences day and night because of its rotation with respect to the Sun, so does the Moon rotate and experience a day and night. We see the Moon where the sunlight hits it, i.e. we only see where it is day on the Moon. The part of the Moon that is in shadow -its night side, appears invisible to us. Because of this the Moon changes appearance in the sky. Sometimes it is full, sometimes it is half and sometimes it is barely there at all. These different appearances are called phases, or lunar phases. The Moon's phases depend upon, where the Moon is relative to the Sun as seen from the Earth. Or in other words: A lunar phase depends solely on how much of the Moon's illuminated part is visible from Earth. Figure 1.2 shows the lunar phases and their corresponding names.

The Moon orbits the Earth in about 4 weeks, during which the Moon completes a full cycle of its phases. During new moon (when the Moon is not visible, i.e. where the Moon's night side is facing us) the Moon is on Earth's day side. Since Earth rotates once every 24 hours, the whole Earth will notice that the Moon is not visible. In Figure 1.2 we see that the Moon is positioned between us and the Sun during new moon. When the Moon is on the opposite side of the Earth, that is, when we are between the Moon and the Sun, we will see the Moon as full. It is not possible for us to see the full moon during the day. Consequently the full moon rises at sunset and sets at dawn.

Although the Moon changes phases, it always keeps the same side facing the Earth. Consequently, Earth-based observers always see the same craters and lunar mountains, regardless of the Moon's phase. This is because the Moon's orbital period around Earth matches its rotational period. The Moon spins around its own axis once every 28 days, and it orbits Earth once every 28 days. That the rotation period of the Moon matches its orbital period is no coincidence and is by no means unique in nature. The phenomenon is called tidal locking (sometimes also gravitational locking or captured rotation) and it arises when orbiting bodies influence each others rotation via tidal forces. Tidal forces are basically gravitational forces acting between objects of size. The gravitational pull of a planet on its moon alters the moon's shape from its original sphere. The moon also creates, what is called a tidal bulge on its parent planet. On Earth, the effect is what causes the tides.

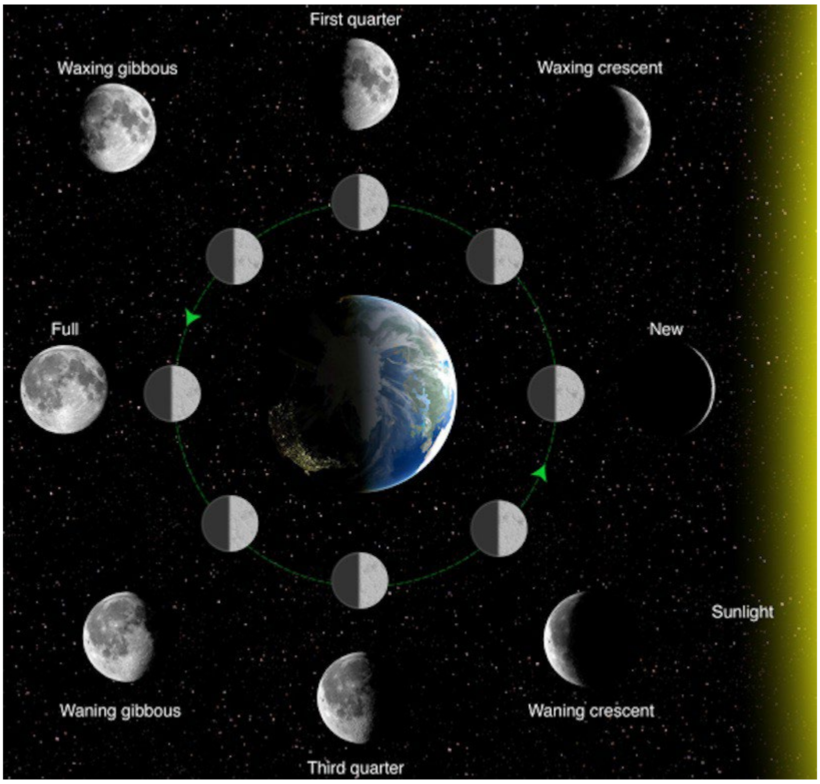


Figure 1.2: Lunar phases and the Moon's position with respect to the Earth and the Sun. Illustration courtesy of NASA.

In the Earth-Moon system the rotation period of Earth is lengthening at the rate of 0.0016 seconds per century, the energy of its rotation being lost into friction from the tidal interaction. Similarly, the Moon is drifting away from us at a rate of 3-4 cm per year. This is a consequence of conservation of angular momentum, which is lost from the Earth's rotation and is gained in the orbital angular momentum of the Moon. Because of these tidal effects, far into the future, Earth will have slowed down its rotation so that the same side of the planet always faces the Moon, just like the same side of the Moon is now always facing us. By then a day will be 47 current days length long, or 47 × 24 hours, and the Earth-Moon system will be double tidally locked.

Tidal locking is the fate of every orbiting system having tidal forces. The smaller body in the interaction will always get locked first, but the larger one will eventually follow. Tidal locking has already happened for several moons in the solar system for example Saturn's moon Triton and the moons of Mars. The dwarf planet system of Pluto and its moon Charon is double tidally locked, with both bodies always showing the same side to each other, while they orbit a point halfway between them.

1.4 SOLAR AND LUNAR ECLIPSES

A lunar eclipse occurs when the Moon passes through the Earth's shadow. This can happen only when the Sun, Earth, and Moon are sitting on a straight line. From the considerations about lunar phases, it becomes apparent, that a lunar eclipse can only happen, while it is full moon. If the Moon was orbiting Earth in the plane that spans Earth's orbit we would have eclipses every new moon and every full moon. That is not the case. The Moon's orbit spans a plane, which is tilted by 5 angular degrees with respect to Earth's orbital plane. In this set up eclipses occur whenever the Sun, Earth, and the Moon are aligned.

The shadow cast by the Earth has two distinct parts: The darkest part (full shadow) is called the umbra. In the full shadow no part of the Sun's surface can be seen. The lesser dark part of Earth's shadow (half shadow) is called the penumbra. In the half shadow part of the Sun's surface can be seen. The difference between these two shadows is so big, that when the Moon only travels through the penumbra, it looks only a little dimmer, and consequently it is easy to miss the eclipse all together. This type of lunar eclipse is called a penumbral eclipse.

A partial lunar eclipse is when the Moon travels into the full shadow of the Earth. When the Moon enters the umbra it looks like a bite has been taken out of the Moon. In a partial eclipse the Moon will be visible throughout. On the other hand a total lunar eclipse is when the Moon travels completely into the umbra. In such an eclipse the Moon will seem to disappear.

When a lunar eclipse appear red, it is called a blood Moon -not a scientific term. What happens is that the light illuminating the Moon is predominately red. When sunlight, which is white and thus contains all colours, travels through the Earth's atmosphere it is filtered and reflected in such a way, that the green to violet light on the spectrum scatter more strongly than the red light. With the other colours in the sunlight gone, the remaining red light is what shines on the Moon.

A solar eclipse occurs when the Earth passes through the Moon's shadow. From Earth this looks like the Moon moves in front of the Sun, but in reality we are moving into the Moon's shadow. A solar eclipse can only happen when the Sun, Moon, and Earth are aligned. However, the Moon must be between the Earth and the Sun for a solar eclipse. We have from the phases of the Moon in Figure 1.2 that solar eclipses only happen when the Moon is in its new moon phase (i.e. not visible).

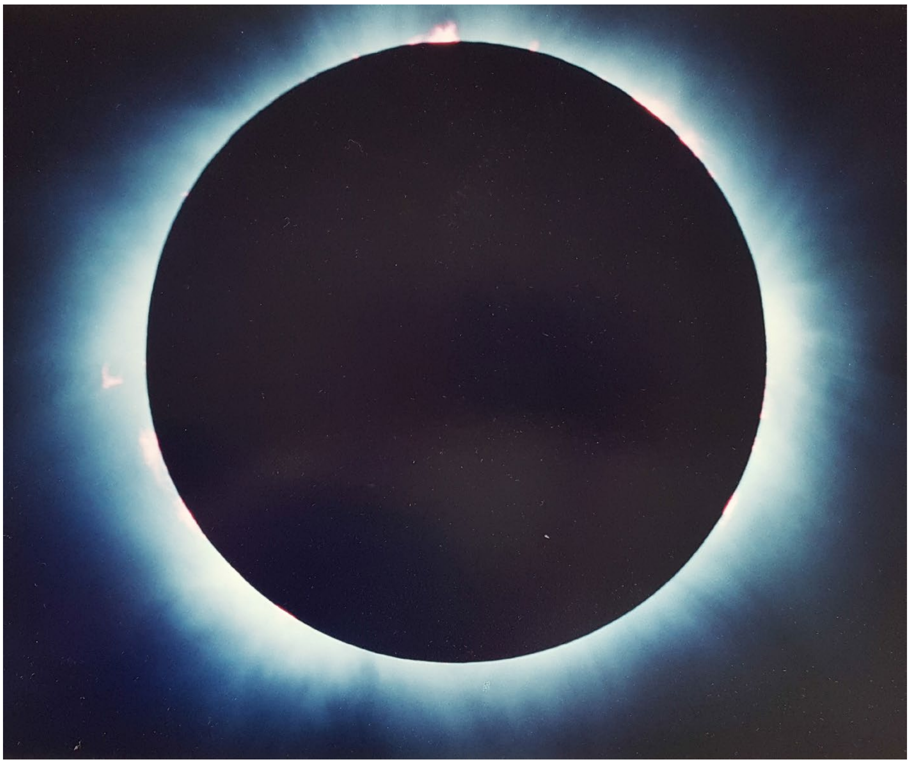


Figure 1.3: Solar eclipse 1999, picture courtesy of Dr. Michael Bauer.

There are at least two solar eclipses per year, with a maximum of five. It changes where on Earth it is possible to see an eclipse. A solar eclipse is very short compared with a lunar eclipse. That is because the shadow cast by the Moon is much smaller than the shadow cast by the Earth.

A total solar eclipse is when the Moon completely covers the Sun. A partial solar eclipse is when the Sun is visible throughout the eclipse. A partial eclipse looks like bites are bitten off the Sun. The Moon passing in from of the Sun remains invisible, as it is in the new moon phase. Your position of Earth determines if you will see a total eclipse or a partial eclipse. The shadow of the Moon that we travel into is so small that often travelling just 1,000 km on Earth's surface can change the experience from a partial eclipse to a total eclipse. Figure 1.3 show a picture from the 1999 total solar eclipse in central Europa.

The Moon's orbit is also an ellipse, meaning that the Moon will appear larger when it is closest to us and smaller when it is further away. This change in apparent size of the Moon is very significant in total solar eclipses as the Moon is not always large enough to cover the Sun completely. A total solar eclipse, where the Moon overlays the Sun, so only the rim of the Sun is visible is called an annular eclipse. That happens when the Moon is too far away from Earth for its umbra to reach us. As the Moon's distance to us is ever increasing, total solar eclipses will some day not happen any more.

Whenever viewing a solar eclipse you have to be cautious, as the Sun's diminished brightness relaxes your intuitive reluctance to look directly at the Sun. The Sun, however, might still be bright enough to blind you.

1.5 BASIC CELESTIAL MECHANICS

Celestial Mechanics is what we call the laws of motion that govern the movements of planets and moons.

1.5.1 KEPLER'S LAWS

We start our understanding of celestial mechanics by examining Kepler's laws:

- Kepler's First Law: A planet orbits the Sun in an ellipse, with the Sun at one focus of the ellipse.

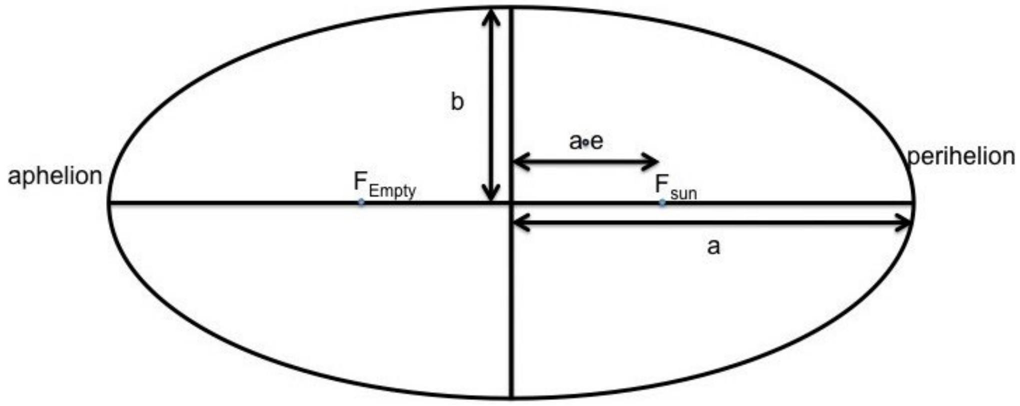


Figure 1.4: An ellipse.

  An ellipse is a squeezed circle, with the longest axis (a line drawn through the centre) being 2×a long, where a is called the semimajor axis. Similarly, b is called the semiminor axis, and an ellipse's shorter axis is 2×b long. See Figure 1.4. The eccentricity of an ellipse, i.e. how oval it is, is denoted as e, which essentially is a number between 0 and 1. An ellipse has two focal points, both situated on the semimajor axis, and both with distance from centre of a × e. Kepler's first law states that the Sun is situated in one of these focal points, it is implied that the other focal point is just empty space. A circle is a special case of an ellipse, where e= 0. Most of the planetary orbits in our solar system are very circular ellipses, with eccentricities between $0.01 < e < 0.24$. The position on a planets orbit that comes the closest to the Sun is called perihelion, and the opposite point on the orbit, i.e. where the planet is the farthest from the Sun, is called aphelion.
- Kepler's Second Law: A line connecting a planet to the Sun sweeps out equal areas in equal time intervals.

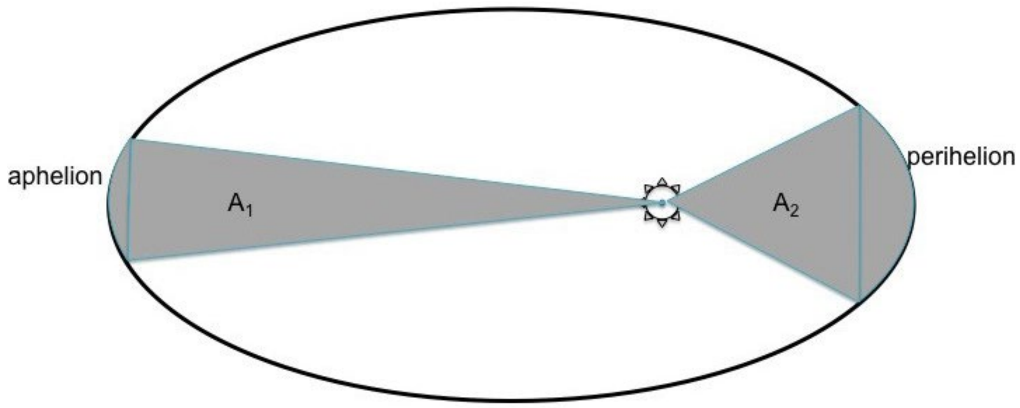
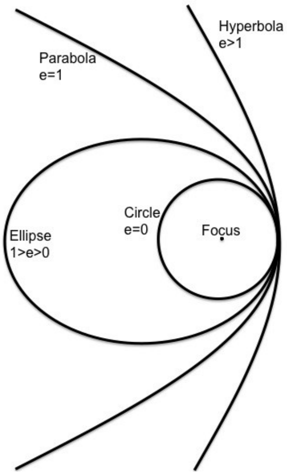


Figure 1.5: Kepler's Second law.

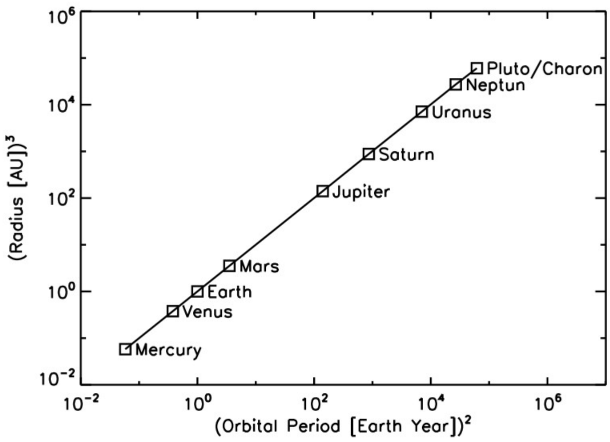
In other words, the orbital speed of a planet depends on its location in that orbit. An illustration of Kepler's second law is shown in Figure 1.5, where the areas A1 and A2 are of equal size. The law states that it takes equal amounts of time for the planet to sweep out equal areas. Consequently, the areas A_1 and A_2 are swept out by the planet in equal amounts of time. In an ellipse the area swept out by the planet close to the Sun is short and wide and the area swept out when the planet is further from the Sun is long and narrow. However, the path traced out by the planet on the elliptical orbit in aphelion is much shorter than that in perihelion. When the planet traces out different path lengths in the same time, its orbital speed must be changing. The orbital speed is larger in perihelion, where the path that is travelled is larger.

That the planet is moving faster in perihelion can also be understood intuitively, by realising that an object closer to the Sun is accelerated by the Sun's gravitational pull, and thereby speeding up when it is closer.

All orbital motions in our Solar system can be described with basis in the ellipse. The planets all have elliptical orbits, so do the recurrent comets, however with a very large eccentricity. Some comets' movements describe that of a parabola ($e = 1$) and some even that of a hyperbola ($e > 1$). The hyperbolic orbit features only one passage close to the Sun. In that process the object can be accelerated away from the Sun, which is sometimes called 'slingshot'. Figure 1.6a shows the different orbital motions.



(a) Orbital motions.



(b) Kepler's Third law.

Figure 1.6: Illustrations to orbital motion and Kepler's third laws.

Kepler's third and last law is also called the harmonic law. The law relates the orbital period, T, in Earth years, to the average orbital distance, r:

- Kepler's Third Law: $T^2 = r^3$,
  where r is in Astronomical Units (AU). 1 AU is defined as the distance between the Sun and the Earth. Kepler found his third law from meticulous observations made by Tycho Brahe (1546-1601). Figure 1.6b shows Kepler's third law for all the planets in our solar system. This law, together with Newton's three laws lead to a generalised expression of the universal law of gravity, which Newton published in 1687. It is noteworthy that Kepler made his realisation of his third law without knowing of the concept of gravity. Newton who lived from 1642-1727 never met Kepler, who lived from 1571-1630.

1.5.2 NEWTON'S LAWS

Newton's three laws are:

- Newton's First Law:
  The Law of Inertia: An object at rest will remain at rest and an object in motion will remain in motion in a a straight line at a constant speed unless acted upon by an unbalanced force.
- Newton's Second Law:
  The net force (the sum of all forces, F) acting on an object is proportional to the object's mass, m, and its resulting acceleration, a. $F = m \times a$
- Newton's Third Law:
  For every action there is an equal and opposite reaction.

Newton's great insight was that the force that keeps the Moon in its orbit, is the same as that which makes an apple fall to the ground. Since the planets are not moving at constant speed along straight lines, they must be under the influence of a force. Newton realised that the acceleration that the planets have in their orbits around the Sun could be calculated using Kepler's laws. In order to calculate this Newton generalised Kepler's laws, so they could describe any orbital system.

First we assume that the orbits are all circular (special case of Kepler's first law where e=0). Kepler's second law (the law of areas) demands that the speed of the planet in the orbit is constant. Remember we are now looking at circular orbits. Using vector analysis (beyond the scope of this book) we arrive at the acceleration, a, directed towards the Sun (the centre of the orbit) with the magnitude:

$$\mathrm{a} = \frac{\mathrm{v}^2}{\mathrm{r}}, \tag{1.1}$$

where v is the velocity (which is constant because it is moving in a circle and not an ellipse) and r, the radius.

In order to derive Newton's general law of gravity, we need to examine circular orbits further. The circumpherence of a circle, the path a planet has to travel to make it once all the way around, is 2πr, where r is the radius of the circle. The time, T, it takes a planet to complete an orbit is given by:

$$\mathrm{T} = \frac{\mathrm{length}}{\mathrm{speed}} = \frac{2\pi \mathrm{r}}{\mathrm{v}} \tag{1.2}$$

Inserting that into 1.1, we get:

$$\mathrm{a} = \frac{\mathrm{v}^2}{\mathrm{r}} = \frac{(\frac{2\pi \mathrm{r}}{\mathrm{T}})^2}{\mathrm{r}} = \frac{4\pi^2 \mathrm{r}^2}{\mathrm{T}^2 \times \mathrm{r}} \tag{1.3}$$

Kepler's third law can be generalised to hold true for any planetary system introducing the constant, C. The law is then $T^2 \times C = r^3$, which inserted into the above equation gives:

$$a = \frac{4\pi^2 r^2}{T^2 \times r} = \frac{4\pi^2 r^2 \times C}{r^3 \times r} = \frac{4\pi^2 \times C}{r^2} \tag{1.4}$$

Newton's second law, F = m × a, then looks like this:

$$\mathrm{F} = \mathrm{m} \times \mathrm{a} = \mathrm{m}\frac{4\pi^2 \times \mathrm{C}}{\mathrm{r}^2}. \tag{1.5}$$

The constant C may appear mysterious. C is really just a constant that changes value depending on the orbital system in question and takes the mass of the central object into account. We would like to have the mass of the central object, M, directly in the equation, so we let $C = M \times G/4\pi^2$, where G is a universal gravitational constant. Newton's law of attraction between two masses, m and M, then assumes this neat form:

$$F = G\frac{mM}{r^2}, \tag{1.6}$$

where G is a universal constant that only depends on choice of units. This formula holds true for any two masses. It implies that Earth is also pulling on the Sun, just no where near as strongly as the Sun is pulling on us.

1.5.3 ESCAPE VELOCITY

Intuitively we understand that an object needs a certain velocity in order to escape the gravitational field of a planet. Too slow and whatever we were throwing up in the air would come back down.

Imagine a child throwing a ball straight up into the air. The ball would start out with a velocity, which would get smaller and smaller as the ball feels the pull of gravity from the Earth. The initial velocity would then have been the maximum velocity. The ball continues up as far as it can get, while it is slowing down. When it is at its maximum height, its maximum distance from the ground, it comes to a halt. Now its velocity is zero. It then starts to fall, speeding up, accelerating, until it lands safely in the hands of the child once again. Just before it is caught again it will have gained its maximum velocity again. This is one of the reasons, why it is a bad idea to fire a gun straight into the air; when the bullet reaches you again, it is almost as fast as when it left the barrel of the gun, since the air resistance is not enough to slow it down significantly.

When we examine this ball game in terms of energy, we only have to consider two kinds of energy for the ball. The kinetic energy, E_{kin}, that relates to the velocity, v, of the ball and m, its mass:

$$E_{kin} = \frac{1}{2} m \times v^2, \tag{1.7}$$

and the potential energy, E_{pot}, that relates the mass of the ball to its distance, r, from the centre of the planet:

$$E_{pot} = -\frac{G \times M \times m}{r}, \tag{1.8}$$

where G is the universal gravitational constant, M the mass of the planet, and m the mass of the ball.

There are a number of conservation laws in nature. The conservation of energy means that the total energy of a system remains constant. The energy in a system may change from one kind of energy (chemical energy, thermal energy, radiation energy etc.) to another. The energy, however, cannot just disappear or pop into existence. In our case with the ball it starts out with a large kinetic energy, only to have that energy be 0 when it has lost all movement at its maximum distance from the child. What the ball has lost in kinetic energy it has gained in potential energy.

The ball, at rest at infinite distance ($r = \infty$) must have a total mechanical energy of zero. Since the total energy is constant it must be zero everywhere.

$$E_{total} = E_{kin} + E_{pot} = 0 \tag{1.9}$$

Inserting the expressions for the energies above:

$$\frac{1}{2}m \times v^2 - \frac{G \times M \times m}{r} = 0 \tag{1.10}$$

Since we want to find the escape velocity of the ball, we substitute v = V_{escape}. The ball with mass m is escaping from the surface of a planet with radius r=R and mass M. Inserting that in the previous expression and we get:

$$\frac{1}{2}m \times V^2_{escape} = \frac{G \times M \times m}{R} \tag{1.11}$$

And the escape velocity is:

$$V_{escape} = \sqrt{\frac{2G \times M}{R}} \tag{1.12}$$

This is a general formula and with it we can find the escape velocity for the Moon, Earth, Sun, Jupiter, Mars and so on. The escape velocity does not depend on the mass of the object you are trying to shoot into space. This might sound counterintuitive, but when you consider giving a small tennis ball a certain speed and you try to give a huge space shuttle the same specific speed, you immediately realise that the energy you need to spend to achieve this will be very different for the shuttle and for the ball.

When we insert the known values for Earth we get an escape velocity of:

$$V_{escape} = 11.2 km/s \tag{1.13}$$

And for the Moon:

$$V_{escape} = 2.38 km/s \tag{1.14}$$

Bibliography and Links

The physical laws described in this chapter are all very old and references to the original papers by Newton and Kepler would not be of much use to the reader. Any physics textbook at university introductory level will have at least some of the same formulae. For additional material I refer the reader to read up on conservation laws, as that is the very heart of physics.

The National Aeronautics and Space Administration (NASA) has a great website with details for upcoming astronomical events: http://eclipse.gsfc.nasa.gov/eclipse.html

Another nice eclipse website is: http://www.timeanddate.com/eclipse/

2 THE SOLAR SYSTEM

Your goals for this chapter are to learn about:

- The formation of the solar system.
- The composition of the solar system.
- The definition of a planet and a dwarf planet.
- The 5 known dwarf planets.
- Astroids, Comets, Meteors, Meteoroids and Meteorites.

2.1 FORMATION OF THE SOLAR SYSTEM

The solar system was formed from a large cloud of dust and gas, which contracted under its own gravity. In the contraction potential energy for the material was converted into kinetic energy and thus the temperature increased. The initial temperature was 10-50 K. The contraction continued until the density in the centre was large enough that the temperature could reach a critical ignition point, where the Sun turned on.

Because of the conservation of angular momentum the coalescing material settled into a rotating disk around the Sun. Such a disk is called a proto-planetary disk, and from this the planets formed. Dust particles in the disk formed larger particles after colliding and sticking to other dust particles. Particles clumped together into larger chunks, which again bound to other chunks. Eventually the chunks became the size of planets, which then were so large that they would accrete dust and gas in their orbits, see Figure 2.1.

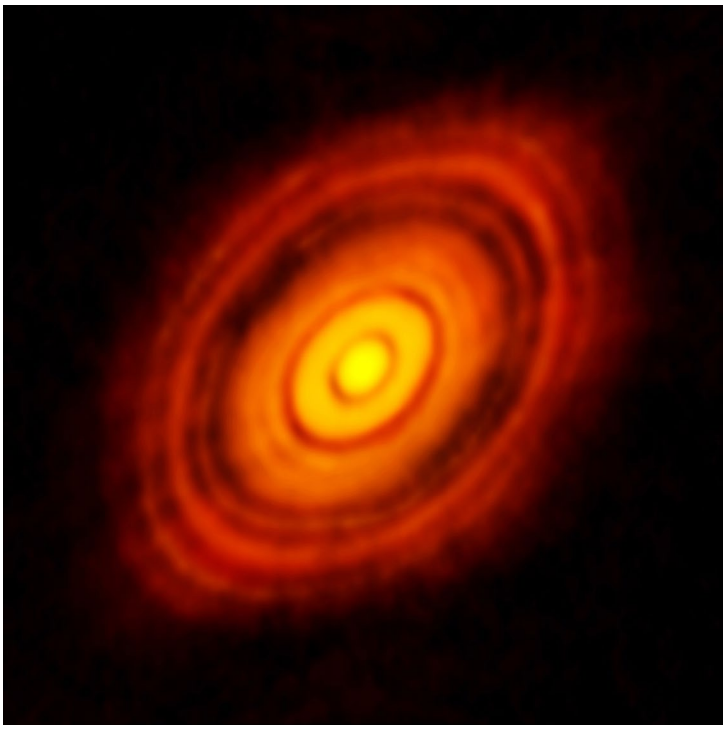


Figure 2.1: Image of the young star HL Tauri located about 450 light-years away. The surrounding dusty disk has gaps, where planets are being formed. Image from ALMA courtesy of ESO.

The newly formed Sun had a strong solar wind, which pushed any extra gasses away. At this point the planets had no more material to accrete and the planet formation was over. All the planets orbit the Sun in the same direction as the Sun rotates, and despite a few exceptions they all rotate in the same direction.

The different composition of the planets can be explained by the temperature distribution in the proto-planetary cloud. Towards the centre the temperature was much higher than further out. A material changes phase, when the temperature changes. From the familiar water we know the solid state ice, the liquid state water, and the gaseous state steam. For each element and chemical composition the temperature for the phase change is different. The inner planets have an overabundance of the heavier elements, because the temperature at which they change into solids is relatively high. The lighter elements could not condense close to the Sun and remained in the volatile gaseous state and thus evaporated.

Soon after the planets formed they went through an intense meteorite bombardment, in literature called the Late Heavy Bombardment. This was an important source of internal heat, which, together with heat from shortlived radioactive isotopes, caused the planets to melt. Before the melting the planets were relatively homogenous, but, when melted, the heavier material sank to the centre. The planets thus became differentiated, as the material settled into layers of different chemical composition. The timescales for planetary formation are very long and almost incomprehensible for us. The age of the solar system is about 4×10^9 years, and the oldest rocks on Earth are 3.7×10^9 years old.
From the beginning of the Sun's formation to the first traces of life about 800 million years passed. The timeline is summarised in Table 2.1.

Era	Time [years]	Description
Pre-solar Nebula Era	0.0	Collapse of cloud to form flattened disk
Astroid Era	3 million	Formation of large astroids up to 200 km across ends
Gas Giant Era	10 million	Rapid formation of Jupiter and Saturn ends
Solar Birth Era	50 million	Sun's nuclear reactions start to produce energy in the core
Planetessimal Era	51 million	Formation of numerous small planet-sized bodies ends
T-Tauri Era	80 million	Solar winds sweep through inner solar system and strip off primordial atmospheres
Ice Giant Era	90 million	Formation of Uranus and Neptune
Rocky Planet Era	100 million	Formation of rocky planets by mergers of 50-100 smaller bodies
Late Heavy Bombardment	600 million	Migration of Jupiter disrupts asteroid belt sending large asteroids to impact planetary surfaces in the inner solar system
Ocean Era	600 million	Comets rich in water impact on Earth and form the oceans
Life Era	800 million	First traces of life found in fossils on Earth date back to this time

Table 2.1: Solar System Formation Timeline. Courtesy of NASA's Space Math.

The topic of solar system formation has previously been explored using computer models, where, by inputting physics and different starting conditions,possible outcomes are analysed and evaluated against observations in nature. With the arrival of large sub-mm arrays (very large radio telescopes), which can see through dust and gas surrounding newborn stars, direct observations of planetary formation are now possible, see Figure 2.1. The research into formation of planetary systems has, with the advancement of telescopes, become a new and exciting frontier in astronomy.

Planetary formation is a slow process, so we cannot just sit back and watch, instead astronomers benefit from the vastness of space, and observe as many systems as possible. The systems will be at different stages in their formation, and combining them we get a coherent image.

2.2 COMPOSITION OF THE SOLAR SYSTEM

The solar system consists of a central star called the Sun, 8 planets, 5 dwarf planets, 173 moons, 3,319 comets, and 670,542 asteroids. The planets can be divided into three physically distinct groups, the rocky terrestrial planets, the gaseous Jovian planets, and the smaller dwarf planets. All motions in the solar system are governed by gravitation.

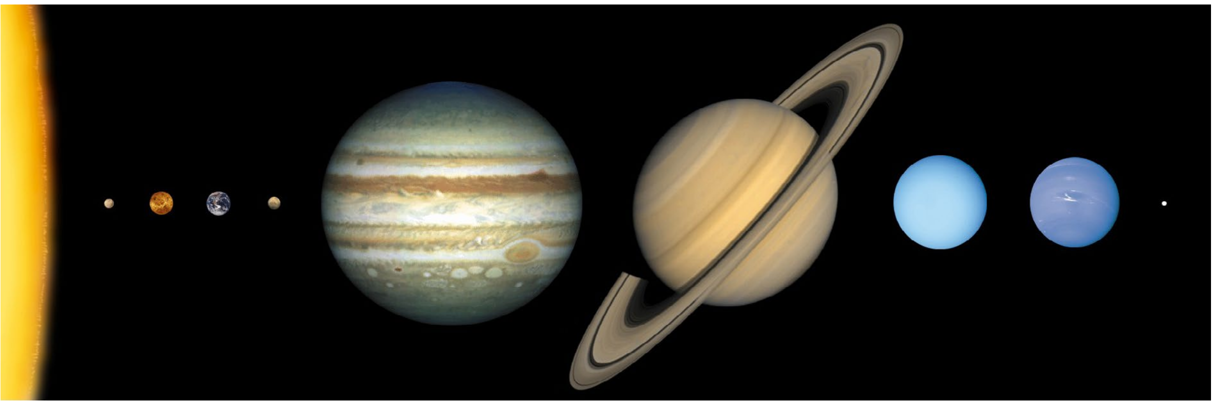


Figure 2.2: This illustration shows the approximate sizes of the planets relative to each other. Outward from the Sun, the planets are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune, followed by the dwarf planet Pluto. The planets are not shown at the appropriate distance from the Sun. Courtesy of NASA.

2.2.1 THE TERRESTRIAL PLANETS

We call the inner 4 planets (Mercury, Venus, Earth, and Mars) terrestrial planets because they all have things in common with Earth (Terra in Latin).

The terrestrial planets have a solid surface and are relatively dense. They all have an iron-nickel core surrounded by a mantle consisting of silicates (i.e. compounds of silicon). Their average densities are between 3,500 and 5,500 kg/m^3, which implies that a piece of terrestrial rock would sink -water's density is 1,000 kg/m^3. Read more about the terrestrial planets in Chapter 3.

2.2.2 THE ASTEROID BELT

A region between Mars and Jupiter, between 2 AU and 3.5 AU from the Sun is filled with asteroids. This region is called the asteroid belt. Asteroids are basically large rocky lumps. The asteroid belt contain thousands of asteroids in many sizes, who all orbit the Sun. Minor bodies, like these rocky asteroids, remain irregular in shape as long as they are fairly small. Bodies larger than 1,000 km in diameter are deformed into a symmetric shape, like a sphere, due to gravity. The largest body in the asteroid belt has a diameter of 1,000 km and is called Ceres. Ceres is spherical and has been classified as a dwarf planet. The asteroids were formed simultaneously with the planets, however, they never merged together to form a single planet. Due to collisions the initially larger chunks broke up into smaller lumps and the asteroid belt we see today is the debris.

2.2.3 THE JOVIAN PLANETS

The 4 outer planets (Jupiter, Saturn, Uranus, and Neptune) are called the Jovian planets because of their resemblance with Jupiter. The Jovian plan- ets are much larger than the terrestrial planets and their density distinctly lower. Most of the volume in a giant planet is hydrogen and helium, and a piece of average Saturn (density of 700 kg/m3) would float. Jovian planets have relatively small silicate (rocky) cores surrounded by a large extended atmosphere of gas. The Jovian planets are also divided into two subcate gories, where Jupiter and Saturn are gas giants and Uranus and Neptune are ice giants. Read more about the jovian planets in Chapter 4.

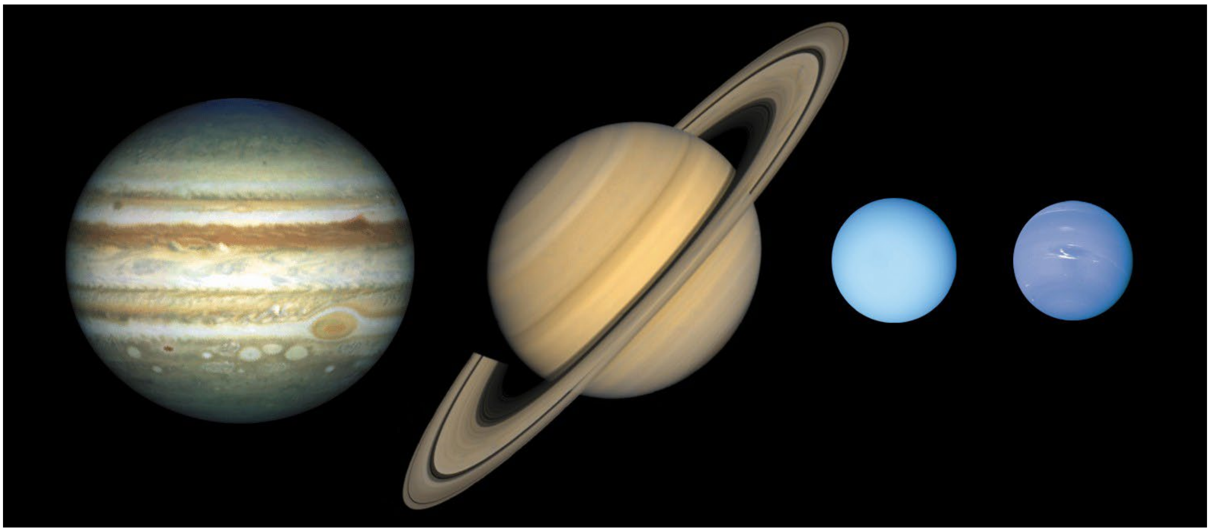


Figure 2.3: Jupiter, Saturn, Uranus, and Neptune are known as the jovian planets. Here the giants are pictured next to each other with their relative sizes. Image courtesy of the Lunar and Planetary Institute.

2.2.4 KUIPER BELT

The Kuiper Belt is a disc-shaped region of icy bodies, which extends from about 30 to 55 AU, i.e. just beyond Neptune's orbit. The Kuiper belt houses 4 confirmed dwarf planets (Pluto, Haumea, Makemake, and Eris) and most likely many more, as there are thought to be thousands of icy bodies larger than 100 km across. It is estimated that there are a trillion or more comets in the Kuiper belt, and the short period comets, which have an orbit of less than 200 years, probably originate there.

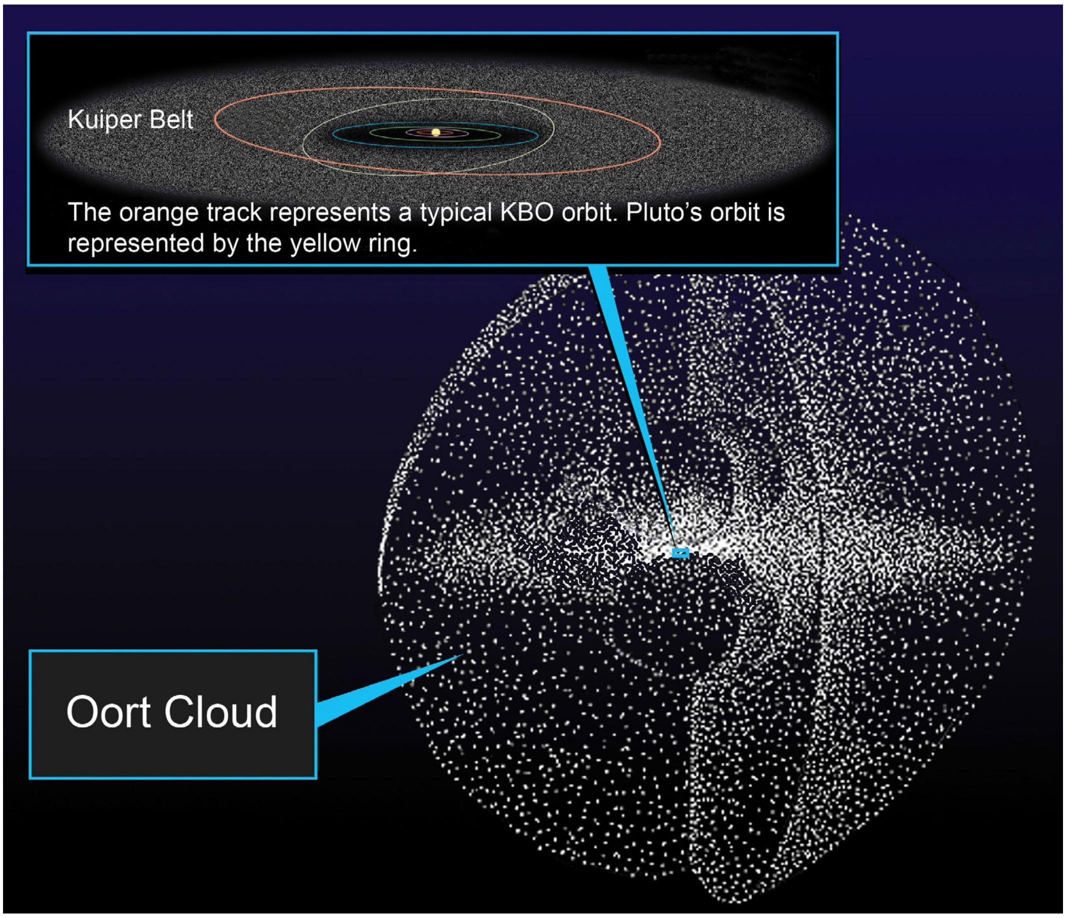


Figure 2.4: An illustration of the Kuiper Belt and Oort Cloud in relation to our solar system. KBO=Kuiper Belt Object. Image courtesy of NASA.

Currently the spacecraft New Horizons is exploring the Kuiper belt and the dwarf planets residing therein. It took the vessel 9.6 years to travel there from Earth.

2.2.5 OORT CLOUD

The Oort Cloud is an almost spherically symmetric shell enveloping our solar system. The cloud consists of a swarm of icy bodies, probably numbering 0.1-2 trillion, all orbiting the Sun. The Oort cloud is in a region between 5,000 and 100,000 AU from the Sun, and thus at the boundary of the solar system. The outer part of the Oort Cloud is where the Sun's gravitational influence can be overpowered by that of other stars. Occasionally, stars passing nearby disturb the orbit of one of the ice bodies in the Oort Cloud, causing it to travel into the inner solar system, where we see it as a long period comet. These comets have very large, eccentric orbits and are observed in the inner solar system only once. No objects still residing within the Oort Cloud have ever been directly observed.

The objects in the Oort Cloud and in the Kuiper Belt are presumed to be remnants from the formation of the solar system.

2.2.6 HELIOSPHERE

The Heliosphere is the very edge of solar system. Here is the boundary between the Sun's magnetic realm of influence and interstellar space. The solar wind consisting of charged particles fills the space between the planets. This wind creates a magnetic bubble that shields the solar system from most of cosmic rays.

NASA's Voyager 1 has now entered interstellar space at a distance of 130 AU. Voyager 2 is located at 110 AU, which puts it in the region, where the solar wind slows and compresses as it interacts with the interstellar gas. This region is called the heliosheath, and it is the last major boundary before interstellar space.

2.3 DEFINITION OF A PLANET

Astronomy, like most other sciences, relies on international cooperation. Astronomers are organised in The International Astronomical Union (IAU), which was founded in 1919. Its mission is to promote and safeguard the science of astronomy in all its aspects through international cooperation. In 2006 astronomers, through the IAU, agreed upon a definition system for planetary objects.

August 24, 2006, Prague: IAU News Release - IAU0603:
The IAU resolves that planets and other bodies, except satellites, in the Solar System be defined into three distinct categories in the following way:

1) A planet is a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (c) has cleared the neighbourhood around its orbit.
2) A "dwarf planet" is a celestial body that (a) is in orbit around the Sun, (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, (c) has not cleared the neighbourhood around its orbit, and (d) is not a satellite.
3) All other objects, except satellites, orbiting the Sun shall be referred to collectively as "Small Solar System Bodies".

A satellite is an object that orbits a planet or a star. Moons are often referred to as "natural satellites", whereas the manmade objects, which sup- port our telecommunication, are called "artificial satellites".

2.4 DWARF PLANETS

Currently the IAU recognises five dwarf planets. Ceres in the astroid belt. Pluto, Haumea, Makemake, and Eris in the Kuiper Belt. There are likely many more dwarf planets in the Kuiper Belt. The dwarf planets have similar compositions and are solid with icy surfaces. Currently NASA is leading the New Horizons mission, which after having visited Pluto and its moons will explore deeper into the Kuiper Belt.

2.4.1 CERES

With a diameter of 1,000 km, Ceres is the largest body in the astroid belt and the only one to be identified as a dwarf planet. Had Ceres not been round it would have been classified as an astroid and not a dwarf planet. Ceres orbits the Sun, but it is not a planet as its orbit (the astroid belt) is shared with thousands of other objects. Ceres was discovered in 1801 by an italian monk Giuseppe Piazzi, and it contains about 30% of the combined mass of the asteroid belt.

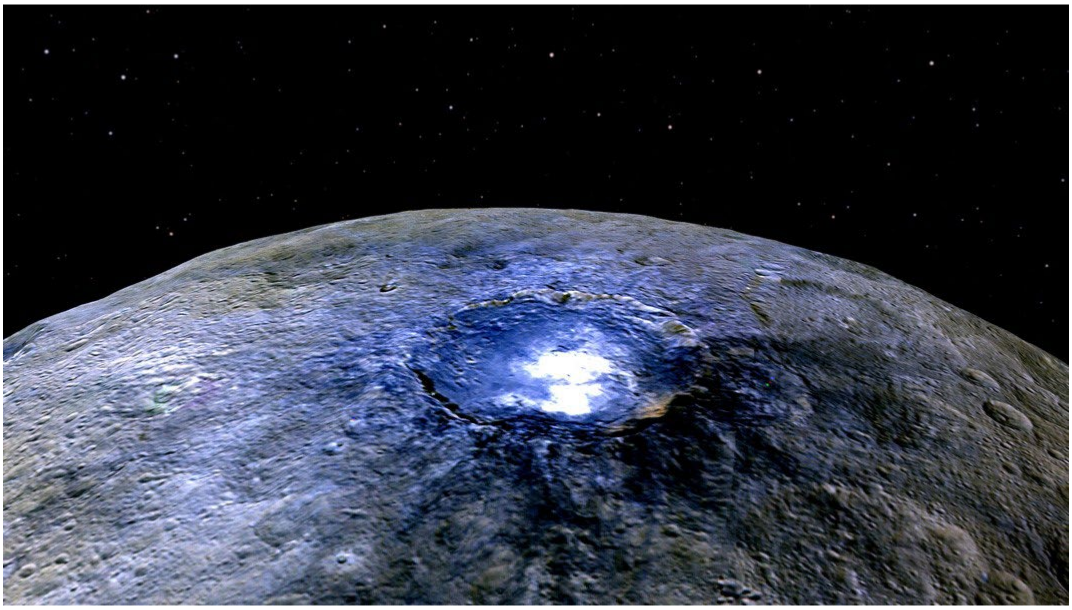


Figure 2.5: Dawn mission image of Ceres and its white spots. Image Credit: NASA/JPL-Caltech/UCLA/MPS/DLR/IDA

Ceres has many bright white spots (see Fig. 2.5), one of which recently was investigated in detail by the Dawn spacecraft (2015-2018 by NASA).

It turns out that the spots are places where the planet is evaporating and creating a haze. The bright spots are associated with craters, so it is likely that impacts have scraped off surface material exposing a more volatile sub-surface layer. The presence of ammonia salts in the the haze raises doubts about the formation of Ceres, as we would not expect ammonia in the composition of a body that close to the Sun.

2.4.2 PLUTO

Pluto was from its discovery in 1930 up until the IAU definition in 2006 considered to be a planet. Its 'demotion' to the status of a dwarf planet has not been popular amongst the general public. However, as we investigate deeper into the Kuiper Belt -the location of Pluto, we have to recognise that Pluto is not unique. Pluto's orbit around the Sun is highly eccentric, bringing it briefly inside Neptune's orbit. It takes Pluto 248 Earth years to make one orbit around the Sun, and out of that only 20 Earth years are spent inside Neptune's orbit. Pluto's eccentric orbit brings it as close as 29.3 AU to the Sun and up to 49.3 AU away.

Pluto's radius is about 66% of the Moon's radius, but weights only 16% of the Moon. This low density is the result of a rocky core and an icy mantle. Where the mantle is probably made of water ice, its surface is coated in methane and nitrogen ice. Pluto has a terrain of mountains 2-3 km high and large craters with diameters up to 260 km. Pluto's atmosphere, which consists of nitrogen, methane, and carbon monoxide, contracts and falls as snow on the surface when Pluto is furthest away from the Sun.

One full rotation of Pluto takes 6.4 Earth days. That is the exact same time it takes Charon, Pluto's largest moon, to make one orbit around Pluto. The Pluto-Charon orbital system is double tidally locked, so that Charon is always showing the same side to Pluto and Pluto always showing the same side to Charon. This implies that Charon is only ever visible from one half of Pluto. The Earth-Moon system is only partially tidally locked, i.e. the Moon is always showing Earth the same side, but the Moon sees all of Earth in the course of its orbit. Because Charon is a somewhat large moon compared to its parent planet, it is half the size of Pluto, the two together are often referred to as a double planet system. The distance between them is a mere 19,640 km.

Pluto has 5 known moons: Charon, Nix, Kerberos, Hydra, and Styx, listed in increasing distance from Pluto. They are thought to have formed from a large collision between Pluto and a similar sized body early in the formation of the solar system. Pluto and its moons has recently been visited by the New Horizon mission's spacecraft, which has given us spectacular images of the surface. The white and almost heart shaped area on Pluto has been named 'Sputnik Planitia' after the first space satellite in 1957. Figure 2.6 shows an image of Pluto taken by New Horizon.

2.4.3 HAUMEA

Haumea is also located in the Kuiper belt and thus orbits the Sun once every 285 Earth years. It is almost egg shaped because of its fast rotation-once every 4 hours. There are speculations about Haumea having had an early large impact, giving it its fast spin and also its two moons Hi'iaka and Namaka. Haumea is likely made of rock with a coating of ice. It was discovered in 2003 and its moons in 2005.

Figure 2.6: A new Horizon image of Pluto. Image courtesy of NASA.

2.4.4 MAKEMAKE

Makemake is also located in the Kuiper Belt and orbits the Sun once every 310 Earth years. It was discovered in 2005 and it is slightly smaller than Pluto. Its surface shows signs of frozen nitrogen, methane and ethane.

2.4.5 ERIS

Eris has an unusual orbit, as its orbital plane is highly inclined with respect to the ecliptic, where all the other planetary bodies are located. Its distance to the Sun is at times further out than the Kuiper Belt, making one orbit take 557 Earth years. Eris has a small moon called Dysnomia, which orbits Eris every 16 days. Eris is smaller than the Moon, and maybe even smaller than Pluto.

2.5 SMALL SOLAR SYSTEM BODIES

Although there are smaller bodies all over the solar system, most of them are located in either the asteroid belt, the Kuiper belt or the Oort cloud. Asteroids, meteoroids, and meteorites all belong to the category of small solar system bodies. The terminology for these objects is a bit messy as there are no clear borders between the different kinds of objects. Some asteroids might have the same composition and structure as a comet and some near Earth asteroid might be a remnant of a comet.

2.5.1 METEOROIDS, METEORS, AND METEORITES

Solid bodies smaller than asteroids are called meteoroids. However, there is no defining size where an object would be either one or the other. Some meteoroids are even debris from asteroid collisions.
Meteoroids are either made of ordinary stone, the more fragile carbon stone, or of cometary material such as ice and snow.

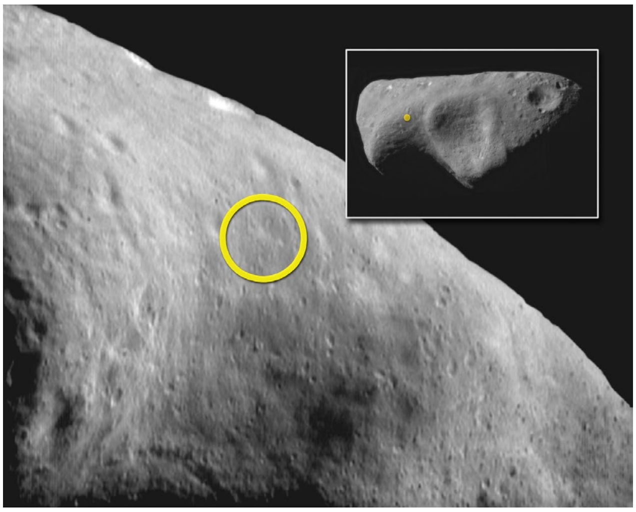
If a meteoroid enters Earth's atmosphere it will glow due to the ram pressure of the air. Essentially, a shock wave is generated in front of the meteor by the extremely rapid compression of air. This compressed air in the shock wave is heated up, and that then heats up the meteor, flowing around it. When heated up the meteor glows and becomes visible to us.
When we see a meteoroid trailing a glowing streak across the sky we call it a meteor or a 'shooting star'. Meteors become visible between about 75 to 120 km above the Earth's surface and they usually disintegrate at altitudes of 50 to 95 km.

A comet passing Earth's orbit on its path towards the Sun can leave be hind a trail of debris as it gets heated up. When Earth passes into this debris we will have a meteor shower.
If a meteoroid survives being a meteor and lands on the ground, the remaining piece of rock is called a meteorite. The more fragile meteoroids do not survive the passage through Earth's atmosphere, which explains why we encounter disproportionally many iron rich

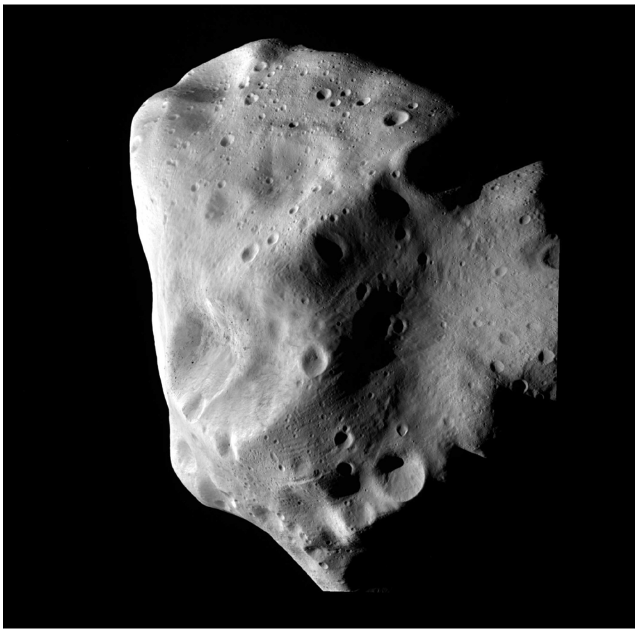
meteorites and none out of snow and ice. Meteorites can be divided into 3 distinct groups: Iron meteorites, stone meteorites, and stone-iron meteorites. By analysing the composition and structure of meteorites we can learn about the conditions under which they were formed.

2.5.2 ASTEROIDS

Asteroids are rocky bodies orbiting the Sun and most of them, by far, are located in the asteroid belt. Outside of the asteroid belt, associations or families of asteroids are found. Most significantly are the Trojans, which share orbit with Jupiter. Earth is surrounded by asteroids (called Near Earth Asteroids), of which about 500-1,000 are larger than 1 km in diameter. The asteroids have crater filled bodies with a fine dust of pulverised rock on their surfaces, see Figure 2.7. Some are two merged into one, some look like one body, but is really two bodies in close contact. They all originate from the very beginning of the solar system.



(a) Landing site on Eros, courtesy of NASA/JPL/JHUAPL.



(b) Asteroid 21 Lutetia, courtesy of ESA.

Figure 2.7: Pictures of asteroids showing their odd shapes and cratered surfaces.

Asteroids can be divided into 3 main types according to their composition:

- C-type: C asteroids are carbon rich and appear grey. The material is the same throughout and they belong to the most primordial bodies of the solar system. They resemble stony meteorites.
- S-type: S astroids are silicon rich and appear grey with a green or red shine. They are made of silicate materials and iron-nickel. They resemble stone-iron meteorites.
- M-type: M Asteroids consist mainly of iron and nickel and they look reddish.They have undergone at least a partial differentiation as evident by an internal layering of different chemical composition. M- type asteroids resemble iron meteorites.

In 2001 NASA's spacecraft NEAR became the first space probe to land on an asteroid. It landed on the near earth asteroid Eros after having studied it for a year, Figure 2.7a shows a picture of Eros and the landing site. Japan's space agency (JAXA) made history in 2006 with their space probe Hayabusa, as it was the first spacecraft to land and take off from an asteroid. Furthermore, it returned home to Earth with samples from the asteroid.

2.5.3 COMETS

Comets have been observed since ancient times, their passage often documented in paintings and literature. Comets orbit the Sun in very elliptical orbits, and become visible to us when they travel closer than 2 AU to the Sun. The Sun heats up the comet, which then gets a spectacular long tail as it evaporates. This tail will always point away from the Sun, because of the radiation pressure and the solar wind. When a comet approaches the Sun the tail is behind it, and when the comet is flying away from the Sun again, it is flying into its own tail.

Comets are about 10 km in diameter or less and they consists of ice, snow, and dust. The nucleus contain icy chunks and frozen gasses with embedded rock and dust, which at the centre might have a small rocky core. When the Sun heats up the comet, the evaporated gasses and dust create an envelope around the nucleus, which is called the coma. The radiation pressure and the solar wind push dust and ionised gas away from the comet creating tails. A comet has two tails, one out of dust -which is the brightest, and one out of ionised gasses. The dust tail is bright because it reflects sunlight, whereas the ion/gas tail's glow comes from excited atoms.

Figure 2.8: Comet C/2014 Q1 with its two tails and the Moon, courtesy of Dr. Yuri Beletsky of www.facebook.com/yuribeletskyphoto, who also provided the cover image for the book.

Some comets have relatively short orbital periods, like the famous Halley's comet recurring every 75 years. The short period comets (i.e. with orbits less than 200 Earth years) originate in the Kuiper belt, and the long-period comets originate in the Oort cloud.

In 1994 scientist all over the world combined efforts to observe the comet Shoemaker-Levy crash into Jupiter. As the Earth turned and one observatory lost sight of Jupiter another observatory took over and followed the event. The comet had broken up during a passage a year previously, meaning that Jupiter suffered multiple impacts.

In 2014 ESA's Rosetta spacecraft rendezvoused with the comet 67P/Churyumov-Gerasimenko. Rosetta's lander Philae was the first to successfully land on a comet. Rosetta still accompanies the comet on its orbit around the Sun, while plans are made to make a controlled impact of the spacecraft on the comet.

Bibliography and Links

NASA is a great resource for background knowledge about planets and planet formation. Most of the facts in this chapter originate from the NASA website.

The image in Figure 2.1 is associated with ESO' press release number 1436, where one can read a more detailed account of planet formation:
www.eso.org/public/news/eso1436/

The recent investigation of Ceres' bright spots are covered in more depth in the press release here:
www.eso.org/public/news/eso1609/
And in the Dawn mission website here:
solarsystem.nasa.gov/missions/dawn

The ongoing New Horizons mission and its exploration of the Kuiper Belt can be followed on this page:
https://www.nasa.gov/mission_pages/newhorizons/main/index.html

The comet crashing into Jupiter is covered on the following pages:
solarsystem.nasa.gov/planets/pshoemakerlevy9/indepth
www.eso.org/public/news/eso9402/

3 THE TERRESTRIAL PLANETS

Your goal for this chapter is to learn about:

- Individual traits of the four inner planets and their moons.
  For a long time all that was known about the planets was derived from visual studies of the surfaces and shapes. A rotating body is always flattened along the rotation axis. Earth is flattened, so that its diameter is shorter from pole to pole, than the diameter at equator. The amount of flattening depends on the speed of the rotation and the strength of the material -a liquid deforms easier than a rock. By observing the shape of a planet and its rotation we can learn about the internal strength of the planet. Surface studies of planets reveal vulcanism, both old and new, continental drift, meteorite impacts, and climate.
  Now the solar system has been surveyed by space missions to remarkable detail, and planetary science has evolved into involving geophysics and ge- ology. By measuring slight changes in the gravitational field around planets and comparing it to physical and mathematical models we get an insight to the interior of the planets below the visual surface.

3.1 MERCURY

Mercury is the planet in our solar system closest to the Sun. Mercury is so close to the sun, it is hard to observe it directly from Earth except during dawn or twilight, where it is not outshone by the brightness of the Sun. When the Sun is high in the sky, Mercury is up there with it, we just do not see it. Using solar telescopes it is possible to observe Mercury, when it passes in front of the Sun. Such an event is called a transit. Mercury's next transit is November 11th 2019 (next one after that is Nov. 13th 2032), something that happens on average 13 times every century (transit predictions courtesy of Fred Espenak, NASA/GSFC).

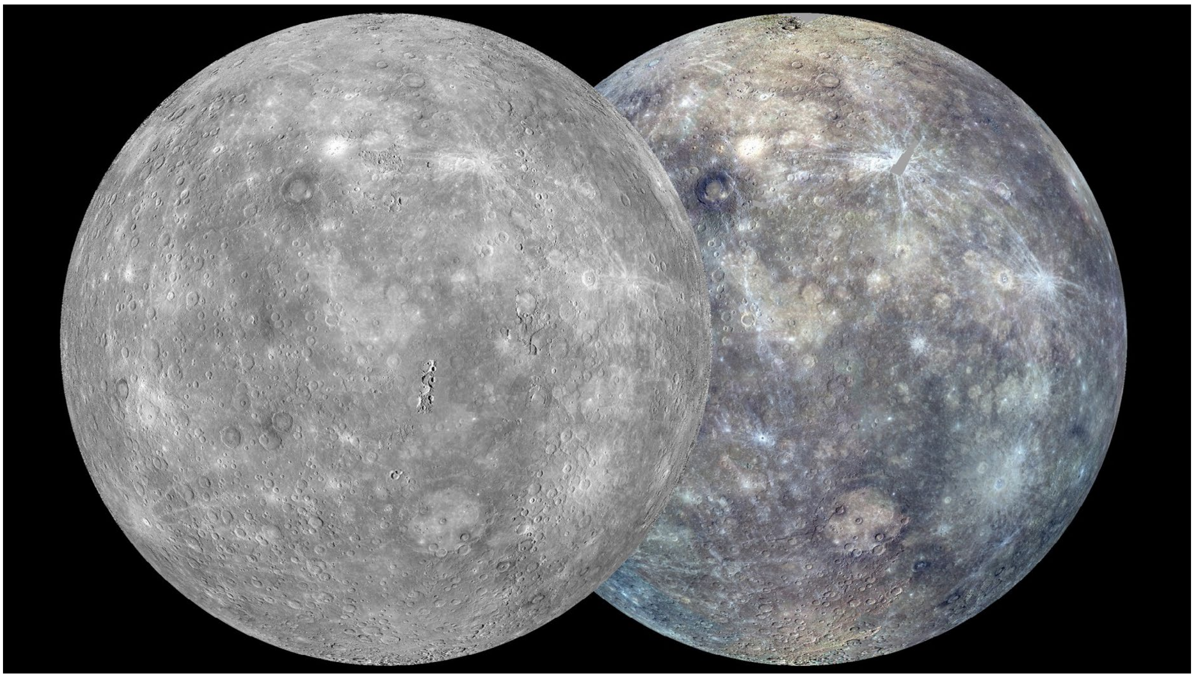


Figure 3.1: Mercury photographed by the MESSENGER mission.Image courtesy of NASA/Johns Hopkins University Applied Physics Labo- ratory/Carnegie Institution of Washington.

Mercury's orbit around the Sun is elliptical with a relatively high eccentricity. When Mercury is closest to the Sun, it is only 47 million km away, and when it furthest away from the sun the distance is 70 million km. Mercury completes an orbit around the Sun every 88 Earth days, which is then the length of a year on Mercury. Due to celestial mechanics (see Chapter 1.5) the closer a planet is to the Sun the faster it is in order to stay in orbit. Mercury is therefore the fastest of the planets.

Mercury, itself, spins once every 175.97 Earth days, making a day on Mercury last longer than a year. It was long thought that Mercury had a permanent day and night side, always facing the Sun with the same side, much like our Moon is always facing us. This is however not the case. Tidal forces have slowed Mercury's rotation down, not so it is always facing the Sun with the same side (as the Moon is tidally locked to Earth), but a resonant state, where its rotational period is 2/3 of its orbital period. That means the planet turns 1.5 times around itself between being in perihelion. After two full orbits around the Sun, the planet has revolved 3 times around itself. This resonate state happened because Mercury has a highly elliptical orbit. Tidal forces and tidal locking are described in Chapter 1.3.

When Mercury is closest to the Sun, the Sun appears more than 3 times as large as when viewed from the surface of the Earth. Mercury is the smallest planet in our solar system and only slightly larger than our Moon. Mercury is the planet that has the largest temperature differences between the day and night side. That is in part due to its slow rotation, as the Sun has a lot of time to heat up the surface, and the surface has a long time to cool down, before sunlight reaches it again. Mercury's non-existent atmosphere and the close proximity to the Sun are other factors that contribute to its extreme temperatures.

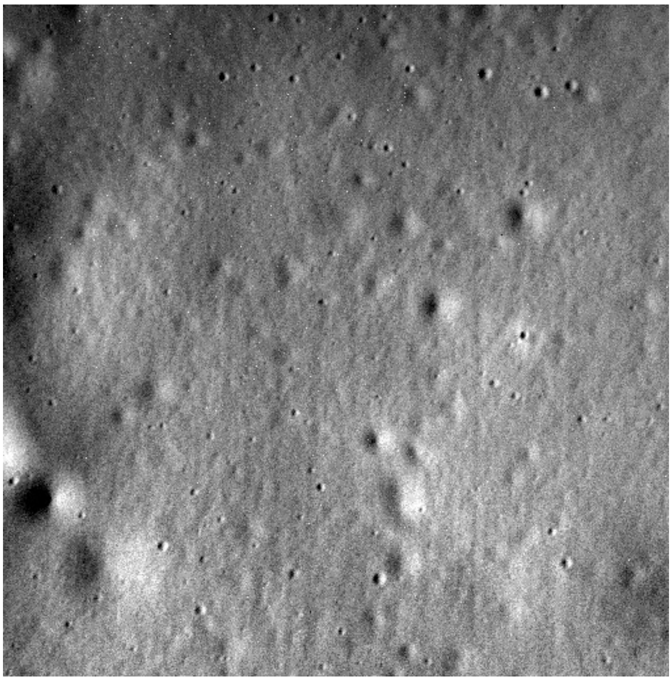


Figure 3.2: MESSENGER's last image. Image courtesy of NASA/Johns Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington.

Mercury's surface has many craters from impacts of meteoroids, just like the Moon, some of which dates back to the very early days of the solar system. Mercury's surface also features lobe-shaped scarps or cliffs, which formed as the planet's interior cooled and contracted. In the deeper craters close to the poles Mercury has permanent shadowed regions, which contain water ice. This was first discovered in 1991 by astronomers using radar observations. The latest space probe to visit Mercury, the MESSENGER, was able to confirm this.

Mercury is the second densest planet after Earth, with a large metallic core having a radius of about 2,000 km, that is about 80 percent of the planet's radius. The core is covered by an outer shell called the mantle, which is again enclosed by a shell called the crust. Mercury's mantle and crust is only about 400 km thick.

In 2007, researchers used ground-based radars to study the core, and found evidence that it is partly molten. Since Earth's magnetic field is generated by flowing magma outside the solid planet core, this discovery could lead to an understanding of Mercury's global magnetic field. Mercury's magnetic field is only one percent of the strength of Earth's magnetic field, and it is possible that Mercury's present field is a weak remnant of an earlier stronger field.

Mercury has been visited twice by space missions. The first spacecraft to visit Mercury was Mariner 10 in 1974, which imaged about 45 percent of the surface in flybys.

In 2008-2009, NASA's MESSENGER (MErcury Surface, Space ENvironment, GEochemistry, and Ranging) mission flew by Mercury three times and was orbiting the planet from 18 March 2011 until its inevitable demise. This mission was a big success and led to an abundance of new discoveries, like that the magnetic field is offset relative to the planet's equator and that the iron core has a liquid layer. The surface of the planet has also been imaged to a large level of detail, revealing that Mercury has been shaped both by vulcanism and impacts. MESSENGER had been planned to orbit Mercury for 1 year but lasted 3 additional years. When MESSENGER finally ran out of fuel on April 30th 2015, it was crashed into Mercury. The last image MESSENGER sent is shown in Figure 3.2.

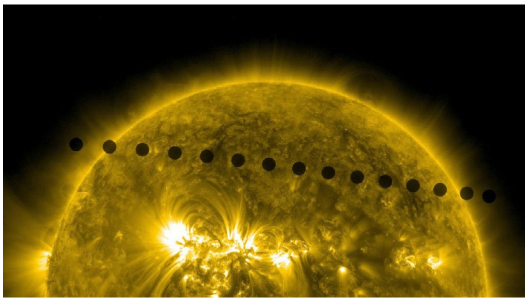
In October 2018 ESA's and JAXA's joint Mercury mission BepiColombo left Earth. The spacecraft will have several planetary flybys of Earth, Venus and Mercury before settling into an orbit around Mercury in 2025.

Mercury's orbit is one of Einstein's classic tests of General Relativity. Mercury's orbital motion has an irregularity in that the position of its perihelion rotates around the Sun, which in scientific term is called perihelion precession. The other planets in the solar system tug at Mercury but not enough to account for the full amount of precession the orbit shows. The remaining precession can be explained by general relativity, where gravitation is mediated by the curvature of spacetime. Einstein showed that general relativity agrees closely with the observed amount of perihelion shift.

3.2 VENUS

Venus is the second planet from the Sun in our solar system. Venus is always found close to the Sun in the night sky either at sunset or sunrise. This is the reason Venus often is called the morning star or evening star. Looking at Venus through a hobby telescope reveals that Venus is not a point source like the stars, but a planet, which have changing phases just like the Moon. We can even observe Venus, when it passes in front of the Sun. Such an event is called a transit, see Figure 3.3b for the latest transit observed with a solar telescope.

(a) Picture of cloud covered Venus.



(b) Transit of Venus.

Figure 3.3: Images courtesy of NASA/Solar Dynamic Observatory (SDO).

Venus is covered in featureless yellow and brown clouds, obscuring the surface completely, see Figure 3.3a. This cloud cover made it difficult to establish the rotation period, as it was the movement of the clouds that was recorded and not the planet.

In 1962, Venus was investigated using radar, which can look through the clouds, and the rotational period was determined. One day on Venus lasts 243 Earth days. Not only does it rotate very slowly, it also rotates the other way around than the other planets in the solar system. That means the Sun rises in West and sets in East. This is expressed in the inclination of the rotation axis being 177°, if the rotation axis was exactly opposite (take a spinning ball and turn it upside down) it would have been 180°. It is thought that this abnormal rotation is a result from an impact or merger with a large moon/meteorite. Venus has no moon.

If we measure the temperature at the top layer of the clouds it is only 250 K. At the surface, however, the temperature is 750 K, that is above the melting point of lead. Venus is so hot because of the greenhouse effect. Incoming radiation (sunlight) is absorbed and reemitted as infrared radiation, which is blocked by atmospheric carbon dioxide. Since the radiation cannot escape the planet, it stays and heats it up. Carbon dioxide is the main component of the Venusian atmosphere, which also contains H_2SO_4 (sulphuric acid).

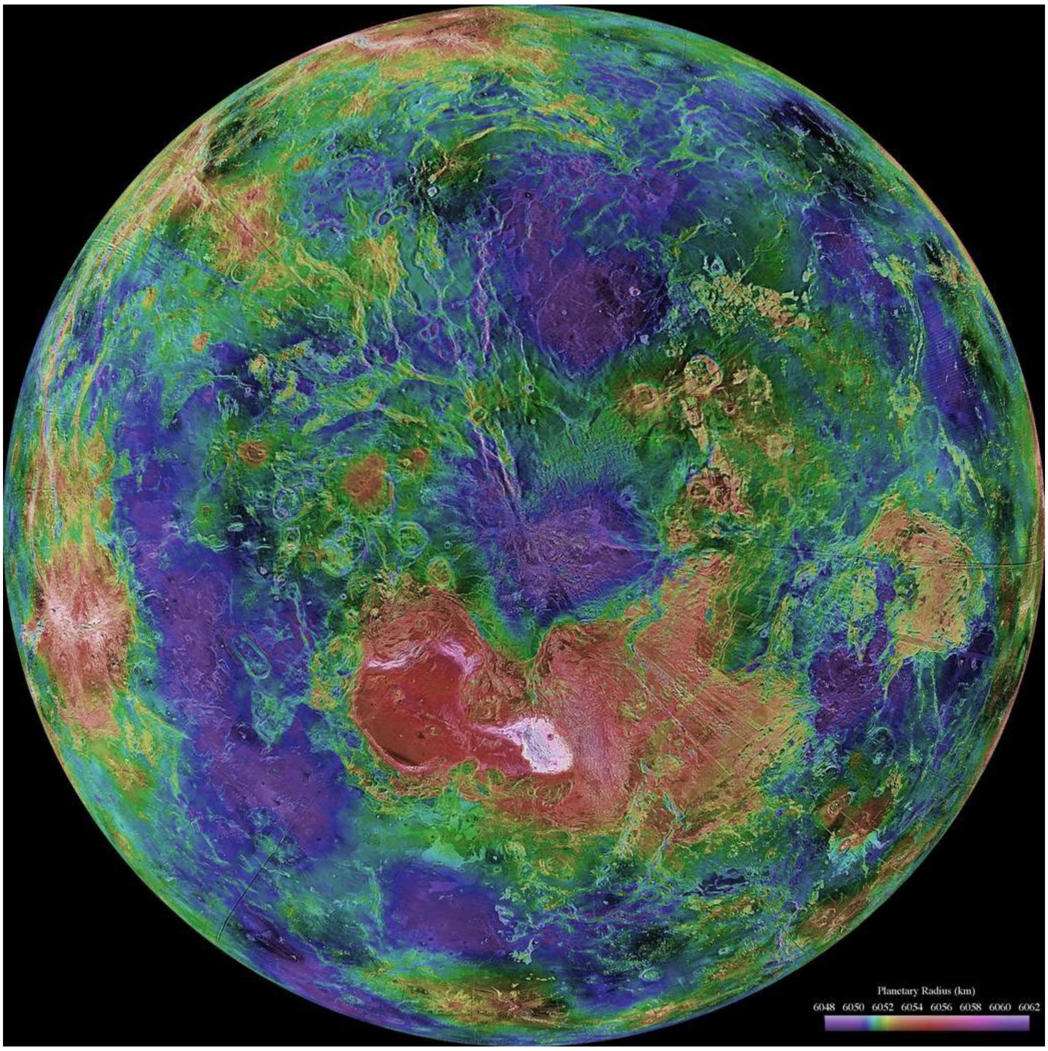


Figure 3.4: Radar image of Venus showing the terrain under the clouds. Courtesy of NASA/Johns Hopkins University Applied Physics Labora- tory/Carnegie Institution of Washington.

The atmospheric pressure at the surface is 90 times higher than on Earth (at sea level). This is because the atmosphere is thick and heavy. Imagine standing at the surface and all the air (matter) above you, all the way to space, is weighing down on you. This is the case on Earth as well, here we just do not sense it, as it is natural to us. The densest cloud layer is 50 km above the surface and is about 2-3 km thick. Above that layer is a hazy layer, which form the perceived appearance of Venus. The uppermost clouds move rapidly travelling once around the planet in 4 days.

From radar maps we know the Venusian surface to have canyons, mountains, craters, and volcanoes. There is no water on Venus. The surface can be divided into lowland (20%), lava flows and hilly upland (70%), and highland (10%). Impact craters on Venus are young (less than 500 million years), i.e. showing not sign of alteration from erosion or vulcanism. There are no craters smaller than 1.5-2 km because smaller meteoroids would have burned up in the thick atmosphere.

The Earth and Venus are similar in size, and their interiors are thought to be similar too. The iron core of Venus is 3,000 km in radius and covered by a rocky molten mantle about 2,000 km thick. The surface (crust) of Venus is similar to terrestrial granite and basalt. Venus has no magnetic field.

Mariner 2 (a NASA spacecraft, which operated in 1962) was the first spacecraft to encounter the planet, five years later the Russian Venera 4 sent back data from below the cloud cover. In 1975 Venera 9 and 10 sent back the first pictures of the surface. US Pioneer Venus 1 spent 18 months mapping the surface using radar in 1980. Finally, the best and most complete maps of the surface come from the Magellan spacecraft in 1990-1994 using radar observations. The resolution of the radar maps is 100m along the surface and the resolution of heights of surface structures (mountains etc.) is 30m, see Figure 3.4.

3.3 EARTH

Earth, our own planet, is the third planet from the Sun in our solar system. It is a rocky planet with a dynamic surface consisting of oceans and various terrain such as mountains, valleys, canyons, plains, deserts, ice sheets and so on. Only a few craters are visible, this is because most meteoroids burn up in the atmosphere. The few large enough to get through to the surface have left craters behind, which then erode relatively fast due to the dynamic surface involving weather, volcanoes, and/or plate tectonics.

71 % of Earth's surface is water, in fact Earth is one of the few places in the solar system where liquid water can be found. Further away from the Sun, water is present in its solid form, ice. Closer to the Sun, on Venus, water can be found in its gaseous state, an invisible vapour. Liquid water is thought to be important for the development of life, which is why the zone around a star, where planets can have liquid water, is called the habitable zone.

The structure of Earth's interior can be probed using seismic waves generated in earthquakes. There are two main types of waves produced by an earthquake, P-waves (primary/pressure) and S-waves (secondary/shear). Where the P-waves can travel through both solids and liquids, the S-waves are restricted to only travelling trough solids. This has revealed that Earth has a solid inner core, a molten outer core, a thick mantle and on top a thin crust.

Figure 3.5: A picture of Earth taken Dec. 7, 1972 by the Apollo 17 crew on their way to the Moon. Courtesy of NASA.

Earth's atmosphere is primarily nitrogen (78 %) and oxygen (21 %). The atmosphere stretches out 500 km above ground and consists of several layers each having a different density, pressure and temperature. The ozone layer is a thin layer of O_3 (ozone), which reflects the Sun's UV radiation. Another important layer is the ionosphere, where the gas is in the form of fully ionised plasma. This layer reflects radio waves on Earth and is thus important for broadcasting corporations. It is also the layer where the aurora (i.e. Northern/ Southern lights) happens.

Earth has a global magnetic field, which originates in the core and stretches out to beyond the atmosphere where it is called the magnetosphere. The magnetosphere is the outer boundary of a planet and it protects us from the solar wind. The magnetic field axis is offset from the rotation axis by 11°. The magnetosphere is limited by the magnetopause, flattened on the solar side and extended to a long tail on the opposite side. Charged particles inside the magnetopause are captured by the magnetic field. The region of space containing trapped charged particles, the radiation belts around Earth, are called van Allen's belts. The number of trapped particles increase after strong solar bursts. Some of these particles leak into the atmosphere at the poles and result in the auroras. Similar phenomenas have also been detected on Jupiter, Saturn and Uranus.

Earth's magnetic field is generated by its planetary dynamo. The require ment for a planetary dynamo to generate a magnetic field is that the planet is rotating and has a convective layer of electrically conductive material. In the case of Earth, the convective layer is the outer molten core, consisting of iron and nickel, both elements known to be electrically conductive. Generally speaking convection is a movement within a liquid with temperature differences, the topic will be covered deeper in Chapter 6 about the Sun.

Figure 3.6: Aurora from space courtesy of NASA.

3.3.1 THE MOON

Earth's Moon probably formed when a large body about the size of Mars collided with Earth, ejecting a lot of material into orbit. Debris from the early Earth and the impacting body accumulated to form the Moon approximately 4.5 billion years ago -the age of the oldest collected lunar rocks.

Figure 3.7: Near and Far side of the Moon courtesy of NASA.

The surface of the Moon is covered in craters, all of which are from meteorite impacts. The Moon doesn't have an atmosphere, vulcanism, or tectonic activity, so the craters will not be destroyed. That also implies that many of the craters are very old.
On its near side the Moon has several almost circular dark areas called maria (plural of mare). The maria are areas of lower altitude, which were formed when large collisions penetrated the thin crust and allowed molten rock to flow across the surface. There is only one mare on the far side of the Moon, this is because the crust is a lot thicker on that side, and only one impact was strong enough to penetrate the crust.

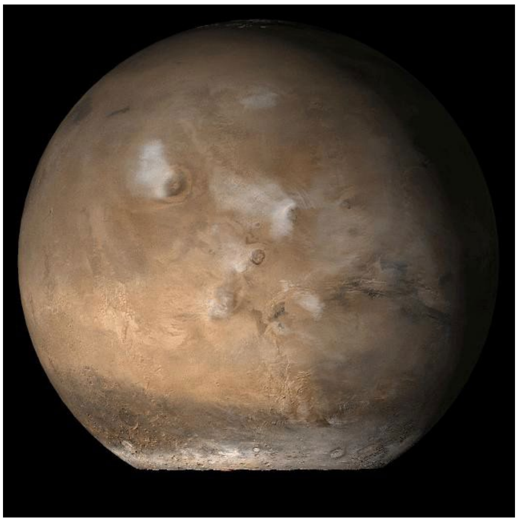
The Moon has seismic activity, which in part is associated with tidal forces from the Earth acting on the Moon. Measurements of the moonquakes gives us knowledge about the interior of the Moon. The Moon has undergone differentiation, where material with different chemical composition and physical properties settle into different layers. Simply put, the denser material sinks to the centre and lighter material rises to the surface. For the Moon this means it has a thin crust, a mantle and a core. Analysis of its rotation suggests that the Moon's interior is partly molten.

The United States' Apollo missions from 1969-1972 saw a total of twelve astronauts walk on the Moon. They also brought back almost 400 kg surface rocks, both from the maria and from the upland. Analysis of the rock samples determined the age of the upland surface to be 700 million years older than the maria. This then also explains that the maria have fewer craters than upland areas.

Besides the manned missions, the Moon has been studied by more than 100 space missions originating not only from the US but also from Russia, China, and India to name just a few.

3.4 MARS

Mars is the fourth planet from the Sun. Because Mars is further away from the Sun than Earth, Earth overtakes it on the inside. On the night sky it then looks like Mars first moves in one direction and then in the opposite direction. A day on Mars is 24.617 Earth hours (i.e. almost the same as Earth) and a year 687 Earth days.



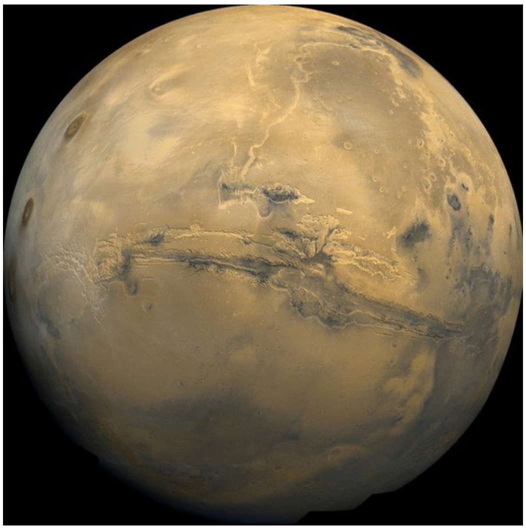


Figure 3.8: Mars. Images courtesy of NASA/JPL-Caltech/MSSS.

Mars is often called the red planet. This dry planet is covered in a fine dust with a reddish tinge to it due to the high iron content. Mars has polar ice caps and a vast canyon called Valles Marineris running along its equator. The canyon is in places up to 7 km deep and it is 4,000 km long, that is about 20% of the circumference of Mars. Mars has, like Venus and Earth, a dynamic surface, which erodes smaller meteorite craters. Volcanoes, impact craters, crustal movement, and atmospheric conditions such as dust storms have all altered the surface of Mars.

Mars' atmosphere is extremely thin and consists mainly of carbon dioxide (95%). Mars has weather, thin clouds and dust storms, and due to a tilt of its rotation axis (25%), it also has seasons. The polar ice caps on Mars grow and recede with the seasons. However, due to the thin atmosphere liquid water cannot exist on the surface for long. Mars is a desert world and although there are signs of ancient floods, the only evidence for water are the ice caps and the thin clouds.

The mean temperature is -50ºC and on warm summer days the temperature at equator can rise close to 0ºC. Mars' orbit is rather elliptical causing a 30º C difference between aphelion and perihelion.

The interior of Mars is not well known. Most likely it has a dense core, a molten rocky mantle and a thin crust. Mars has no global magnetic field today, but its crust in the Southern Hemisphere is highly magnetised, suggesting that it once had a magnetic field. The lack of a magnetic field suggests that the core might be solid, as there then would not be the needed convective and conductive layer to generate a planetary dynamo.

Although Mars is half the size of Earth it has the largest volcano in the solar system namely Olympus Mons, which at 20 km high is 3 times as large as mount Everest. Mars shows no signs of plate tectonics, has no mountain chains, nor any global patterns of vulcanism. The lava flows on Olympus Mons are less than 100 million years old.

Mars has in the past had visits by space probes and landers, which communicated their findings back to Earth. Currently Mars is being explored by 10 crafts and rovers. NASA has the Mars Reconnaissance Orbiter, MarCo, MAVEN, Mars Express and Mars Odyssey in orbit around Mars. ESA has the ExoMars Trace Gas Orbiter and India the Mars Orbiter Mission in orbit around Mars. On the ground NASA's rovers Opportunity and Curiosity investigate the surface and since November 2018 InSight is probing the interior of Mars. Figure 3.9 are two pictures of the Martian surface taken by Curiosity, one is a selfie the other of active sand dunes, which migrate up to about one meter per year.

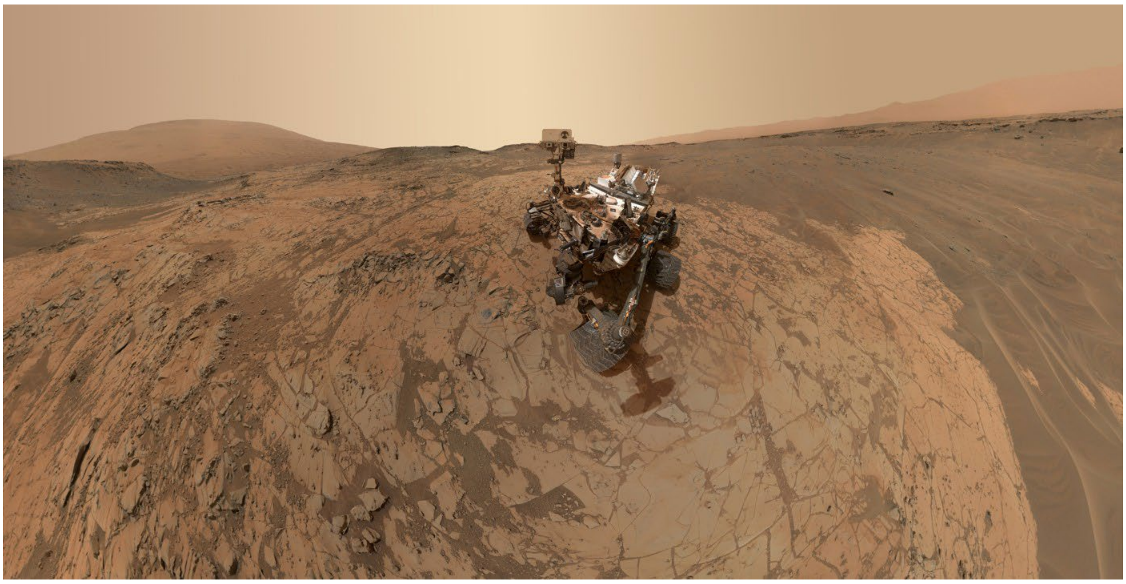
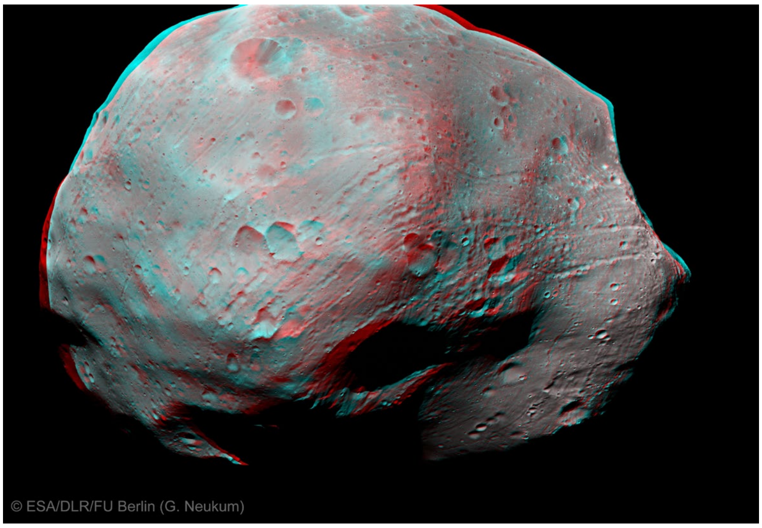


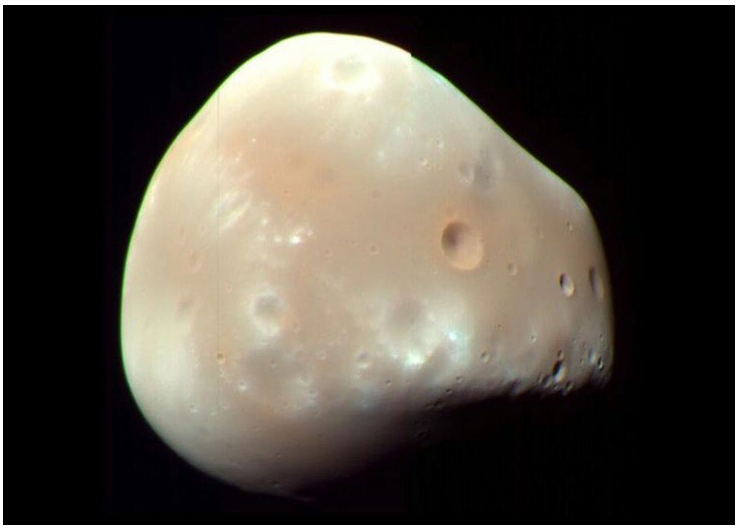
Figure 3.9: NASA's curiosity rover on Mars. Top: A selfie from February 2015. Bottom: The rover starts to investigate dark sand dunes December 2015. The dunes can be up to two stories tall. Pictures courtesy of NASA/JPL-Caltech/MSSS.

3.4.1 THE MOONS OF MARS

Mars has two small moons Phobos and Deimos, see figure 3.10, which might be captured asteroids. They are shaped like potatoes, as they have too little mass for gravity to make them spherical.



(a) Phobos.
Image courtesy of ESA/DLR/ FU Berlin (G. Neukum).



(b) Deimos.
Imagecourtesy of NASA/JPL- Caltech/University of Arizona.

Figure 3.10: Mars' two moons.

	Unit	Mercury	Venus	Earth	Mars
Mass, M Mass, M	[$M_{\otimes}$] [kg]	0.055 3.30×10^{23}	0.815 4.87×10^{24}	1 5.97×10^{24}	0.107 6.42×10^{23}
Radius, R Radius, R	[$R_{\otimes}$] [km]	0.38 2.44×10^{3}	0.95 6.05×10^{3}	1 6.37×10^{3}	0.53 3.39×10^{3}
Density, ρ Density, ρ	[$\rho_{\otimes}$] [g/cm3]	0.98 5.43	0.95 5.24	1 5.513	0.714 3.934
Inclination	[degrees]	0.0	177.3 R	23.44	25.2
Eccentricity		0.206	0.00678	0.0167	0.0934
τ rotation	[$Day_{\otimes}$]	175.97	243.02	1	1.03
Dist. to Sun Dist. to Sun	[AU] [km]	0.39 5.79×10^{7}	0.723 1.08×10^{8}	1 1.50×10^{8}	1.52 2.28×10^{8}
τ orbit τ orbit	[$Year_{\otimes}$] [$Day_{\otimes}$]	0.24 88	0.62 225	1 365.26	1.88 687
Surface Temp.	[K]	100/700	735	185/331	120/293
Moons	[number]	0	0	1	2

Table 3.1: Facts and stats of the inner 4 planets, numbers from NASA.
$\otimes$: In Earth units.
Inclination of the planet with respect to ecliptic. R=retrograde rotation.
Eccentricity, e, as defined in Chapter 1.5 about elliptical orbits.
τ rotation is the time it takes the planet to spin on its axis once. Dist. to Sun is the average distance to the Sun.
τ orbit is the time it takes the planet to complete one orbit around the Sun.

Phobos, the innermost moon, is heavily cratered, with deep grooves on its surface. It is orbiting Mars at just 6,000 km from the planet's surface, and thus closer to its parent planet than any other known moon in our Solar System. The moon's proximity means that it hurtles around Mars faster than the planet rotates: for an observer on the surface of Mars, Phobos would appear to rise and set twice a day. The moon's orbit is decreasing and in some 50 million years time it will likely break up to form a debris ring around Mars, before colliding with the planet's surface.

3.5 SUMMARY OF THE TERRESTRIAL PLANETS

The physical parameters for the inner 4 planets have been summarised in Table 3.1.

Bibliography and Links

The source of most of the information in this chapter came from NASA,
which has a website for each body in the Solar System:
http://solarsystem.nasa.gov/planets/mercury
http://solarsystem.nasa.gov/planets/venus
http://solarsystem.nasa.gov/planets/earth
http://solarsystem.nasa.gov/planets/mars

Johnson et al., 2015, 'Low-altitude magnetic field measurements by MES- SENGER reveal Mercury's ancient crustal field', Science, Vol.348, Issue 6237, pp.892-895.

Transits (the passage of Mercury and Venus across the face of the Sun) are calculated by Fred Espenak, NASA/GSFC). Upcoming transits can be found here:
Mercury: http://eclipse.gsfc.nasa.gov/transit/catalog/MercuryCatalog.html
Venus: http://eclipse.gsfc.nasa.gov/transit/catalog/VenusCatalog.html

4 THE JOVIAN PLANETS

Your goal for this chapter is to learn about:

- Individual traits of the outer gaseous planets and their moons.

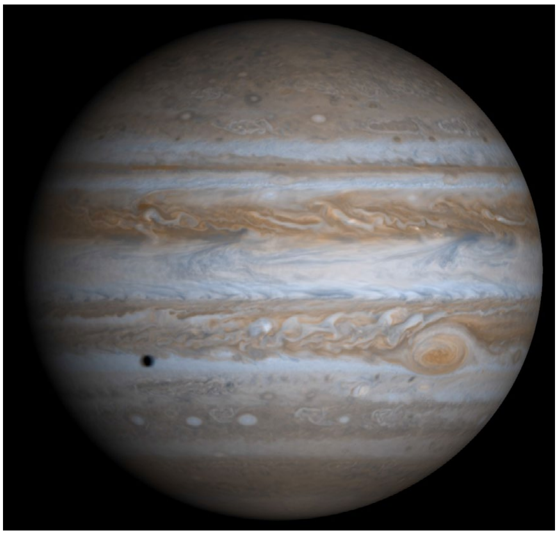


Figure 4.1: These true-color images of Jupiter were taken by NASA's Cassini spacecraft. The black dot on the left image is the shadow of Jupiter's moon Europa. The right hand image is from the South pole. Images courtesy of NASA/JPL/University of Arizona/Space Science Institute.

4.1 JUPITER

Jupiter is the fifth planet from the Sun and also the largest planet in our solar system. It consists mainly of hydrogen and helium, with a relative abundance similar to the Sun's. While Jupiter is roughly 1/1000 the mass of the Sun it's density is the same as the Sun's 1.330 g/cm^3.

The surface of Jupiter is stripy with red belts and white zones, see Figure 4.1. These belts and zones are stable cloud formations parallel to the equator, which explains why they appear as rings in the view from the pole. The gas flows upwards in the white zones and downwards in the red belts.

Figure 4.2: Voyager I image of Jupiter's Great Red Spot. Courtesy of NASA.

There is a distinct Great Red Spot in a belt just below the equator, see Figure 4.2. The spot is a huge cyclone rotating counter clockwise, which has the same size as 3 Earths next to each other. The picture in Figure 4.2 is from Voyager I's flyby in 1979. When we compare the region around the Great Red Spot on the Voyager image with that taken by Cassini from 2000 we can see that the cloud formation has changed in the intervening time. The clouds change on a much shorter timescale than 30 years. This became clear in 1979 as Voyager II flew past Jupiter only 4 months after Voyager I, and showed changes in this short time. However, the Great Red Spot is relatively stable, as it has been there since the first observations of Jupiter, with a high enough resolution to see it, almost 400 years ago.

Jupiter, not surprisingly, does not behave as a rigid body, meaning that different layers can rotate with different speeds. Jupiter has a rather strong magnetic field, and by measuring that, scientists have established that Jupiter makes a full rotation every 10 hours. A rotational speed this fast causes a significant bulge, meaning that the planet is squeezed at the poles and fatter at the equator.

Jupiter has an inner iron nickel core, which weights a few tens of Earth masses. This core is surrounded by a layer of singular hydrogen, also called metallic hydrogen. Hydrogen is most common in its molecular state (H_2), but this deep into Jupiter the weight and pressure from the layers above (the pressure is 3×10^6 atm) means hydrogen becomes dissociated into single atoms. Hydrogen in this state is electrically conductive and serves in Jupiter's case as the location for its magnetic field. Most planets with a significant magnetic field have in their interiors some conductive moving material, be it liquid or gaseous, powering or driving the magnetic fields. Figure 4.3 shows a picture of an aurora on Jupiter.

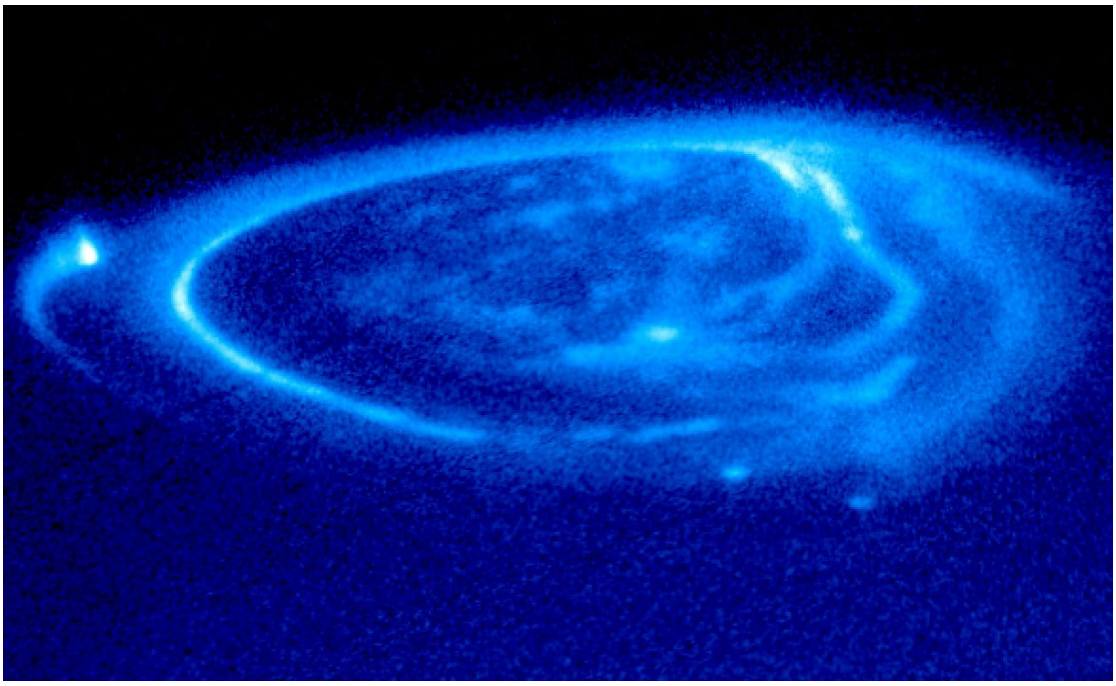


Figure 4.3: Hubble image of the aurora on Jupiter. Courtesy of NASA.

Jupiter radiates twice as much energy as it receives from the Sun. This is because the planet has a relatively large internal heat stemming from its formation. Furthermore, the planet is still cooling and contracting, converting potential energy to kinetic energy and thereby to heat.

4.1.1 JUPITER'S MOONS AND RINGS

Jupiter has three rings and 79 confirmed moons. The most significant moons are the 4 largest, also collectively called Galilean moons, Io, Europa, Ganymede, and Callisto. The Galilean moons are named so because they were discovered by Galileo in 1610. Many of Jupiter's outer moons have highly elliptical orbits and orbit backwards (opposite to the spin of the planet).

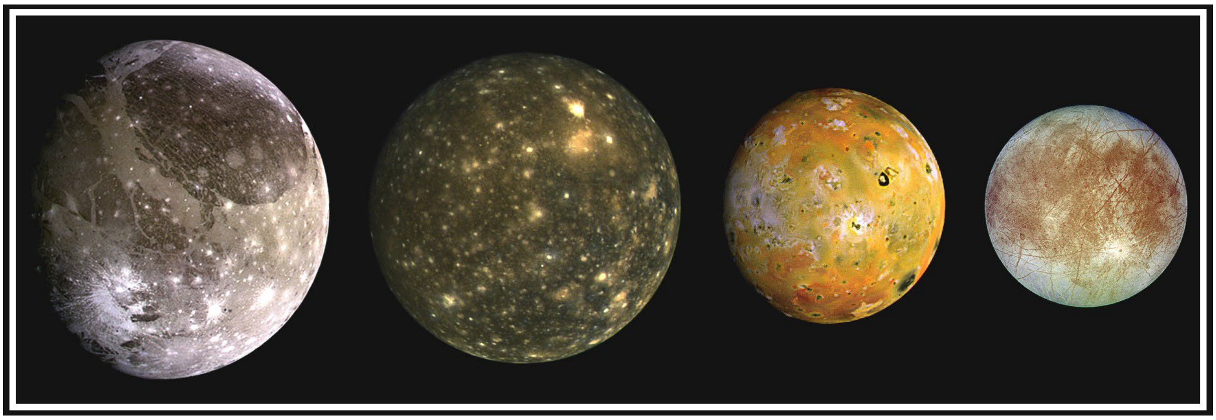


Figure 4.4: The Galilean Moons to scale. From left to right, the moons shown are Ganymede, Callisto, Io, and Europa, courtesy of NASA Planetary Photojournal.

Figure 4.4 shows the four largest moons next to each other to scale, ordered after size. In order of increasing distance from Jupiter, Io is closest, followed by Europa, Ganymede, and Callisto. The different distances of the moons to Jupiter help explain some of the visible differences amongst the moons.

The moon Io is subject to strong tidal forces from Jupiter. This creates internal heating, which is released at the surface through volcanoes. Io is tidally locked and always shows the same side to Jupiter, but perturbations by Europa and Ganymede make Io's orbit an irregular ellipse rather than a circle. Because of this irregular elliptical orbit, the tidal forces from Jupiter vary and cause Io's surface to bulge up to 100 m. With tidal forces as a source of internal heat, the subsurface crust is liquid, relieving pressure through volcanoes. The surface is thus constantly changing and impact craters don't last long before they are filled with molten lava. Data from the Galileo spacecraft indicates that an iron core may form Io's centre, thus giving Io its own magnetic field.

The moon Europa is a moon with an icy surface, an iron/rocky core and potentially large bodies of water in between those two layers. The icy sur face is smooth and bright, consisting of water ice crisscrossed by long, linear fractures. Europa's orbit is also tidally locked, making Europa always show Jupiter the same side, but because the orbit is not perfectly circular, Europa has tides, as it is sometimes closer or further away from Jupiter. Also the difference in gravity from the near to the far side of Europa causes tidal forces. The tidal forces stretch and relax Europa's surface causing the linear fractures along its surface. If Europa's ocean exists, the tides might also create volcanic or hydrothermal activity on the seafloor, supplying nutrients that could make the ocean suitable for living things.

The moon Ganymede is the largest moon in our solar system. It is larger than Mercury and Pluto, and three-quarters the size of Mars. Ganymede has a metallic core, which is where its magnetic field is generated, a rocky mantle, surrounded by a thick layer (~ 800 km) consisting mostly of ice intermixed by larger rocks. The icy surface is a mixture of two types of terrain. Forty percent of the surface of Ganymede is covered by highly cratered dark regions, and the remaining sixty percent is covered by a light grooved terrain, which forms intricate patterns across Ganymede. The light regions are young and smooth with only a few craters, whereas the dark regions on Ganymede are old and rough, and probably the original crust of the satellite. Craters in the light terrain have probably been washed away by water released from within or by tensional faulting.

The moon Callisto is the third largest moon in our solar system and is almost the size of Mercury. Its interior is probably similar to Ganymede except the inner rocky core is smaller, and this core is surrounded by a large icy mantle. Callisto's surface is the darkest of the Galileans, but it is twice as bright as our own Moon. Callisto is the most heavily cratered object in our solar system, and probably also the oldest surface, as it shows almost no sign of resurfacing.

In 1675 the astronomer Ole Rømer carefully studied the Galilean moons' orbits and noticed something strange. The orbits of the moons follow Kepler's third law as their movements are only governed by gravity. Rømer studied the moons movements into the shadow of Jupiter, and he was able to predict future moon eclipses. However, he noticed that when Earth was closer to Jupiter, the eclipses happened earlier than expected. Similarly, when Earth was further away from Jupiter, the eclipses were delayed. Rømer realised that the discrepancy was caused by the differing amounts of time it took the light to travel the changing distance between Earth and Jupiter. He concluded that light required 22 minutes to cross the diameter of Earth's orbit (the actual value is 16.5 minutes).

Figure 4.5: Above a real image of Jupiter's thin rings. Courtesy of NASA/JPL.

Jupiter has three very thin rings, which was not discovered until 1979 by a Voyager flyby. Figure 4.5 shows an eclipse of the sun by Jupiter, as viewed from the Galileo spacecraft. Small dust particles high in Jupiter's atmosphere, as well as the dust particles that compose the rings, can be seen by reflected sunlight. Jupiter's rings were a surprise, as they are composed of small, dark particles and are difficult (but not impossible) to see except when backlit by the sun.

In the same orbit as the rings Jupiter has 4 smaller moons, which are the source of the dust which forms the rings. Impacts by small meteoroids (fragments of asteroids and comets) into these small, low-gravity moons feed material into the rings. Therefore, the three rings are bounded by the small moons, which feed material to the ring closest to them.

Jupiter's faint outer rings are called the gossamer rings. The moons Thebe and Amalthea supply dust to those rings, which are thicker because Thebe and Amalthea orbit Jupiter on inclined paths. The thin, narrow main ring, is bounded by the small moon Adrastea and shows a marked decrease in brightness near the orbit of Jupiter's innermost moon, Metis. It is composed of fine particles knocked off Adrastea and Metis. The innermost and thickest ring is the halo that ends at the main ring.

Since 2016 Jupiter is being explored by NASA's Juno spacecraft. The mission is to investigate the atmosphere, the deeper structure and the magnetosphere.

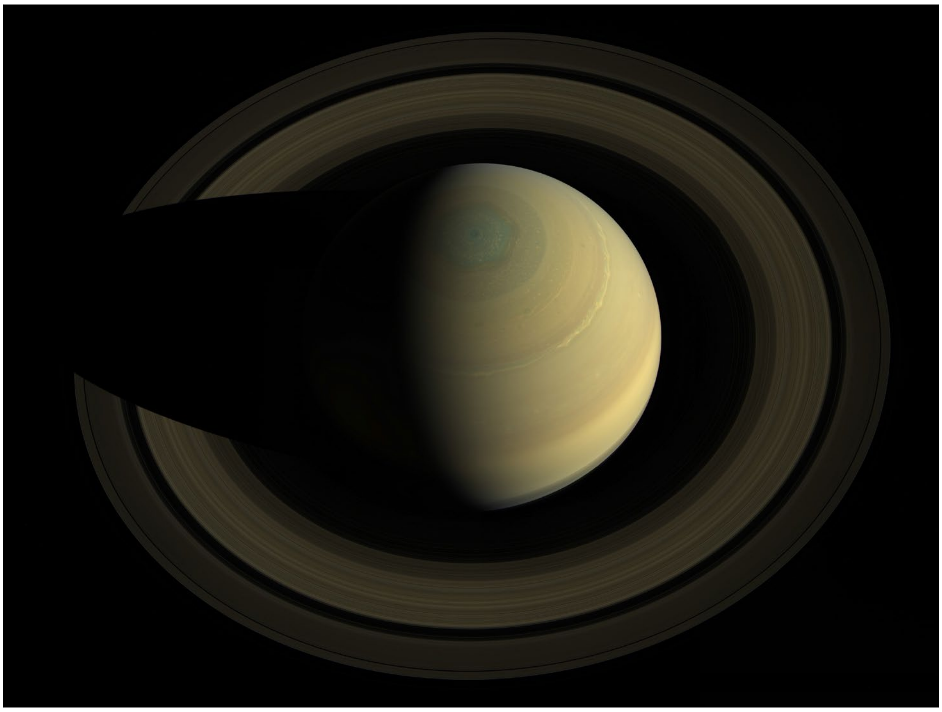


Figure 4.6: NASA Cassini spacecraft caught a view of Saturn's North pole. Courtesy of NASA/JPL.

4.2 SATURN

Saturn is the sixth planet from the Sun, and the second largest in the solar system. What sets Saturn apart from the other planets, is its spectacular ring system observable even with smaller telescopes. As with Jupiter, Saturn is mainly composed of hydrogen and helium and is observed to be of very low density. Saturn's mean density is only about two-thirds that of water. Figure 4.6 is a picture taken by NASA's spacecraft Cassini, which ended its 20 year mission in 2017. The picture is in natural colours and this polar view gives us a clear opportunity to see a hexagonal weather pattern on Saturn's North pole as well as the shadow of Saturn on its rings.

Saturn is yellow in colour because of clouds in its upper atmosphere which consist of hydrogen, ammonium and methane. It turns out that the colour of Saturn changes with season. The hemisphere where it is summer is yellow, and the hemisphere where it is winter is bluish. This change in colour was discovered by the Cassini spacecraft. The image in figure 4.6 shows the northern hemisphere well into spring, which is why there is only a little blue tinge left in one of its bands and on the pole itself.

Saturn's inner structure resembles that of Jupiter, with the only difference being size. In the centre there is a dense rocky core with nickel-iron alloys. It is enveloped by liquid metallic hydrogen, similar to Jupiter's layer but considerably smaller. This is the place and source of Saturn's magnetic field, which is smaller than Jupiter's but still 578 times as powerful as Earth's. On top of the metallic hydrogen is an envelope of hydrogen and helium, which is liquid closer to the core and gaseous furthest out. There is no clear boundary layer between the two states.

Clouds in Saturn's hydrogen-helium atmosphere are layered according to their composition. The very top layer of clouds are those made of ammonia, below those are the clouds made of hydrogen sulphide and below those water clouds. Strong winds in the upper atmosphere cause the bands or stripes on Saturn. Curiously, the energy driving these winds come from differentiation of the planet. Differentiation is when a planet settles into layers of different chemical composition, and in Saturn this is still taking place. For Saturn this is a separation of hydrogen and helium, where the heavier helium is gradually sinking towards the centre, while releasing potential energy in the form of thermal radiation. This energy release drives the strong winds on the surface and also causes Saturn to radiate about 2.8 times the energy it receives from the Sun.

As mentioned, Saturn has a magnetic field, and it occasionally gets auroras, just like Earth does. From the magnetic field the rotation period of Saturn has been measured to be only 10.5 hours. This rapid rotation causes the planet to be flattened, i.e. thicker around the equator than at the poles, which can be seen in all pictures of Saturn.

4.2.1 SATURN'S RINGS AND MOONS

Saturn has a spectacular ring system and 53 known moons plus 9 moons awaiting official confirmation. The rings are located in Saturn's equatorial plane and can be viewed even with a small telescope. The rings are actually many narrow rings, with varying brightness. Some dark lines are in fact gaps between rings. Historically the rings have been named in sequence of their discovery. Therefore the order of the rings, from the planet and out are: D, C, B, A, F, G, and E. This implies that Saturn has 7 rings, however, with the increased level of detail that the space missions provide, it looks like the rings are made up of several narrower rings.

Figure 4.7: Cassini (NASA) image of Saturn's rings from about 20 degrees below the ring plane. The blue dot underneath the rings is Earth. Courtesy of NASA.

The main rings are C, B and A. The Cassini Division is the largest gap (4,700 km) in the rings and separates rings B and A. The D ring is closest to the Planet and it is much fainter than its neighbouring ring. The F Ring is very thin and lies just outside the A Ring. The G ring is much fainter than the F ring and lies a fair bit further out. The final ring, the E ring lies furthest out and is much more fuzzy in its appearance. Figure 4.7 shows a picture taken by the spacecraft Cassini of Saturn's rings and our planet Earth in the same frame. The very bright and narrow ring is the F ring, and the fainter ring further away is the G ring. The E and G rings have in the image been brightened for better visibility.

Saturn's ring system extends up to 282,000 km from the planet, but the rings have a vertical height of 1 km or less. The main rings are generally thinner than the rest. The Cassini spacecraft discovered vertical formations on the rings, where a pile up of particles reached 3 km up.

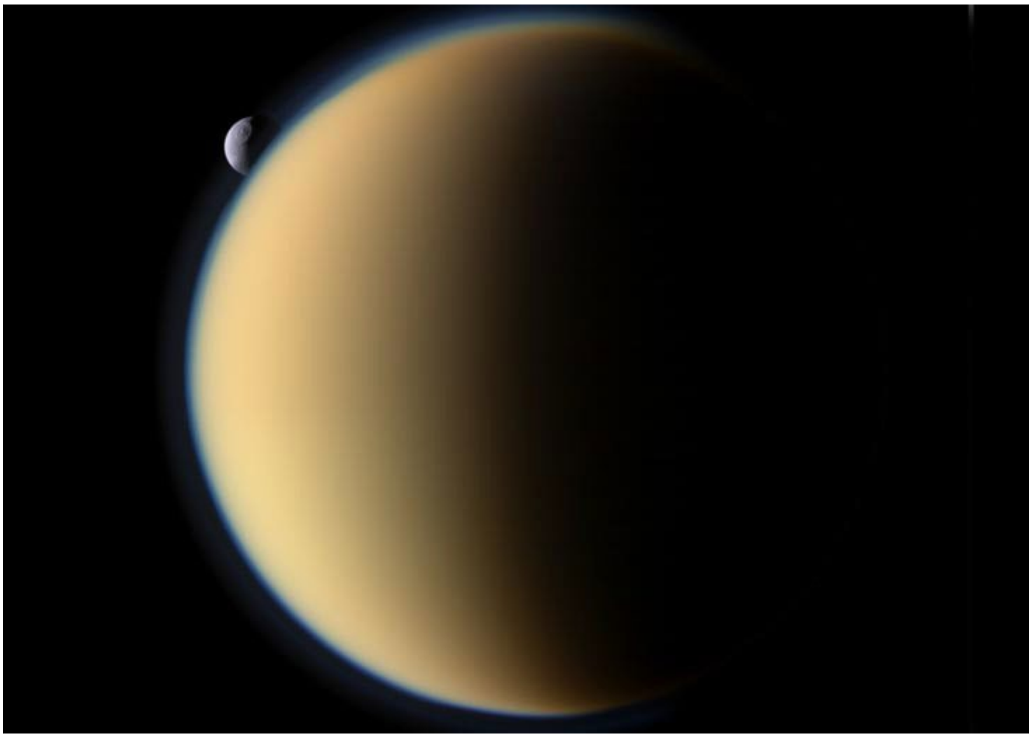


Figure 4.8: Saturn's moon Tethys moves behind Saturn's largest moon Ti- tan. Courtesy of NASA/JPL/ Space Science Institute.

The gaps in the ring system are caused and maintained by smaller moons.

The Cassini division is caused by the gravitational influence by the moon Mimas. The Encke and Keeler gaps are also kept open by two tiny moons orbiting in gaps. Moons with this function are called shepherd moons, and they help keep the rings in line. Other shepherd moons are Pandora and Prometheus. Pan and Atlas cause weak, linear density waves in Saturn's rings, which in turn have yielded more reliable calculations of their masses.

The ring particle sizes range from tiny dust-sized icy grains to a few particles as large as mountains. The composition of the ring particles is dominantly water ice. The origin of the rings is still not settled. It could either be remnant from the planet formation of it could be pieces of comets, asteroids or shattered moons that have been broken up before they reached the planet. Some ice in the E ring comes from the nearby moon Enceladus' geysers.

Saturn's many moons provide us with great variety. For the sake of brevity we will only look at Saturn's largest moon Titan and the volcanic Enceladus in some detail.

Titan is interesting because it has a thick nitrogen-rich atmosphere, which might be similar to what Earth's was like before biology influenced it. Another similarity with Earth is the cycle of liquids flowing across its surface, which seems to follow a seasonal pattern. In Figure 4.8 we see how Titan's atmosphere makes it appear hazy, while the moon Tethys in the background, show what the level of detail could have been for Titan, was it not for fuzziness of its atmosphere.

NASA's Cassini spacecraft carried the European-built Huygens probe, which was sent to the surface of Titan. The temperature at Titan's surface is about –178° C and the pressure is about 60 % larger than Earth's, so the flowing liquid is not water. Titan has instead, as

Cassini-Huygens determined, lakes and seas of liquid methane and ethane near its poles. These bodies of standing liquids appear to change in size in a seasonal way, as the evaporation of liquid and the condensation of gas into liquid depend on sunlight. The mission has also found creases on the surface that were carved by flowing liquid. From the orbit around Titan, Cassini's radar instrument revealed that large swaths of the surface near the equator are blanketed by dune fields, similar to the Namibian desert on Earth. The mission has also found that Titan has an internal ocean of liquid water.

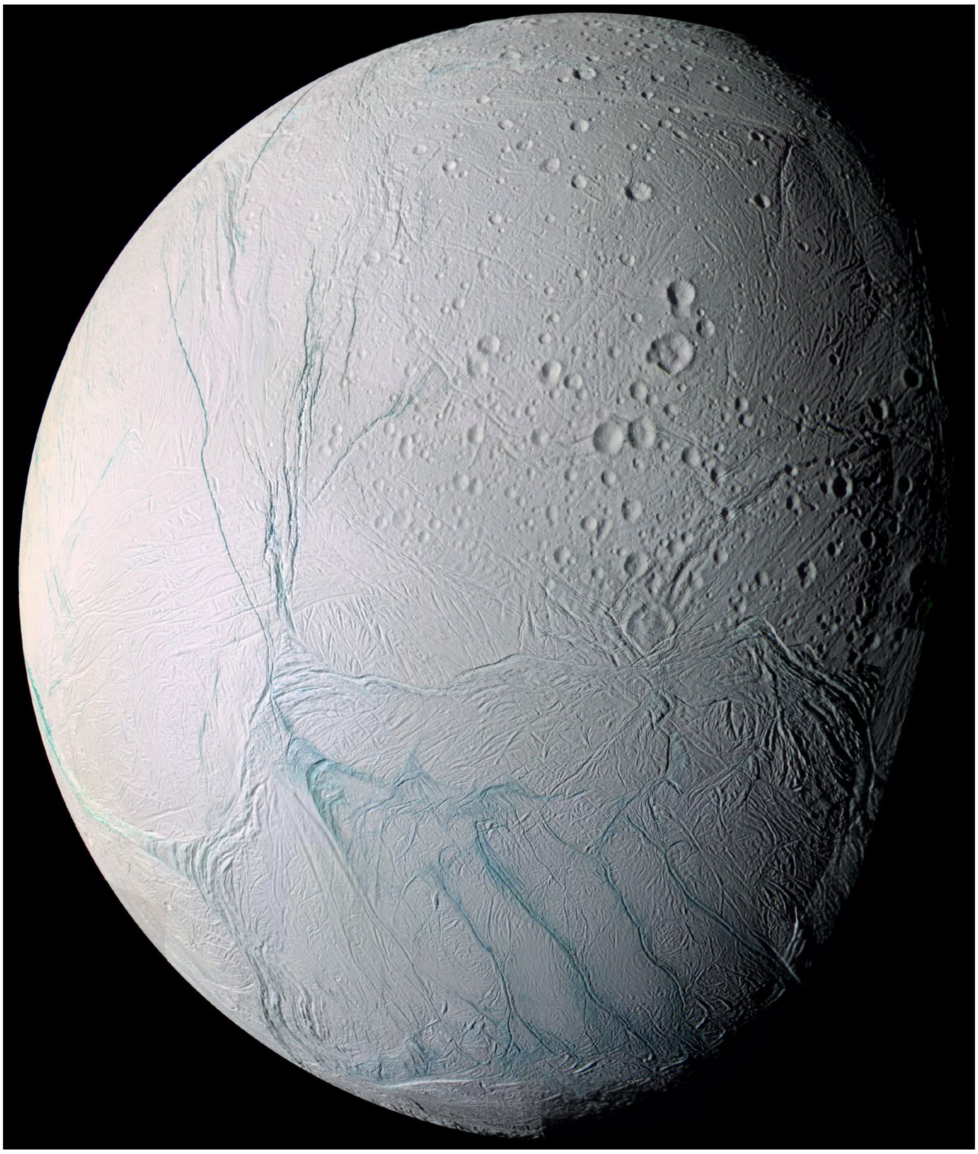


Figure 4.9: NASA Cassini took this spectacular picture of Enceladus on a flyby. Image courtesy of Cassini Imaging Team/SSI/JPL/ESA/NASA.

The moon Enceladus has a slight wobble in its orbit, suggesting that only its surface is frozen and probably covering a global ocean. This global ocean feeds the plumes spouting from Enceladus' surface. Figure 4.9 shows a picture of Enceladus taken by Cassini on a flyby. The false coloured blue stripes are warm fractures from which gasses from the interior escapes in a process like geysers on Earth. The plumes consist of water vapour, carbon dioxide, methane, perhaps a little ammonia and either carbon monoxide or nitrogen. The material released in the geysers is refreshing the surface and causing an enormous halo around the moon and contributing material to Saturn's E-ring, the one furthest out. Given Enceladus' distance to the Sun the moon should be frozen to the core, however, it is likely that it is heated by tidal friction in the same way as Jupiter's moon Io.

4.3 URANUS

Uranus is the seventh planet from the Sun and it is about 20 times as far away from the Sun as the Earth. It takes Uranus 84 Earth years to make one orbit around the Sun. Voyager 2 made a flyby of Uranus in 1986 and it is the only spacecraft (thus far) to have visited Uranus.

Uranus has an unusual spin in that it rotates from East to West. Furthermore, rather than having a rotation axis perpendicular to the orbital plane, like the rest of the planets, Uranus' rotation axis is tipped on its side (97.8°). This gives the impression that the planet is not spinning but rolling along on the surface of the orbital plane. The unusual rotation is thought to be resulting from a collision with a very large body early in the planet's formation. When a pole points towards the Sun, the whole hemisphere has sunlight throughout that whole season. Each season lasts 21 years, and the pole not bathed in sunlight is having a dark winter without any light at all.

Uranus is a gas planet composed mainly of hydrogen and helium, see figure 4.10 for a Hubble Space Telescope image of the planet. It belongs to a sub-category of gas giants called ice-giants, because 80% of its mass is tied up in materials such as water (H_2O), methane (CH_4), and ammonia (NH_3). Actually, Uranus gets its blue/greenish colour from the methane in its atmosphere, as methane absorbs the red part of the solar spectrum and reflects the blue part back to us.

Uranus is thought to have an inner rocky core (composed of nickel and iron alloys) enveloped by an extended liquid core, which is hot and dense because of the high pressure. The outer liquid core is composed of water, ammonia, and methane.

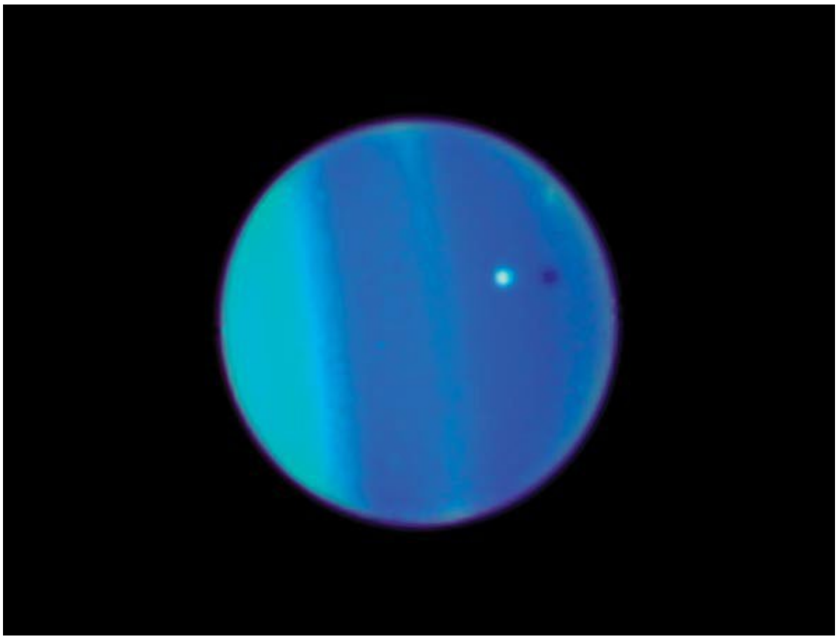


Figure 4.10: Uranus with its moon Ariel in front (pale blue dot) and Ariel's shadow (darker spot to the right of Ariel). Picture by Hubble Space Telescope, courtesy of NASA/Space Science Institute.

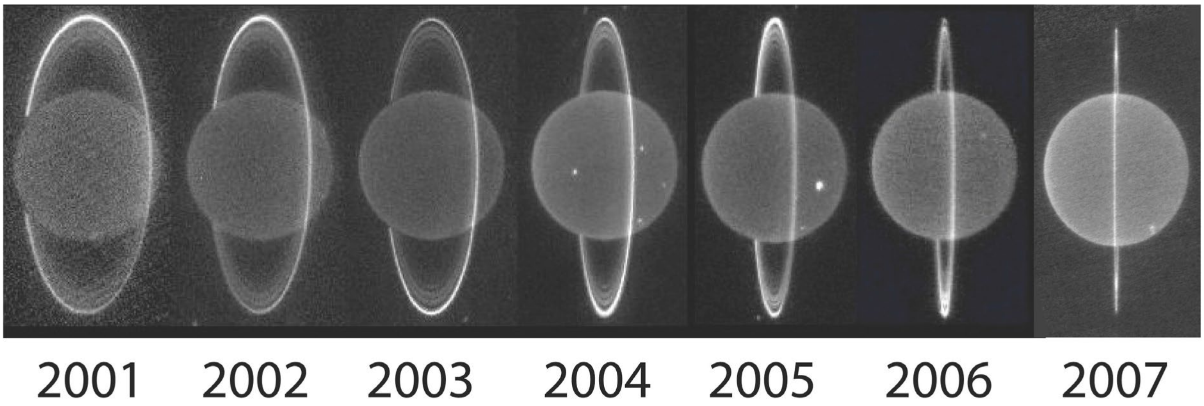


Figure 4.11: Uranus' rings change perspective during equinox. Images by KECK Observatory, courtesy of KECK and NASA.

The pressure in Uranus is not high enough to disassociate hydrogen and thereby create a conductive layer to carry the magnetic field. However, Uranus has a layer, just above its core, consisting of water, methane, and ammonia, which acts as a molten salt. This layer is electrically conductive and the seat of Uranus' magnetic field. The magnetic field axis is tilted around 60 degrees from the rotation axis and it is offset from the centre of the planet.

Uranus has 11 inner rings and 2 outer rings. The rings are aligned with the equator and during equinox we can observe the rings changing orientation with respect to our line of sight, see figure 4.11. Some of the larger rings are surrounded by belts of fine dust.

Uranus has 27 known moons, which are named after characters in Shakespeare's and Alexander Pope's literary works. The five largest moons, also called the inner moons, were the first to be discovered. They are Miranda, Ariel (on the picture in figure 4.10), Umbriel, Titania, and Oberon. The inner moons are thought to be half water ice and half rock. Although they are the largest moons of Uranus, they are rather small. The largest moon Titania has a radius only half of our Moon's radius. Moons with orbits beyond the orbit of Oberon are probably captured asteroids. The moons Cordelia and Ophelia are shepherd moons, that guide the outermost of Uranus' inner rings.

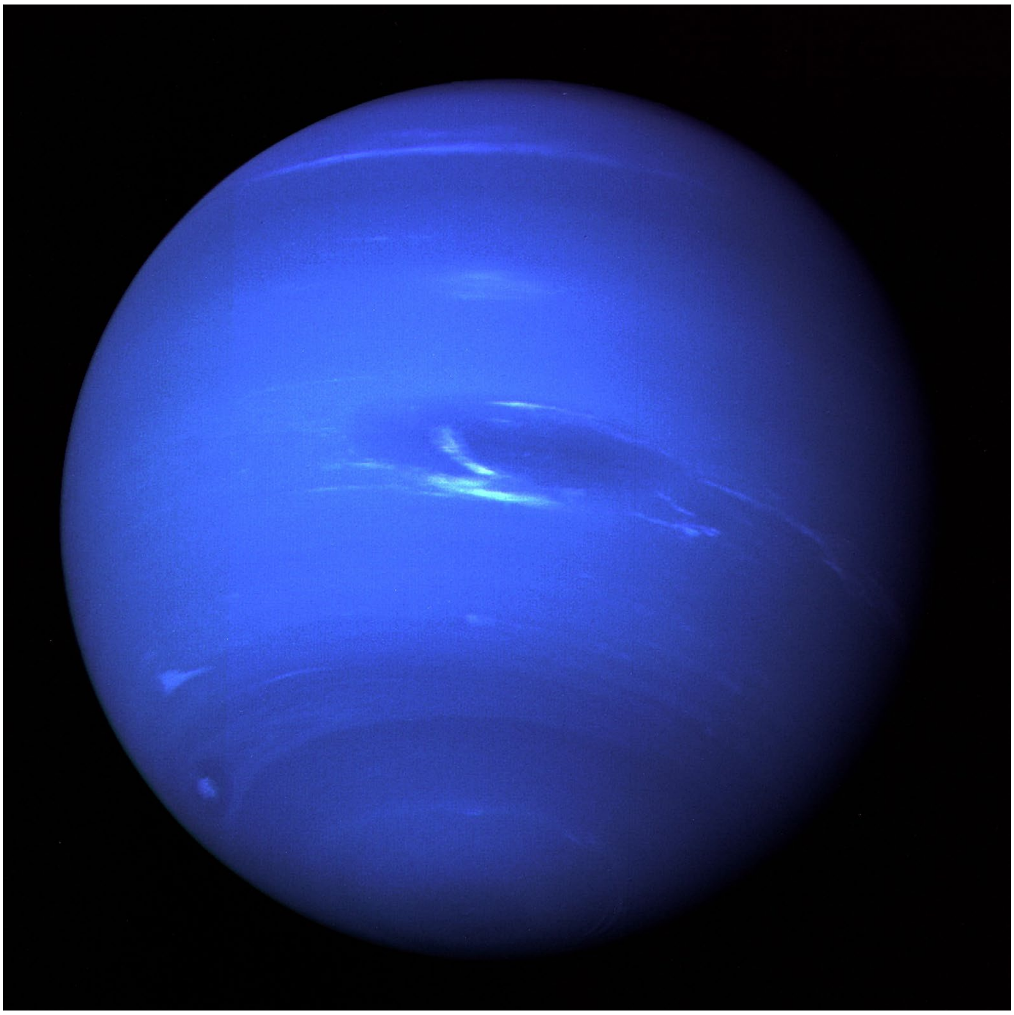
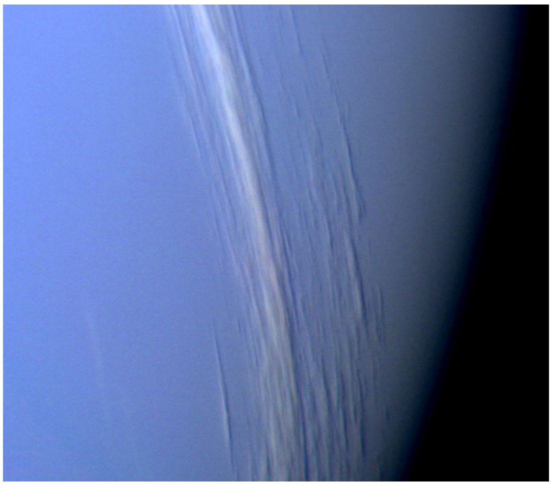
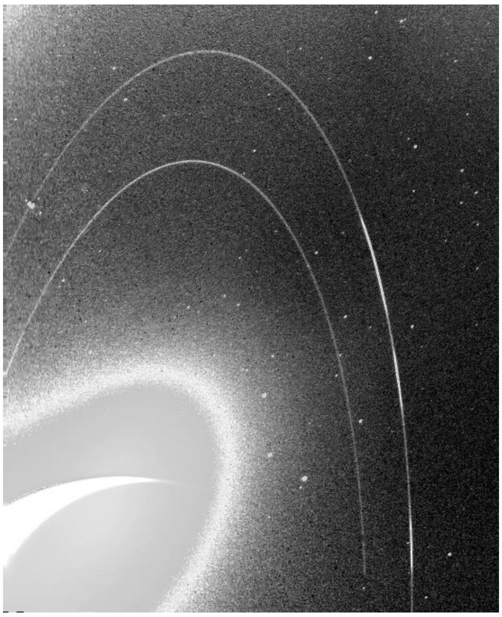


Figure 4.12: Neptune, with its great dark spot. Image courtesy of NASA.



(a) Clouds on Neptune.



(b) Neptune's rings.

Figure 4.13: Images taken by Voyager 2 during the 1989 flyby, courtesy of NASA, NASA Planetary Photojournal, and the Jet Propulsion Laboratory.

4.4 NEPTUNE

Neptune is the eighth planet from the Sun. It is often described as a twin to Uranus as they share many physical features. It takes Neptune almost 165 Earth years to orbit the Sun, which means Neptune has only made one orbit since its discovery in 1846. Neptune's distance to the Sun is about 30 AU, i.e. 30 times further away from the Sun than Earth, and it is therefore impossible to see in the sky without the aid of telescopes.

Neptune is blue due to methane in the atmosphere that absorbs the red light and reflects the blue. There is a slight colour difference between Uranus and Neptune, where Uranus has a green tinge to it. This difference must be cause by a slight difference in the chemical composition. Neptune is an ice-giant and consists mainly of hydrogen and helium, and, like Uranus, methane, ammonia and water.

Neptune has a cloud cover and a gaseous atmosphere, see Figure 4.13a. From the shadows casts by the clouds on the lower atmosphere the altitude of the clouds has been deduced. Neptune also has strong winds, and in Figure 4.12 we see that Neptune has a great dark spot, which is a storm system, much like Jupiter's. The great dark spot on Neptune is larger than Earth, but unlike Jupiter's spot, which has endured for centuries, Neptune's spot has dissipated. Newer storm systems have developed and since then dissipated.

Below the gaseous atmosphere Neptune is warmer and the material there is in a liquid state. The lower part of this liquid layer may be metallic hydrogen, as Neptune is a lot denser than Uranus and therefore has a higher pressure there. Neptune has the highest density of all the jovian planets, leading us to believe that it has proportionally the largest core.

Neptune has a magnetic field, which is generated in the electrically conductive metallic hydrogen and ions of methane and ammonia. The magnetic field axis is inclined with 47° with respect to the axis of rotation. Neptune's magnetic field is 27 times more powerful than Earth's magnetic field.

Neptune has six known rings, see Figure 4.13b. The rings have thicker and thinner regions, giving the ring system an arc like appearance. The rings are thought to be relatively young and short lived. Rings are expected to become uniform with time because of the laws of motion, however the presence of the moon Galatea might be the responsible for confining the arcs.

Neptune has 13 known moons and one awaiting confirmation. The moons have been given names from water gods and nymphs in Greek mythology to match the name of Neptune, which is the Roman god of the sea. The three largest moons are Triton, Proteus, and Nereid. Triton orbits Neptune in the opposite direction as all the other moons, which suggest that it is a captured moon and therefore not assembled out of pre-planetary material during the formation of Neptune. The capture of a moon would have disrupted the orbits of the other moons, and could be the reason the moon Nereid has such an eccentric orbit. When Nereid is furthest away from Neptune, it is seven times as far away, as when it is nearest to Neptune.

Triton is the largest of Neptune's moons, it is in fact as large as the dwarf planet Pluto. It is extremely cold on Triton (–235° C), probably because its icy surface reflects almost all of the sunlight that reaches it. Ice geysers or volcanoes eject a mixture of liquid nitrogen, methane and dust up to 8 km into the air, where it freezes and subsequently snows back onto the surface. Because of Triton's retrograde orbit it is being slowed down by Neptune's tidal forces. As the moon slows down it will drop closer to Neptune, where it at some point (million years from now) will be torn apart by tidal forces.

Neptune and its moons have only been visited by the Voyager 2 spacecraft (1989), and many questions remain unanswered.

4.5 SUMMARY OF THE JOVIAN PLANETS

The physical parameters for the outer 4 planets have been summarised in Table 4.1.

Bibliography and Links

The source of most of the information in this chapter came from NASA, which has a website for each body in the Solar System:

	Unit	Jupiter	Saturn	Uranus	Neptune
Mass, M Mass, M	[$M_\otimes$] [kg]	317.828 1.90×10^{27}	95.161 5.68×10^{26}	14.536 8.86×10^{25}	17.148 1.02×10^{26}
Radius, R Radius, R	[$R_\otimes$] [km]	10.97 6.99×10^4	9.14 5.82×10^4	3.98 2.54×10^4	3.86 2.46×10^4
Density, ρ Density, ρ	[$\rho_\otimes$] [g/cm^3]	0.241 1.326	0.125 0.687	0.230 1.270	0.297 1.638
Inclination	[degrees]	3.1	26.7	97.8 R	28.3
Eccentricity		0.04838624	0.05386179	0.04725744	0.00859048
τ rotation	[$Day_\otimes$]	0.41467	0.445	0.72	0.67
Dist. to Sun Dist. to Sun	[AU] [km]	5.2 7.78×10^8	9.54 1.43×10^9	19.19 2.87×10^9	30.07 4.50×10^9
τ orbit τ orbit	[$Year_\otimes$] [$Day_\otimes$]	11.86 4,333	29.45 10,756	84.01 30,687	164.79 60,190
Surface Temp.	[K]	125	95	57	59
Moons	[number]	79	53	27	13
Rings	[number]	3	7	13	6

Table 4.1: Facts and stats of the outer 4 planets, numbers from NASA.
$\otimes$: In Earth units.
Inclination of the planet with respect to ecliptic. R=retrograde rotation.
Eccentricity, e, as defined in Chapter 1.5 about elliptical orbits.
τ rotation is the time it takes the planet to spin on its axis once.
Dist. to Sun is the average distance to the Sun.
τ orbit is the time it takes the planet to complete one orbit around the Sun.

http://solarsystem.nasa.gov/planets/jupiter
http://solarsystem.nasa.gov/planets/saturn
http://solarsystem.nasa.gov/planets/uranus
http://solarsystem.nasa.gov/planets/neptune

Additional reference (about Neptune's rings):

Murray, C. D. & S. F. Dermott, 1999, 'Solar System Dynamics', Cambridge University Press.

5 THE PHYSICS OF LIGHT

Your goals for this chapter are to learn about:

- The connection between temperature and movement.
- The Planck Curve and blackbody radiation.
- The photon.
- Emission and Absorption Lines.
- Doppler Shift and Doppler Broadening.

Almost everything we know about the Universe, we know from analysing the light we receive here on Earth. Light consists of photons, which is often referred to as electromagnetic radiation. To understand the Universe it is vital that we understand the physics governing electromagnetic radiation.

Everything in the Universe emits photons because of its temperature, that type of light is called thermal radiation.

Additionally, photons are also emitted when an atom (most often an elec tron associated with an atom) changes energy state from a higher to a lower state. The energy carried away by the photon corresponds with the energy difference between the two states.

Finally photons are also emitted under other special circumstances, like when a charged particle moves in a magnetic field (synchrotron radiation), or when a charged particle is decelerated by another charged particle (free-free emission/Bremsstrahlung), there it is the loss in kinetic energy that corresponds with energy of the emitted photon.

5.1 THERMAL RADIATION

Heat and temperature is movement on an atomic scale. The hotter something is the faster its molecules and atoms vibrate and move around. This is easily seen when we look at a gas in a container. There the temperature of the gas dictates the speed of the molecules and consequently the pressure of the gas. The pressure of the gas is really the molecules hammering into the sides of the container, and with increased temperature, i.e. increased speed, the force the gas exerts on the container walls increases.
The molecules move around randomly and they collide not only with the walls of the container, but also with each other, thus transferring momentum. For an ideal gas, temperature is proportional to the average kinetic energy of the random motions of the molecules.

Imagine water through the two phases ice and liquid. When water is frozen the water molecules are stuck in a lattice. The movement and vibration of the molecules are slow and the temperature is low. As a liquid the molecules can move around more freely and the average speed of the molecules is higher and so is the temperature.

5.1.1 TEMPERATURE

When we talk about temperatures, we will often be using the Kelvin scale, with the unit being Kelvin (K). The coldest theoretical temperature is absolute zero (0° K), at which the thermal motion in matter would be zero. However, an actual physical system or object can never attain a temperature of absolute zero. Not even space is that cold. Cold dark space is around 2.78 K, which is the heat signature of the Big Bang. On earth we have created gasses colder than this, by using lasers to trap the atoms so they can barely move, and thus are almost at absolute zero. Absolute zero is denoted as 0° K on the Kelvin scale, which is –273.15° C on the Celsius scale, and –459.67° F on the Fahrenheit scale.

Everything that has a temperature above absolute zero will be emitting energy in the form of electromagnetic radiation i.e. photons. The amount of energy emitted depends on its temperature. The hotter an object is the more energy it emits. Something really hot will emit more energy and something cold will emit less energy.

When two systems are in thermal contact and they have the same temperature no heat transfers between them. When one is warmer than the other, heat will flow from the warmer system to the colder, until they both have the same temperature. When that happens, we say that the systems are in thermal equilibrium. The heat is transferred via photons (thermal radiation) or direct transfer of momentum between molecules (or atoms or particles).

5.1.2 BLACK BODY RADIATION

When we talk about temperatures in astronomy, we often assume that the object we look at is a black body. Black is the colour that absorb the best, i.e. it reflects the least. A perfect black body absorbs all radiation that falls on it, and all emission from it comes from itself, because of its temperature. This emission is referred to as black body radiation. A black body is in this context not necessarily black, it will have a colour according to its surface temperature. Stars are good black bodies.

In 1879 Josef Stefan experimentally found a relation between energy emitted from an object and the object's temperature. 5 years later Ludwig Boltzmann derived this relation analytically by making some basic assumptions. The law is now called the Stefan-Boltzmann law, and it looks like this:

$$F = \sigma \times T^4, \tag{5.1}$$

where F is the energy flux emitted with the unit: $[Js^{-1}m^{-2}]$. T is the surface temperature in Kelvin and σ is the Stefan-Boltzmann constant. From the Stefan-Boltzmann law we see that any body with a temperature above 0° K will emit radiation.

An object emits radiation over a wide range of wavelengths, but there is always a particular wavelength (λ_{max}), at which the emission is the strongest. For stars it is this wavelength that gives it its colour. In 1893 Wilhelm Wien discovered a simple relation between the temperature, T, of a black body and the dominant wavelength λ_{max}. The relation is called Wien's law:

$$\lambda_{max} = \frac{0.0029}{T}. \tag{5.2}$$

Hotter bodies will emit the more energetic short wavelength photons, and colder bodies the less energetic longer wavelength photons. A body at room temperature will emit most of its radiation in the infrared, and an object, which emits most visible light has a surface temperature of a couple of thousand Kelvin.

In the year 1900 Max Planck found a description of black body radiation depending on the surface temperature of the black body. Rather than looking at the total energy emitted (all wavelengths combined), like the Stefan Boltzmann law, Planck found a wavelength dependent energy output law. The spectral intensity of a black body, B, describes the amount of energy it gives off as radiation at different wavelengths:

$$B_\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1}, \tag{5.3}$$

where T is the surface temperature, k_B the Boltzmann constant, h the Planck constant, and c the speed of light.

Planck's law can also be expressed in frequency. The spectral intensity of black body radiation with a surface temperature of T at a specific frequency ν is given by Planck's law:

$$B_\nu = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/(kT)} - 1}. \tag{5.4}$$

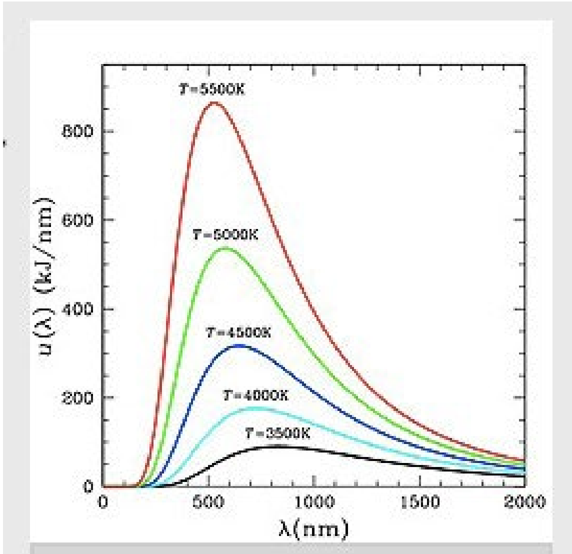


Figure 5.1: Planck function showing how black body emission varies with wavelength. Curtesy of Dr. Sten Odenwald for NASA.

Planck's law is also called the thermal radiation law. Both Wien's law and Stefan-Boltzmann's law can be derived from Planck's law.

A spectrum is a plot of the intensity or emission versus wavelength. Figure 5.1 shows the Planck law for black bodies with different temperatures. We call this resulting shape of the emission spectrum a black body curve or the Planck curve. In the figure it is clear that stars with different temperatures peak at different wavelengths. Red stars are not as hot as blue stars, and our yellow Sun lies somewhere in between. Furthermore, the area under the curve is the total energy output for a blackbody, and there is a clear relation that hotter bodies have a larger total output.

It is important to note that the blackbody spectrum is a continuum, mean- ing there is emission at all wavelengths continuously. Later we will discuss spectral lines, which occur at very defined wavelengths, and therefore display a discrete spectrum.

5.2 QUANTUM DEFINITION OF LIGHT

Light is often referred to as electromagnetic radiation. That is because the light particle, the photon, is the particle that communicates the Electromagnetic Force, i.e. a photon is a boson. When an electron moves in a magnetic field or changes energy level it emits a photon. By analysing the energy of the photon, that is its wavelength or frequency, we can learn about what circumstances lead to the photon being emitted. A photon is massless, the only other massless particle is the gluon (the boson communicating the strong force).

A very fundamental property of matter is that it will always seek the lowest energy state. An atom or a molecule can be in an excited state (i.e. a higher energy state) by having a vibration or the extra energy can be carried by electron(s). The atom or molecule will spontaneously enter a lower energy state and in the process emit a quantum of energy corresponding to the energy difference between the start and end state. That quantum of energy is a photon, and the energy can be calculated as:

$$\Delta E = \frac{h \times c}{\lambda}, \tag{5.5}$$

where h is the Planck's constant, which is the fundamental constant of quantum theory that determines the scale of the small-scale world, and λ is the wavelength of the photon and c the speed of light -a constant in vacuum.

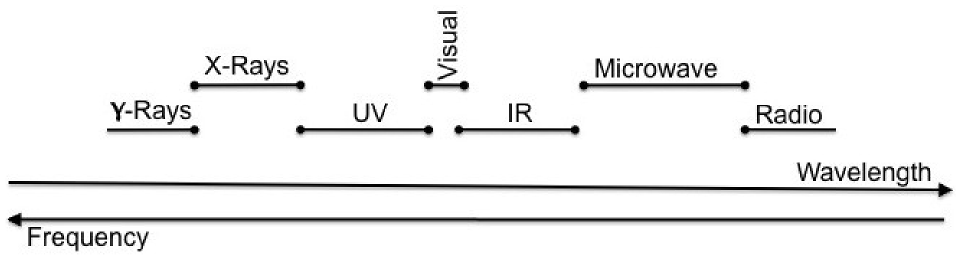


Figure 5.2: The electromagnetic spectrum and the different wavelength areas with respect to each other.

When we talk about a photon being absorbed or emitted, we describe it is a particle. When we describe its energy we look at it as if it was a wave with a certain length. The truth is that a photon shows both wavelike and particle like features. When we describe light as a wave, we can assign a wavelength for the light, λ. The length of a wave is the distance between wave-peaks. The frequency, ν, of light is the number of wave peaks passing a given point in one second, therefore the unit is per second or s^{-1}. λ and ν relates to each other:

$$c = \lambda \times \nu \tag{5.6}$$

The shorter the wavelength is, the higher the frequency is, because more wave peaks pass by per second. The energy formula 5.5 can then be express as:

$$\Delta E = h \times \nu = \frac{h \times c}{\lambda}, \tag{5.7}$$

We see from this formula that as the wavelength gets longer the energy gets smaller, and conversely smaller wavelength corresponds with higher energy. The energy-wavelength relation means that radio waves, which have wavelengths between 10 and 30 meter have really small energies and photons with shorter wavelength, like X-rays, are more energetic. If it feels counterintuitive that photons with small wavelengths have large energies it helps to look at the frequency instead.

Generally photons can be absorbed and therefore detected by objects of the same size as the wavelength. A photon with an energy corresponding to a radio wave will pass over you. The human body is too small to be seen by the photon. In order to catch a radio wave you need something really long, hence the length of radio antennas. Conversely, a photon with a high energy and thus a short wavelength (for example an X-ray photon), travels through almost everything. Our skin is not dense enough to stop it, but our bones are.

Microwave-ovens use light with wavelength of the order of a micrometer, μm (10^{-6}m), to heat up your food. Light with that wavelength is easily absorbed by water molecules, which increases the internal energy. With the extra energy the molecules vibrate and the temperature has been increased. Mobile phones send and receive light signals with wavelengths around 333- 375 μm (or a frequency ν ≈800-900 GHz).

If the wavelength of light is between 375 and 780 nm (10^{-9}m), then we call it visual light and our bodily receptors can detect it, i.e. we can see it with our eyes. Bees and other pollinating insects (and some birds) have their vision extended into the ultraviolet, UV. That means they can also see shorter wavelengths. Many flowers vying for the attention of pollenating insects or birds light up in UV, when they are ripe with pollen. With night googles we can see further into the infrared, IR, enabling us to see heat signatures.

5.3 SPECTRAL LINES

When an atom in an higher energy state decays to a lower energy state it will emit a photon. Similarly, when an atom absorbs a photon, it will go from a lower energy state to a higher energy state. The difference between the lower and higher energy states, in both cases, is exactly the energy of the photon.

In astronomy we look at emission graphically by plotting the amount of energy versus wavelength, this is called a spectrum. The most commonly known spectrum is that of the Sun. In a spectrum a spectral line stands out as a very defined wavelength with an abundance of emission (or a deficit of emission). A spectral line is either an emission line (an emitted photon) or an absorption line (an absorbed photon).

The photon emitted or absorbed has an energy that exactly corresponds with the atom or molecule that emitted it. That implies that emission and absorption give rise to a spectrum of many discrete values rather than a continuum like that from the thermal emission. Like a fingerprint, there is a limited set of energy values for photons which belong to each atom or molecule. Figure 5.3 shows the line spectrum of Hydrogen, Carbon and Helium, which clearly illustrates that each set of lines are unique.

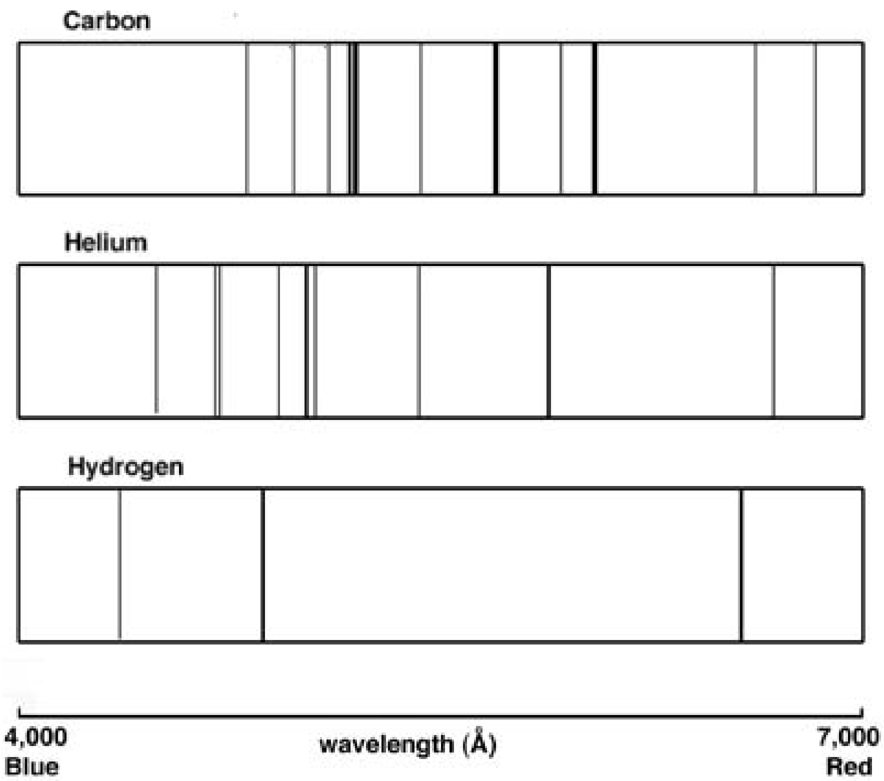


Figure 5.3: Spectral line 'fingerprint' of Carbon, Helium and Hydrogen. Image courtesy of NASA

5.3.1 EMISSION LINES

Hydrogen is the most simple atom, which consists of a proton in the core and one electron bound to it. When energy is added to the hydrogen atom the extra energy is carried by the electron, which then becomes less bound. When the atom decays from this higher energy state it emits a photon and the electron becomes more tightly bound. We say an atom is in the ground state, when all the electrons associated with it are as bound as they can get.

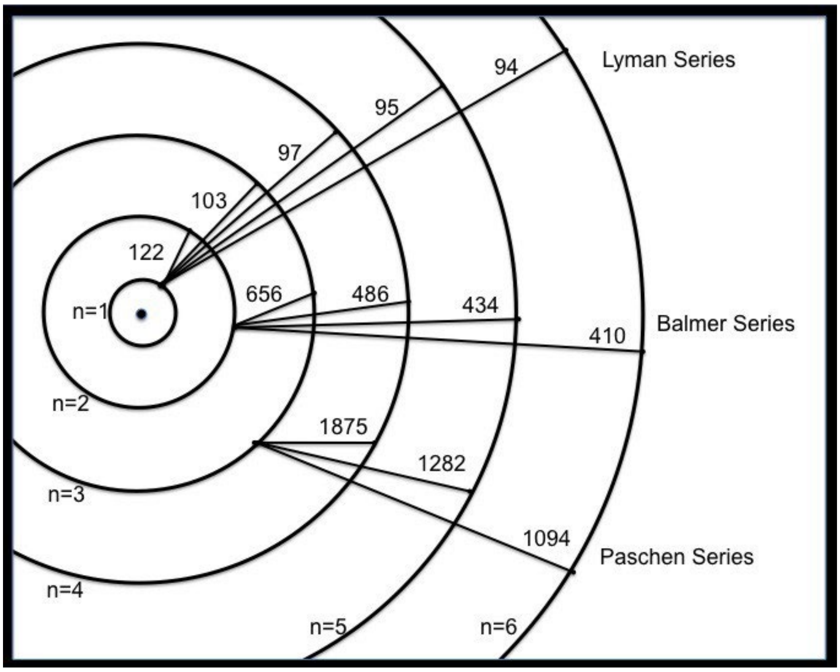


Figure 5.4: Energy levels in the hydrogen atom, each level denoted with an 'n' and the corresponding quantum number. Each series of transitions ends at the same energy level, but originates at different levels. The numbers next to each transition are the wavelength (in nm) of the photon corresponding to energy difference between the two levels bridged by the transition line.

Figure 5.4 shows a simplified model of the hydrogen atom. In this model the electron's allowed states are shown as concentric rings around the core, often these rings are referred to as shells, but keep in mind it is still at simplified model of reality. Each 'shell' in the atom is given a quantum number n, where n=1 is the innermost shell, and therefore the energy state in which the electron is most tightly bound. The electron can jump between shells, and when it jumps inwards, i.e. going from a shell with a higher number to one with a lower number, it will emit a photon. The energy of the emitted photon is related to the wavelength, see formula 5.5.

An emission line spectrum is a discrete spectrum having only emission at very defined wavelengths, Figure 5.3 illustrates this point well. The emission line spectrum for helium looks different than that for hydrogen, because the core of the helium atom boasts two protons (positive charged particles). Therefore, the electrons (negative charged particles) in the helium atom have a binding energy different from that of the electron in the hydrogen. Consequently, the emission spectrum for helium is different from the emission spectrum for hydrogen. Since the number of protons in the core is unique for each element, we get a unique emission spectrum for each element. The same is true for molecules, i.e. two or more atoms bound together.

5.3.2 ABSORPTION LINES

A photon can only be absorbed by an atom (electron) if the photon interacting has the right energy, i.e. wavelength. Absorption will not happen if the photon to be absorbed is too energetic or not energetic enough. However, that does not make absorption something highly unlikely. Blackbody radiation, as discussed previously, emits radiation at all wavelengths, and the most predominate source of light are stars, which are essentially emitting as blackbodies. The implication here is, that for absorption to take place, there needs to be an energy source emitting the photons in the first place.

Absorption in the simplified shell model shown in Figure 5.4 is when the electron jumps from an inner shell to an outer one, i.e. to a higher quantum number. In order to become less tightly bound the electron needs extra energy, which it gets from an absorbed photon.

Absorption is really emission just in reverse. Therefore, when we see absorption in a spectrum it is the lack of emission at a given wavelength, that alludes to the fact that the photons have been absorbed.

5.3.3 KIRCHHOFF'S LAWS

In the 1860s Gustav Kirchhoff summarised spectra in 3 important state- ments, which we today call Kirchhoff's 3 laws:

- Law 1: "A hot opaque body, such as a hot and dense gas produces a continuous spectrum".
  This law is about thermal radiation (black body radiation). The object has to have a high density (i.e. be thick or opaque) in order to warrant thermal radiation. A hot and dense object emits photons at all wavelengths and we say that the spectrum is continuous. There are no holes or breaks in the spectrum and there are no singular spectral lines.
- Law 2: "A hot transparent gas produces an emission line spectrum".
  This law is about emission lines. An emission line spectrum has only emission at very specific wavelengths, typically corresponding to energy transitions in atoms. The gas has to have a low density (i.e. be thin or transparent), otherwise it would be emitting thermal radiation as well.
- Law 3: "A cool, transparent gas in front of a source of continuous spectrum produces an absorption line spectrum".
  This law is about absorption lines. An absorption spectrum is a continuous spectrum, where specific wavelengths are missing due to absorption by the medium the light had to travel through. The absorbing gas has to be cold, or it would be causing emission lines itself. The absorbing gas has to have a low density (i.e. be thin or transparent) or it will also emit a thermal continuous spectrum.

Kirchhoff's laws are illustrated in Figure 5.5.

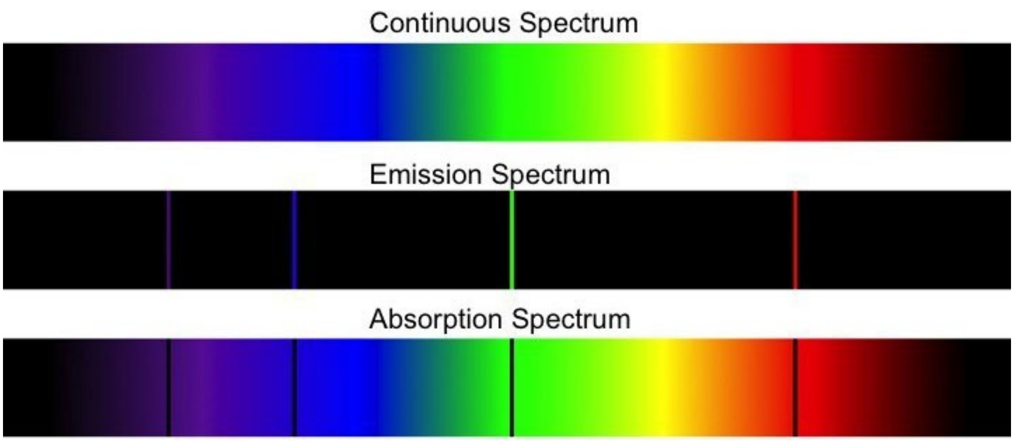


Figure 5.5: A continuous spectrum looks like a complete rainbow of colours without any spectral lines. An emission line spectrum looks like a series of bright spectral lines against a dark background. An absorption spec- trum looks like the continuous spectrum, but where specific wavelengths are missing.

5.4 OTHER CAUSES FOR EMISSION

Besides thermal radiation, emission, and absorption lines, there are a few other radiation mechanisms. The three most important ones are:

- Fusion: When atomic nuclei meld together, there is often an energy release in the form of a photon. This is explained in more detail in Chapter 6.1 about energy production in the Sun.
- Fission: When larger atomic nuclei split up under the release of energy in the form of photons. This is the driving mechanism behind nuclear power plants and nuclear weapons.
- When a charged particle is accelerated it will emit photons and consequently its energy will change. This is seen in several different circumstances described below.

5.4.1 SYNCHROTRON RADIATION

Synchrotron radiation occurs, when a particle or atom with an electric charge is accelerated by a magnetic field. Charged particles moving in a magnetic field are forced to move along certain curves. As seen from the direction of the magnetic field the particles actually move in circles. We know from our basic celestial mechanics (Chapter 1.5.2) that this circular motion implies that the particles are accelerated. The accelerated charge will emit radiation, and this particular radiation we call synchrotron radiation. Often this is observed in connection with strong magnetic fields, like around neutron stars.

Synchrotron radiation is observed as X-ray photons, where the radiation maps out the magnetic field of the object. A good example of synchrotron radiation is the Crab nebula, which is a supernova remnant, see Figure 5.6. Here the powerful magnetic field, which is wound up around the neutron star (the remnant core of the star that exploded), causes synchrotron radiation X-rays. The X-rays are also absorbed by the outer part of the supernova remnant, which powers emission there but at different wavelengths.

In astronomy we often deal with extremes. Masses, energies and magnetic fields are bigger and stronger than anything on Earth. Synchrotron radiation is actually the relativistic case of cyclotron radiation. When the electrons are moving at relativistic speeds, cyclotron radiation is known as synchrotron radiation. The name 'Cyclotron' refers to a type of particle accelerator used in physics research.

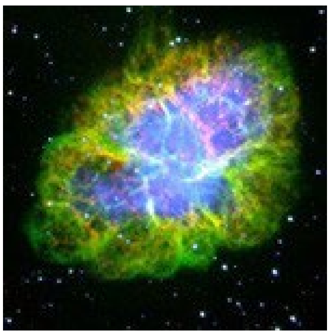


Figure 5.6: The Crab nebular in optical (left) and X-ray (right). The X-rays are emitted because of synchrotron radiation, and therefore trace out the magnetic field. The extend of the X-ray image is only 40% of that of the optical image. Courtesy of NASA.

5.4.2 BREMSSTRALUNG

Bremsstralung is German for braking radiation. As with synchrotron radiation the cause of the radiation is when a charged particle is accelerated. Braking is essentially negative acceleration and it happens to a charged particle when it is deflected by another charged particle. Often an electron is the particle, which is deflected and decelerated by an atomic nucleus. The moving particle loses kinetic energy, which is converted into a photon, thus satisfying the law of conservation of energy. The physical conditions most common for this type of radiation is when the temperature is $T > 10^6$K.

5.4.3 FREE-FREE EMISSION

Bremsstrahlung emitted from plasma is sometimes referred to as free-free radiation. This refers to the fact that the radiation in this case is created by charged particles that are free, i.e. not part of an ion, atom or molecule, both before and after the deflection (acceleration) that caused the emission. Free-Free emission or Bremsstrahlung has a continuous spectrum.

5.5 ANALYSING THE LIGHT

Now that we understand the most common radiation mechanisms we can analyse the light we receive here on Earth. Analysing a spectrum is one of the most powerful tools astronomers have. We look closely at the lines emitted and deviations from the norm are interpreted into physical conditions at the origin of the light. Below are the most important physical conditions, which can be inferred directly from a spectrum.

5.5.1 DOPPLER EFFECT

When a light emitting atom (or particle or molecule) is at rest, the light it emits has a very defined wavelength, $\lambda_{\rm rest}$. If the atom is moving the wavelength of the light will be shifted with the amount $\Delta\lambda$. The shift, also called a Doppler shift, in wavelength relates to the velocity of the atom emitting the light, v_r:

$$\frac{\Delta\lambda}{\lambda_{\rm rest}} = \frac{v_r}{c}, \tag{5.8}$$

where c is the speed of light. The shift $\Delta\lambda = \lambda_{\rm obs} - \lambda_{\rm rest}$, where $\lambda_{\rm obs}$ is the observed wavelength.

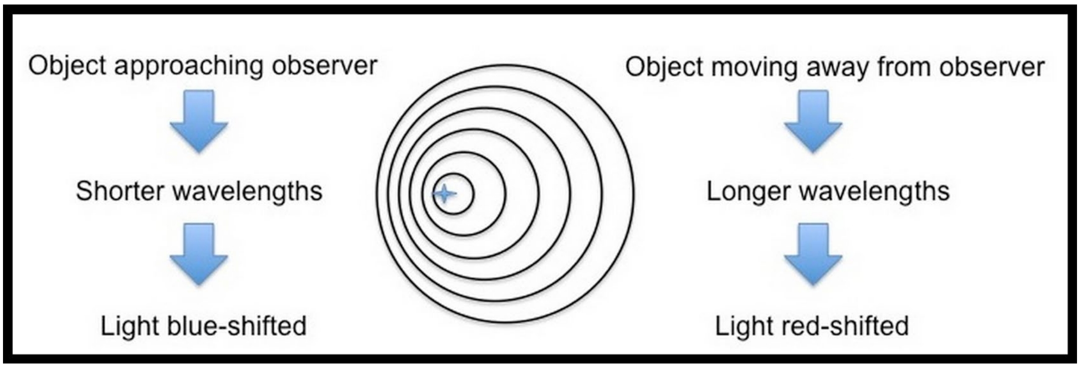


Figure 5.7: Doppler shift illustrated.

The wavelength shift is so that when an object is approaching the light is blueshifted (shorter wavelengths) and when an object is receding the light is redshifted. See Figure 5.7 for an illustration.

Measuring the doppler shift we can find velocities of objects along the line of sight (radial velocity, v_r), combined with their velocity across the sky we have the total spatial direction and speed of objects. This technique is very useful and with it we can establish the movement of stars in our galaxy, galaxies in groups, and with that realise the large scale structures in the universe. With the Doppler shift our simple 2D pictures turn into 3D dynamic views.

Because of Earth's movement around the Sun, and the Sun's movement around the centre of the Milky Way, everything we observe is shifted with an additional factor corresponding to our total movement.

5.5.2 DOPPLER BROADENING

While the Doppler effect causes lines to be shifted it also causes lines to be broadened. This broadening is called rotational broadening and occurs, when the spectrum we get is of both receding and approaching sides of a rotating object.
Stars rotate and when we look at the spectrum of one, an emission or absorption line is the sum of all the emission/absorption in view. That means that we see both the the part of the rotating star that rotates away from us and the part rotating towards us. In total it means the line is broad. The broadness of the line corresponds with the rotation. The same is the case when we look at a spectrum for a galaxy, where we then can infer the rotation of the galaxy.

Additionally to Doppler broadening there is also Natural broadening and collisional broadening. Natural broadening comes from quantum mechanics and is a consequence of Heisenberg's uncertainty principle. Collisional broadening is due to the random motions in the gas, on the surface of the star, which is temperature and pressure dependent.

5.5.3 ZEEMANN EFFECT

The Zeeman effect is the splitting of a spectral line into several components due to a magnetic field. In the atom, the electron can only have pre-determined specific energies. In the presence of a strong magnetic field (for example in sunspots) the allowed energy states are split into several states, causing us to observe a splitting of the spectral line.

Bibliography and Links

The physics described in this chapter is well established and originally published in works over half a century old.

For the advanced reader interested in atoms and energy levels I recommend the book:
Bransden & Joachain, 1984, 'Physics of Atoms and Molecules', published by the Longman Pub Group.

The best resource for accurate wavelengths for different elements is the NIST Atomic Spectra Database:
http://www.nist.gov/pml/atomic-spectra-database

6 THE SUN

Your goals for this chapter are to learn about:

- The structure of the Sun.
- Energy production in the Sun, i.e. fusion.
- Energy transfer in the Sun.
- Surface phenomena of the Sun.
- The Solar Wind.

At the centre of our solar system is our star, the Sun. It is a hot ball of gas, whose energy production influences all the planets. The Sun's energy is essential for our lives and yet, in the larger picture, the Sun is insignificant. There are billions of stars, just like the Sun, in our galaxy, the Milky Way, alone. Furthermore, the Milky Way is one of countless galaxies spread out across the Universe.

Stars (and our Sun), are made up of plasma, which is a state of matter where the electrons are stripped from the atoms. States of matter, like solid, liquid and gas are all temperature dependent, and the disassociation of the electrons is a direct cause of the high temperatures in stars, which make the atoms fully ionised.

The Sun is made up of 75% hydrogen and 25% helium, with an additional 0.1% of heavier elements like carbon, nitrogen, iron etc. Consequently the Sun consists mainly of hydrogen and helium plasma.

The Sun is a medium sized star of the type 'yellow dwarf'. It is 4.6 billion years old, and thus halfway through its life, as it will shed its outer layers and turn into a white dwarf in another couple of billion years. The Sun, and in fact any star, spends most of its life in equilibrium. The gravitational pull that forces all the matter to move towards the centre is perfectly balanced by the outward pressure of the heat generated in the centre.

The conditions for internal equilibrium in a star can be expressed mathematically as 4 equations governing the distribution of mass (density), gas pressure (temperature), energy production (temperature dependent), and energy transport (radiation vs. movement of material). It is far beyond the scope of this book to go into the details of these differential equations, but the causes and implications of them are discussed in the following.

6.1 ENERGY PRODUCTION IN THE SUN

Fusion, which is the way the Sun produces energy, can only take place when the temperature is high enough. Failed stars are not heavy enough for gravity to push material enough together, and thereby reach the temperature needed. Those failed stars are called brown dwarfs. The dividing line between brown dwarfs and giant gas planets, like Jupiter, is blurred.

Fusion is when two lighter elements are fused together into a single heavier element. The Sun melds hydrogen cores into Helium cores. Remember, the temperature here is so hot that the electrons are disassociated from the nucleus of the atoms. A hydrogen core is essentially just a proton, and the fusion process in the Sun is often referred to as the proton-proton chain (or just pp-chain).

About 91% of the energy produced is via the proton-proton chain, which has the following steps:

$$^{1}\mathrm{H} + {}^{1}\mathrm{H} \longrightarrow {}^{2}\mathrm{H} + \mathrm{e}^{+} + \nu_{\mathrm{e}} \tag{6.1}$$

$$^{1}\mathrm{H} + {}^{1}\mathrm{H} + \mathrm{e}^{-} \longrightarrow {}^{2}\mathrm{H} + \nu_{\mathrm{e}} \tag{6.2}$$

^{2}H is deuterium, i.e. a proton and neutron nucleus. ν_e are neutrinos, where the 'e' shows that it is neutrinos of the electron type that are created. The neutrinos interact very poorly with matter, making it possible for them to escape the Sun right after creation. e^+

is a positive electron (positron), which shortly after its creation will be annihilated when merging with a regular electron and creating two photons. While the first link in the chain, formula 6.1, accounts for 99.75% of the deuterium created, the second link, formula 6.2, only accounts for 0.25%.

In fact, creating deuterium by colliding two protons is not a very successful business. The probability of this reaction is so low, that the expected time for it to happen is about 10^{10} years, even at the high density and temperature in the centre of the Sun. Fortunately, the vast number of particles means that the reaction happens at a reasonable rate. If this reaction was more likely, the Sun would have burned out long ago, which implies that this step is the needed break in the chain, for it to burn moderately and long.

The next step involves fusing deuterium with a proton and creating helium:

$$^{2}\mathrm{H} + {}^{1}\mathrm{H} \longrightarrow {}^{3}\mathrm{He} + \gamma, \qquad (6.3)$$

where γ is the chemical sign for a photon. This reaction is, on the other hand, very likely, which means that the number of deuterium atoms is small. ^{3}He is an isotope, where the nucleus consists of two protons and one neutron. The last step of the chain is:

$$^{3}\mathrm{He} + {}^{3}\mathrm{He} \longrightarrow {}^{4}\mathrm{He} + {}^{1}\mathrm{H} + {}^{1}\mathrm{H}, \qquad (6.4)$$

where two protons are released, which are then free to participate (again) in the first steps of the chain. This last step in the chain accounts for 91% of the helium production. There are two other ways this reaction takes place, one involving lithium and one involving beryllium.

In this fusion of hydrogen to helium the energy is released in the form of photons and neutrinos. To calculate the amount of energy created we will use Einstein's formula:

$$\mathrm{E} = \mathrm{m}\ \mathrm{c}^2, \qquad (6.5)$$

where E is the energy, m the mass, and c the speed of light. The mass converted into energy can be found by looking at the masses of the particles going into the reaction and at masses of the end products. Letting the reaction 6.4 happen twice, the above reactions can be summarised into one simple reaction:

$$4\,{}^{1}\mathrm{H} \longrightarrow {}^{4}\mathrm{He} + 2\,\nu_{\mathrm{e}} + \gamma, \qquad (6.6)$$

where γ refers to several photons. In fusion the mass of the end product is lighter than the start products put together. The mass of the end nucleus is smaller than the sum of the masses of all its constituent particles separately. This difference is accounted for in the binding energy of the nucleus. The binding energy per nucleon (nucleus particle) increases towards heavier elements until iron, meaning you can gain energy by fusion of all elements lighter than iron. After iron the nucleus is so large that you gain energy by splitting the nucleus rather than making it larger. Splitting of a nucleus is called fission, and that is what nuclear power plants and nuclear bombs do.

The sun converts 4 million tons of mass into energy every second, but since the Sun is so large it means that in its entire lifetime burning hydrogen it has only converted 0.1% of its mass into energy.

6.2 THE STRUCTURE OF THE SUN

The Sun can be divided into 6 distinct layers or regions each having defining characteristics. At the centre of the Sun the energy creating core resides within 20-25% of the solar radius. Above the core sits the radiative zone, which reaches up to 70% of the solar radius. This is encased in a convective envelope which reaches up to the surface. The atmosphere of the Sun can be divided into two layers: The photosphere and the chromosphere. The final layer is the corona, which we see as the glow around the Sun. Figure 6.1 shows this structure.

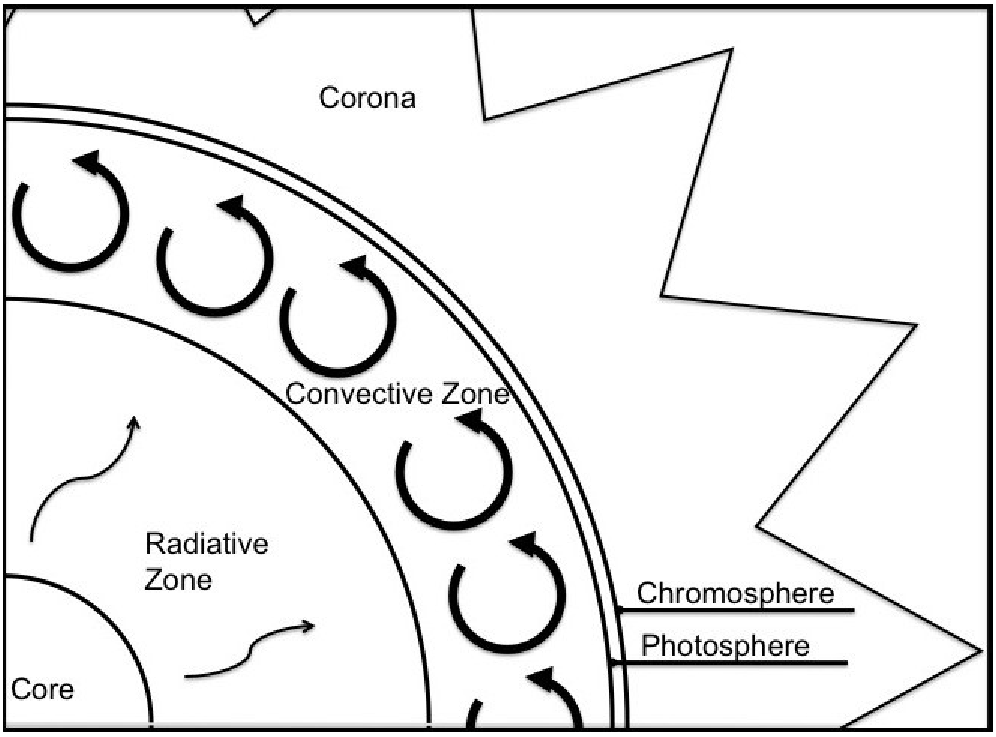


Figure 6.1: The Structure of the Sun.

Since the Sun is not transparent, what we know from its interior structure has been found by comparing observations, such as temperature, energy output, mass etc, with solar models taking all those facts into account. Furthermore the research area of helioseismology contributes to our understanding of stellar structures.

6.2.1 THE CORE

The core is at the centre of the star, it is the place in the Sun, where the pressure, density, and temperature, are at their highest. In the core it is 15 million degrees Celcius, which is high enough for fusion to take place. Actually, the region of the core is defined as the area, where the fusion takes place. The majority of the fusion processes happen within the inner 50% of the core. As the density decreases outwards so does the fusion process and at the edge of the core no fusion takes place.

6.2.2 THE RADIATIVE ZONE

Just above the core the Sun has a layer, which we call the radiative zone. The core releases a huge amount of energy in the form of photons. This energy travels through the radiative layer as photons. However, the photons do not take a straight path. Instead they are emitted in hotter parts of this region and absorbed in cooler parts. Where the photons are absorbed the plasma heats up, becoming the new hotter part, which then emits photons. This process of emission and absorption depends on temperature, density, and chemical composition of a star. For the Sun it is estimated that it takes a photon roughly 170,000 years to make its way through the radiative zone.

6.2.3 THE CONVECTIVE ZONE

The temperature in the star drops steadily with increasing radius. It is at its hottest in the centre and in the convective zone it has dropped below 2 million degrees Celcius. The difference between the radiative zone and the convective zone is the way energy is being transferred. This is temperature dependent and in the convective zone the temperature is so low that energy transport via photons is inefficient. Instead the energy is deposited in the material and thus carried by the particles.

At the boundary layer between the radiative and convective zones the plasma absorbs the photons and increases its internal energy as heat. Now it is the movement of the plasma that carries the energy outwards in the star. Plasma particles move around in groups called cells, which essentially is just a conglomerate, where the member particles have similar temperature and density. These plasma cells (or convection cells) drift upwards (i.e. outwards) if they are hotter than their surroundings. Once they reach the surface the energy is free to radiate away as photons and the plasma cell cools. When a plasma cell is cooler than its surroundings it will sink (or fall) inwards in the star until it reaches the radiative layer, where it sucks up photons and increases in temperature again, which causes it to make another ascent to the surface. This kind of energy transportation is called convection.

Convection causes the material in this layer to be thoroughly mixed. In the radiative zone this is not the case as radiation only moves energy and not material. As plasma is a conductive material, and since it moves around in the convective layer, it creates magnetic fields. These magnetic fields causes several different surface phenomena, which will be discussed later. The convection cells are actually visible at the surface as small bubbles either brighter or darker. Collectively they are called granulation and can be seen in figure 6.2, which shows a real image of the surface of the Sun.

6.2.4 THE PHOTOSPHERE

The photosphere is the lower layer in the Sun's atmosphere and thus also the surface of the Sun. This 400-500 km thick region is where most of the radiation leaves the Sun. The temperature here is about 5, 800° K. The top of convection cells can be seen as granules. The hottest part of a granule is the centre where the colour is also brighter. Around the edge the colour is darker, here the plasma has cooled off and is descending again. A granule is about 1,000 km in diameter and it lasts around 8-20 minutes before it is dissipated.

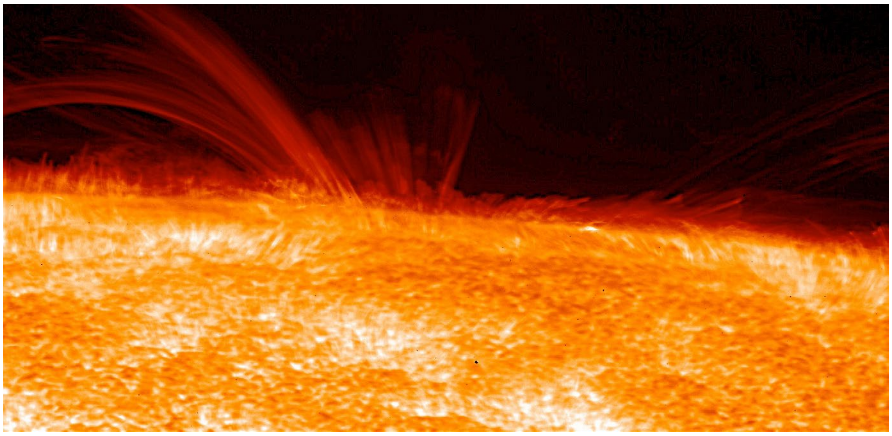


Figure 6.2: The fine scale structure of the Sun's surface taken by Hinode's solar optical telescope. Curtesy of Hinode JAXA/NASA.

As can be seen in the beautiful picture of the Sun's surface by the solar optical telescope Hinode, Figure 6.2, there are many spectacular phenomena here. Sunspots are darker areas on the surface, where a cooled down plasma cell cannot sink downwards, because a magnetic field is blocking its passage. This phenomena is associated with a very strong magnetic field, which can be seen in the spectral lines. Sunspot activity follows an 11 year cycle, which is the overall cycle of the Sun's activity. In the picture we also see a faint prominence, which we call a bridge of plasma, that is flowing along the magnetic field lines. In addition to prominences and sunspots other events like solar flares and coronal mass ejections are also connected to the magnetic activity and cycle of the Sun.

6.2.5 THE CHROMOSPHERE

The granules in Figure 6.2 reside in the photosphere and the red layer just above is the chromosphere. This layer is especially visible when we look at the Sun's horizon in the image. During a total solar eclipse the chromosphere will be visible as a red rim around the Sun. The chromosphere is red because the photons predominately are emitted from a specific transition in the hydrogen atom called Hα (wavelength of 656.28 nm).

The chromosphere reaches from just above the photosphere at 500 km above the surface to around 2,100 km above the surface. In this layer the temperature increases with distance from the surface, just opposite to the inner 4 layers, where temperature decreased with radius. The lowest temperature in the chromosphere is 4,000° K at the boundary of the photosphere, and the highest temperature is at the very top of the chromosphere, where it is measured to be 8,000° K.

6.2.6 THE CORONA

Above the two surface layers is a region of hot plasma called the corona. The corona is too weak to be seen, because the bright photosphere overshadows it. However during a solar eclipse, when the Moon covers the photosphere, the corona can be seen as a white light around the Sun with the shape of a pointy crown. The points of the crown are formed by plasma streaming out from the Sun.

The corona is very hot, about 500,000 - 2 million degrees K, and thus much hotter than the visible surface. Remember from Chapter 5.1 that temperature and heat is movement on an atomic scale. In the corona the particles are moving at high speeds. Plasma is, by definition, atoms stripped of their electrons, and the atoms are in their highest energy states. The physical conditions in the thin corona are extreme with strong magnetic fields and a high flux of photons continuously passing through. Nevertheless, this is not enough to explain the high temperature in the corona. Recent observations (by TRACE) points to a non-uniform heating at the base of coronal loops, which connect sunspots, as the source of the heat in the corona.

6.2.7 THE SOLAR WIND

The corona has no upper bound, it just extends endlessly and contributes to the outward flow from the Sun that we call the solar wind. Material lost this way from the corona is replaced by material from the chromosphere. The mass loss from the Sun into the solar wind is almost insignificant compared to the size of the Sun.

The solar wind is a flux of charged particles, mostly electrons and protons, that could escape the Sun's gravity because of their high velocities. Near Earth the solar wind has a density of about 5-10 particles per cm^3 and it moves with a speed of 500 km/s. However, the speed, density and temperature of the wind vary over time. When the Sun has had more activity than usual we call it a solar storm. The solar wind creates a huge bubble that surrounds the whole solar system, this bubble is called the heliosphere.

6.3 THE ACTIVE SUN

The Sun has many surface phenomena, which all are related to its changing magnetic field. Based on observations of sunspots the Sun's activity has been measured to follow a periodic cycle of around 11 years. The magnetic field is generated in the convective layer, and it might intially be bipolar across the Sun, with a magnetic North at one pole and a magnetic South at the other pole. However, the Sun rotates faster at the equator (27 days) than at the poles (36 days), which causes the field lines to wind up around the Sun like a thread on a bobbin. The magnetic field lines reconnect via sunspots around the surface of the Sun. After the height of activity the Sun quiets down again and starts from scratch, but now with the magnetic poles reversed.

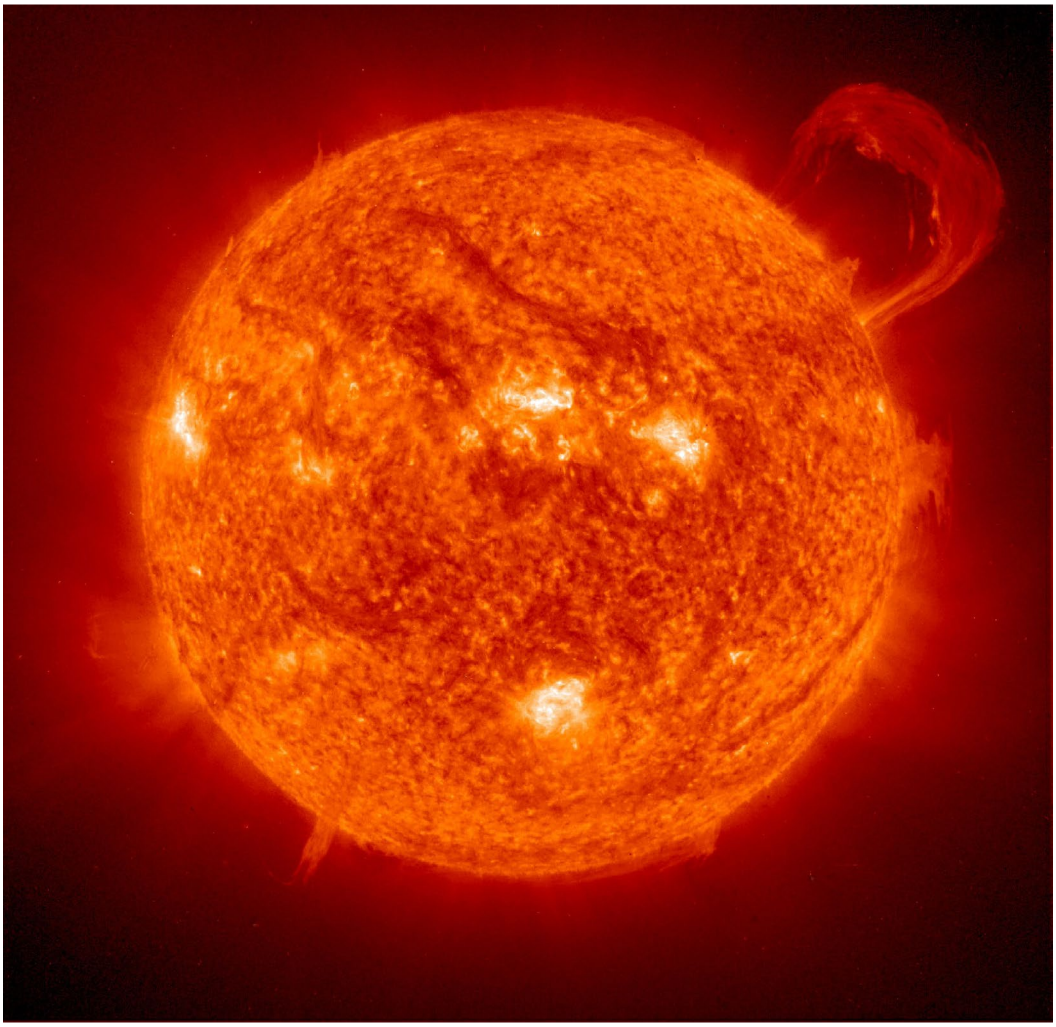


Figure 6.3: The Sun with a prominence about 30 times larger than Earth. Courtesy of NASA and ESA.

In essence a solar cycle is one with a maximum and a minimum. First there is an increased activity, where number of sunspots, flares and solar storms increases to a peak, called a solar maximum, which might last a few years. Afterwards the Sun's activity decreases and eventually rests for a few years with low activity called a solar minimum.

Sunspots are temporary dark areas on the Sun, where concentrated magnetic fields restrict the transfer of energy from below and thus cool the surface. Their diameters range from 16 km to 160,000 km, with the most common size being around 10,000 km. They are typically about 1,500 K colder than their surroundings. Individual sunspots may last anywhere from a couple of days to several months. Eventually, however, they will decay. Often sunspots appear in pairs, where a pair will have opposite magnetic polarity. A sunspot pair will be connected via magnetic field lines, which become visible when plasma flows along them. With plasma flowing along the field lines they appear as glowing loops or streaks.

As can be seen in Figure 6.3 the Sun can have enormous coronal loops (often also called prominences), this one looks almost like a handle on the Sun. The coronal loop is a direct consequence of the twisted magnetic field just below the solar surface. These features are usually associated with sunspots at the base, where they connect with the surface. A prominence hanging in the thin corona gives a stark contrast to the low density there. On the other hand, the dense plasma is relatively cool compared to the hot corona.

Another phenomenon is called a solar flare. That is a burst of radiation at all wavelengths, often accompanied by a cloud of electrons, ions, and atoms, ejected through the corona and into space. Where the light reaches us 8 minutes after leaving the surface of the Sun, the particles typically reach Earth one or two days later. Flares, like all the other events, happen in active regions around sunspots. The driver of a flare is thought to be when a strong magnetic field penetrates the photosphere and link the corona to the solar interior. The magnetic energy stored in the corona is then rapidly released, thus powering the flare. Coronal mass ejections (CME) might be a consequence of the same mechanism, although the relation between CMEs and flares is still not well established.

Coronal mass ejections (CME) are violent events, where solar plasma is thrown away from the Sun. This is the result of long duration solar flares and filament eruptions and thus also a cause of the solar activity. A CME can cause a geomagnetic storm here on Earth, which would be very destructive. The strongest and oldest known geomagnetic storm happened in 1859, also known as the Carrington event, which caused auroras all the way to Sahara and massive disruptions on the telegraph grid. If we had an event today of similar strength all our electronics could potentially be damaged.

Given, the above account of solar activity it is clear that it is vital for us to keep an eye on the Sun's activity. Our solar space missions give us space weather forecasts. The most recent massive CME was in 2012, but luckily for us its direction was not towards Earth. While solar flares are very fast, CMEs are relatively slow.

Bibliography and Links

The Sun has been studied or is currently being studied in great detail by several space missions such as WIND, SOHO, ACE, Solar Terrestrial Relations Observatory (STEREO), and the Solar Dynamics Observatory (SDO (NASA)), and the Parker Solar Probe. The websites for these missions continuously provide updated information about the Sun and its surface phenomena.

NASA also provides many pretty pictures and some informational posters alongside overview facts for the Sun.
http://solarsystem.nasa.gov/planets/sun

The winding up of the Sun's magnetic field is illustrated in this short video (follow the link and choose 'Photosphere (Solar Flare)').
http://alienworlds.southwales.ac.uk/sunStructure.html

For a thorough book on the subject i recommend:
K. R. Lang, 2001, 'The Cambridge Encyclopedia of the Sun', Published by Cambridge University Press.

Generally, topics of solar physic are discussed on an advanced level in this journal: 'Living reviews in Solar Physics'
http://solarphysics.livingreviews.org/

For the advanced reader I recommend Palacios et al.'s detailed paper. It provides additional insight into computer simulations of the Sun's interior, which was only mentioned briefly here. Palacios et al., 2006, 'Dynamical processes in the solar radiative interior', Proceedings of SOHO 18/GONG 2006/HELAS I, ESA SP-624, pp.38.
It can be downloaded for free at:
http://arxiv.org/pdf/astro-ph/0609381v1

7 OBSERVATIONAL ASTRONOMY

Your goals for this chapter are to learn about:

- The positional coordinate system on the sky.
- The magnitude system.
- Luminosity and Flux.
- Distance measurements in astronomy.
- Filters in optical astronomy.
- Stellar classification.

Observational astronomy started in ancient Greece. In modern times the old definitions have been quantified. In this chapter the astronomer's basic tools are explained with an emphasis on the study of stars.

7.1 STELLAR POSITIONS ON THE SKY

Stars and other far away objects are identified by their positions on the celestial sphere. The celestial sphere is a spherical coordinate system with Earth in the centre. Earth's poles are aligned with the poles of the celestial sphere just as the Earth's equator is also aligned with the equator of the celestial sphere.

Coordinates are given in α right ascension, and δ declination.
Declination is given in degrees from the equator (on the celestial sphere), with a negative prefix denoting South of the equator and a positive denoting North. The value thus ranges from -90^{o} to 90^{o}.
Right ascension is given in a time format, as the sphere can be divided into 24 hours, with 60 arc minutes in each, which again span 60 arc seconds. 0 hours right ascension is by convention the right ascension of the sun on the spring equinox (North), i.e. March 21st. The hours are counted towards East.

For example SN 1987A has the coordinates:
α : 05h 35m 28.03 δ : –69o 16' 11.79"
The declination tells me that it is visible from the Southern hemisphere, and the right ascension tells me that the object will be highest in the sky around December. Because of precession (see chapter 1) the coordinates are also given with an epoch (e.g. 1950).

Observatories send out calls for observation proposals either annually or biannually, where astronomers can apply for time on the telescopes. The first step is to match observatory with object, so the calls are often specified with restrictions in declination (position of the telescope on Earth) and in right ascension (a time period, where the objects should be highest in the sky).

7.2 THE MAGNITUDE SYSTEM

The human eye perceives brightness differences logarithmically rather than linearly. That means that if 3 stars A, B and C have brightnesses of 1, 10 and 100, respectively, the difference between star A and B will seem to be just as big as the difference between star B and C.

In 1856 Norman Pogson defined the magnitude system, which we use today. It was based on a system first conceived by Hipparchos, where the brightest stars were counted into the first class (magnitude of 1), the second brightest into second class (magnitude of 2) and so on. A brighter star will have a lower number for its magnitude and a fainter star a larger number. The scale even extends into minus for the really bright objects. The apparent magnitude of the Sun is -26.8, and the brightest star in the sky, Sirius, is -1.5.

Apparent Magnitude

The Magnitude system is a way to attribute a number to the brightness (apparent or absolute) of an object. The apparent magnitude, m, is the observed magnitude and it is related to the flux density, F:

$$m = -2.5 \times \log \frac{F}{F_0}, \tag{7.1}$$

where F0 is the flux density of a star with apparent magnitude m=0. Flux density, often also referred to as just flux, is the amount of energy passing through a square meter per second.

Based on the relation between magnitude and flux it is possible to compare the brightness of two stars, with magnitudes m_1 and m_2 and flux densities F_1 and F_2:

$$m_1 - m_2 = -2.5 \log \frac{F_1}{F_2}. \tag{7.2}$$

Absolute Magnitude

The absolute magnitude, M, is a measure for brightness, when the star is at the specific distance d=10 pc. In other words, if the star is 10 pc away from us its apparent magnitude will equal its absolute magnitude. The absolute magnitude relates to the apparent magnitude, m, from the same star:

$$m - M = 5 \log \frac{d}{10pc}, \tag{7.3}$$

where the distance d is in parsec. This formula is called the 'distance modulus', since it is used to determine distances. If you know a star's absolute magnitude and measure its apparent magnitude you can calculate its distance. The Sun has apparent magnitude m = –26.8 and absolute magnitude M = 4.8.

7.3 FLUX AND LUMINOSITY

The flux, F, is the amount of energy passing through a square meter per second, and thus has the unit: [J/s/m^{-2}]. The flux depends on the surface temperature, T, according to the Stefan-Boltzmann law, see Chapter 5.1.2:

$$F = \sigma \times T^4, \tag{7.4}$$

where σ is the Stefan-Boltzmann constant. The flux is the amount of energy passing through a square meter every second. If two stars have the same surface temperature, they will also have the same flux, regardless of any size differences. That is because flux is only counting the amount of energy passing through a small fraction of the surface.

However, flux is distance dependent. Flux is defined as the amount of energy passing through a surface, further away from the object the less energy will pass through a surface of the same size. The surface of a sphere grows with the radius of the sphere squared (surface area = $4\pi r^2$). Consequently, the flux falls off proportionally with $1/r^2$. This relation is also called the inverse square law.

Luminosity, L, is the total amount of energy radiated by a star every sec- ond. The luminosity of a star depends on its temperature and size. If two stars with the same temperature have different sizes, the larger star will have a larger luminosity, because it has a larger surface area emitting the energy. Because luminosity is all the flux being emitted from a star, the flux is related to the luminosity via geometry. To arrive at an expression for the luminosity, we have to multiply the flux with the surface area of the star.

$$L = 4\pi r^2 \times F, \tag{7.5}$$

where r is the radius in a sphere completely encompassing the star.

The previous statement, that two stars with same temperature but with different sizes will have different luminosities, now makes sense. The star with the larger radius will have a larger luminosity. However, a small hot object can have the same luminosity as a large cold object. The temperature and radius of the stars are determining factors for the luminosity.

By inserting the above expression for the flux into the formula for the Lu- minosity we get:

$$L = 4\pi r^2 \sigma T^4, \tag{7.6}$$

The luminosity is an absolute value for the star, as it does not depend on the distance or any other observer defined restriction.

Where the apparent magnitude is a measure of the star's flux received by us, the absolute magnitude is related to the luminosity. The apparent magnitude is related to flux by introducing a zero point flux, from a star with apparent magnitude of zero. Similarly, the luminosity is brought into the magnitude system by defining a comparison point. The Sun is often chosen as the baseline, since we know its values with great accuracy and since luminosity is often given in units of the solar luminosity, $L_{\odot}$. The absolute bolometric magnitude, Mbol, is the magnitude of a star, where emission at all wavelengths are accounted for. It relates to the luminosity:

$$M_{bol} - M_{bol,\odot} = -2.5 \log \frac{L}{L_{\odot}} \tag{7.7}$$

7.4 DISTANCE MEASUREMENT

A problem in astronomy has always been to determine distances. Is some- thing far away or is it really close but very faint? Is something very close or far away but very bright? Two stars can appear close to each other, when in fact they are not. The constellations on our night sky will undoubtedly look very different when viewed from a different angle.

7.4.1 TRIGONOMETRIC PARALLAX

Parallax is a distance measurement technique, that can be used for foreground stars. In essence we make use of the slightly different perspective we have, when we are on one side of the Sun as opposed to the other. A picture is taken now, and another again half a year later, when the Earth is on the other side of the Sun. The linear distance that Earth has travelled is thus 2 times the distance to the Sun (1 AU), hence 2 AU. A foreground star will appear to have moved with respect to the background stars, which are so far away that the change in perspective did not change their positions.

The closer a foreground star is to us, the further it will seem to have moved between the two perspective views. The angle the star moved is called the parallax, π, and is measured in arc seconds ($''$). It is related to the star's distance, d, in parsec (pc):

$$d = 1/\pi \tag{7.8}$$

A star at 1pc distance, will have a parallax $\pi = 1''$.

Parsec relates to other distance units as: 1pc= 206,265 AU and 1pc=3.26 light years (ly), the distance light can travel in one year, i.e. 1 ly = 9.5×10^{15}.

7.4.2 DISTANCE MODULUS

The 'distance modulus' is a way to find distances to stars, where the absolute magnitudes are known. The distance modulus is:

$$\mathrm{m} - \mathrm{M} = -1.5 \times \log(\frac{\mathrm{d}}{10\mathrm{pc}}), \tag{7.9}$$

where d is the distance in parsec.

If the absolute magnitude of a star is known, and we measure its apparent magnitude, we can deduce its distance by rearranging the distance modulus:

$$\mathrm{d} = 10\mathrm{pc} \times 10^{(\mathrm{m}-\mathrm{M})/5} \tag{7.10}$$

While we can always just measure the apparent magnitude, the tricky part is to know the absolute magnitude. The absolute magnitude of stars can be found by using the HR-diagram (see Chapter 8.1). By inspecting a star's spectrum it is possible to infer where in the HR-diagram it belongs, which will give an absolute magnitude.

7.4.3 MAIN SEQUENCE FITTING

Some stars are located in clusters, often we distinguish between globular clusters and open clusters. A cluster is an association of stars, which share age, movement, and distance. It is possible to find the distance to a cluster by comparing it with a cluster where the distance is known. The known cluster should have calibrated measurements plotted in the HR-diagram (see Chapter 8.1), where the distance and thus the absolute magnitudes of the member stars are known. The unknown cluster is also plotted in a HR-diagram, just with apparent magnitudes instead of absolute magnitudes. The two HR-diagrams are then compared, or rather the two 'main sequences' are compared. The unknown cluster is then moved up or down along the brightness axis so that it covers the known cluster. This fitting is called 'main-sequence fitting'. Often smaller adjustments have to be made to account for metallicity and age. From the known cluster we can then attain the absolute magnitude for the unknown cluster, and thereby its distance using the distance module.

7.4.4 CEPHEID VARIABLES

Cepheids are variable stars that change in magnitude with regular periods. Their periods are 1-50 days and the change in magnitude between 0.1-2.5. In 1912 Henrietta Swan Leavitt published a discovery of a relation between the period and Luminosity. That made it possible to find the absolute magnitudes for these stars, and thereby also determine their distances. This method can be used to calculate the distances to nearby galaxies.

7.4.5 TYPE IA SUPERNOVAE

Supernova type Ia can be used to determine distances further than any other method. A type Ia supernova is a white dwarf that undergoes a thermonuclear explosion that completely destroys the star. These bright supernovae have a clear relation between the maximum brightness and the radioactive decay evident in the light curve. This means that the absolute magnitude can be found based on the light curve alone. The apparent magnitude is, as always, directly measured. With the apparent and absolute magnitudes known, the distance can be derived using the distance modulus.

The extreme luminosity of supernovae means that they can be detected at large distances. They are often nicknamed 'cosmological' supernovae because they can help probe the overall geometry of the Universe and its rate of expansion. In 2011 the Nobel prize in physics was given to two competing collaborations (the High-Z Supernovae Search team & the Supernova Cos- mology Project) doing just that. They found independently of each other that the Universe is accelerating in its expansion.

7.4.6 COSMOLOGICAL REDSHIFT

Beyond a certain distance we no longer use the normal units of distance. Instead the term 'redshift' is used. Cosmological redshift is very similar to the doppler redshift, in that we observe it as a shift in the emitted light, see Chapter 5.5.1. It appears as if all objects far away are moving away from us. Furthermore, as Hubble observationally confirmed in 1929, the further away an object is, the faster it is moving away from us. This is called Hubble's law, and it relates the distance, d, to the radial velocity, v_r:

$$v_r = H_0 \times d, \tag{7.11}$$

where the constant H_0 is called Hubble's constant, which is a constant in space, but not in time.

The radial velocity, v_r, is calculated from the redshift, z, of the lines by:

$$v_r = c \times z, \tag{7.12}$$

where c is the speed of light and the redshift, z, is:

$$z = \frac{\Delta\lambda}{\lambda_{rest}} = \frac{\lambda_{obs} - \lambda_{rest}}{\lambda_{rest}}, \tag{7.13}$$

where λ_{obs} is the observed and λ_{rest} the rest wavelength.

When astronomers navigate the Milky Way, parsec is the unit of choice. The Sun is 8.5 kpc from the centre of the Milky Way. The distance to our neighbouring galaxies are measured partially in parsec, partially in light years. In the local bubble of galaxies we use kpc or Mpc. Beyond that the distance is measured in redshift. Because cosmological distances rely on the observationally determined Hubble constant, it is 'cleaner' to express distances as redshifts. Instead of using Mpc (10^6 pc) we use the unit free redshift z. The furthest galaxy observed (as of June 2016) is at a redshift of z=11.1. If we use H_0 = 72 km/s/Mpc, then we can calculate the distance:

$$d = \frac{v_r}{H_0} = \frac{c \times z}{H_0} = \frac{3 \times 10^5 km/s \times 11.1}{72 km/s/Mpc} = 46250 Mpc \tag{7.14}$$

or 46.25 10^9 pc. It helps to keep it in redshift terms, as the distance then is free of choice of Hubble constant.

7.5 PRACTICAL ASTRONOMY

Because space is not empty, the light emitted from stars might be absorbed by interstellar material. Actually the further away an object is, the more light is likely to be absorbed. The above definitions for magnitudes are valid for the ideal case of no absorption by dust or the Earth's atmosphere. The formulae can then be corrected to include a term, A, that accounts for the absorption, i.e.:

$$m - M = 5 \log \frac{d}{10pc} + A. \tag{7.15}$$

Some absorption might not only make an object dimmer. Light that is absorbed can be re-emitted at lower wavelengths, which would cause a reddening of the light.
Optical astronomy makes use of several filters. A filter is a partially see through thin sheet that is mounted on a telescope. In optical astronomy the most common filters are U (Ultraviolet), B (Blue), V (Visual), R (Red), and I (Infrared). They each let only a part of the electromagnetic spectrum through.

Stars are observed through two or more filters, which makes it possible to find their colours and infer their black body curves. This way the time consuming spectral observations can be avoided. This method is especially useful when the observing target is a group of stars. The apparent magni- tude of a star observed through the B-filter will simply be called: B. The same goes for apparent magnitudes in the other filters. An absolute magnitude in the B-filter is denoted MB (in the V-filter, MV and so on). The colour of a star is given by either U-B or B-V. The colour of a star is a direct indication of its surface temperature.

The colour of a star found this way is not influenced by a dimming due to absorption, if the absorption is flat, i.e. the same at all wavelengths. That is rarely the case, in fact there is a general reddening of the light, as blue light scatters more than red light.
To correct for this the colour excess, E, is introduced:

$$E_{B-V} = (B - V) - (B - V)_0, \tag{7.16}$$

where the subscript 0 indicates B and V colours intrinsic to the star, i.e. without any absorption or reddening.
Studies of the interstellar medium in the Milky Way has found that the ratio between A_V and E_{B-V} is almost constant:

$$\frac{A_V}{E_{B-V}} \approx 3.1 \tag{7.17}$$

We now have the tools in place to sort stars into different groups based on their intrinsic colours.

7.5.1 STELLAR CLASSIFICATION

By observing stars through different filters, their colours can be determined fairly accurately. Stars are grouped into types based on their colours.

Historically there were several classification systems, but now the Harvard spectral classification system is the one, that is being used. The Harvard classification system began in 1872 by Henry Draper, whose work and equipment was donated to Harvard Observatory after his death. In 1918-1924 the HD catalogue was published, of which the main part of the classification was done by Annie Jump Cannon. It contains more than 225,000 stars, where each is identified by a number with the HD prefix. Initially the classification system was based on the strength of the hydrogen absorption lines. Later it was discovered that the line strength depends on the surface temperature.

The stars are divided into classes (also called types) based on their surface temperature. Practically, that means that by finding the intrinsic colour, the surface temperature can be inferred. If the spectrum for a star is available the temperature can also be found by fitting a black body curve to the spectrum.

Here is a short overview of the stellar types:

- O-stars: These blue stars are the hottest stars with surface tempera- tures of 20,000 - 35,000° K.
- B-stars: Blue or white stars that have surface temperatures around 15,000° K.
- A-stars: Almost white in colour and have surface temperatures of 9,000° K.
- F-stars: White or Yellow in colour and have surface temperatures of 7,000° K.
- G-stars: The Sun is a G-type star, i.e. yellow and with a surface temperature of 5,500° K.
- K-stars: Orange or yellow and with surface temperatures of 4,000° K.
- M-stars: The coldest stars are called M stars, they are red and therefore with surface temperatures of 3,000° K.

When ordered by temperature the spectral type sequence is: OBAFGKM. These types are divided further into sub-divisions from 0 to 9 in order to distinguish between slight differences in the stellar spectral patterns, which depend on the stars' temperatures. 0 stars are hotter than 1 stars.

The Harvard system only takes the surface temperature into account and not the luminosity. To distinguish two stars with same colour, but with different luminosities a complementary system was developed. The MKK system (of luminosity classes) was developed at the Yerkes Observatory, and is named after the initials of the creators William W. Morgan, Phillip C. Keenan, and Edith Kellman. The system assigns a Roman numeral in addi- tion to the Harvard type. Luminosity class I is brighter than luminosity class I I. The system is based on an absorption line analysis, where the depths of the lines are directly related to the surface gravity and hence the luminosity.

Combining the two classification systems our Sun has the type G2V, indicating a main-sequence star with a temperature around 5,800K.

The Harvard classification system was settled before a coherent picture of stellar evolution was established, i.e. before it was realised that the HR- diagram is a roadmap of evolutionary paths (see Chapter 8.1). The implication is, that we now know, that certain stellar types are earlier stages for other types. For an observer a star has a type and stays that way. However, as rare as it is to see a change, the stars evolve following paths around in the HR-diagram, and one stellar type is a progenitor stage for another. Our Sun (type G2V) will end up as either a K9III or M0III giant.

Bibliography and Links

Any good astronomy textbook mentions magnitudes and luminosity. The best astronomy textbook is:
Karttunen et al., 2003, 'Fundamental Astronomy', Springer, 4th edition. ISBN:3-540-00179-4
A reference to the discovery of the furthest observed galaxy:
Oesch et al., 2016, 'A Remarkably Luminous Galaxy at z=11.1 Measured with Hubble Space Telescope Grism Spectroscopy', ApJ, Vol.819, pp.129.

8 STELLAR EVOLUTION

Your goals for this chapter are to learn about:

- The Hertzsprung-Russell Diagram.
- The Hayashi Track for newly forming stars.
- Stars on the Main Sequence.
- Stellar evolution for low mass stars.
- Stellar evolution for high mass stars.

In Astronomy the term 'evolution' refers to the different stages a star passes through in its lifetime. This should not be confused with the biological term, which means changes to a population, where hereditary traits are passed on from one generation to the next.

Stellar evolution begins at the onset of nuclear burning in the star's interior. During the hydrogen burning phase the stars are on the main sequence, an evolutionary step that all stars goes through. When the hydrogen has been depleted there are several evolutionary paths that a star can take. The deciding factor is the mass of the star.

8.0.2 STELLAR POPULATIONS

We observe stars as a population, in which there are stars at all stages in their evolution. Stellar evolution works on very large timescales and it is rare to see a particular star change from one evolutionary stage to the next.

When we look at all the stars as a population, we see both young and old versions of the same type of star, i.e. of stars with similar starting conditions. That means, that when we know the type of a star and its evolutionary stage, we can reliable predict its complete evolutionary path.
It is possible to use a stellar population study to determine stellar evolution, because it is mainly the mass of a star that determines its type and subsequently its evolutionary path.

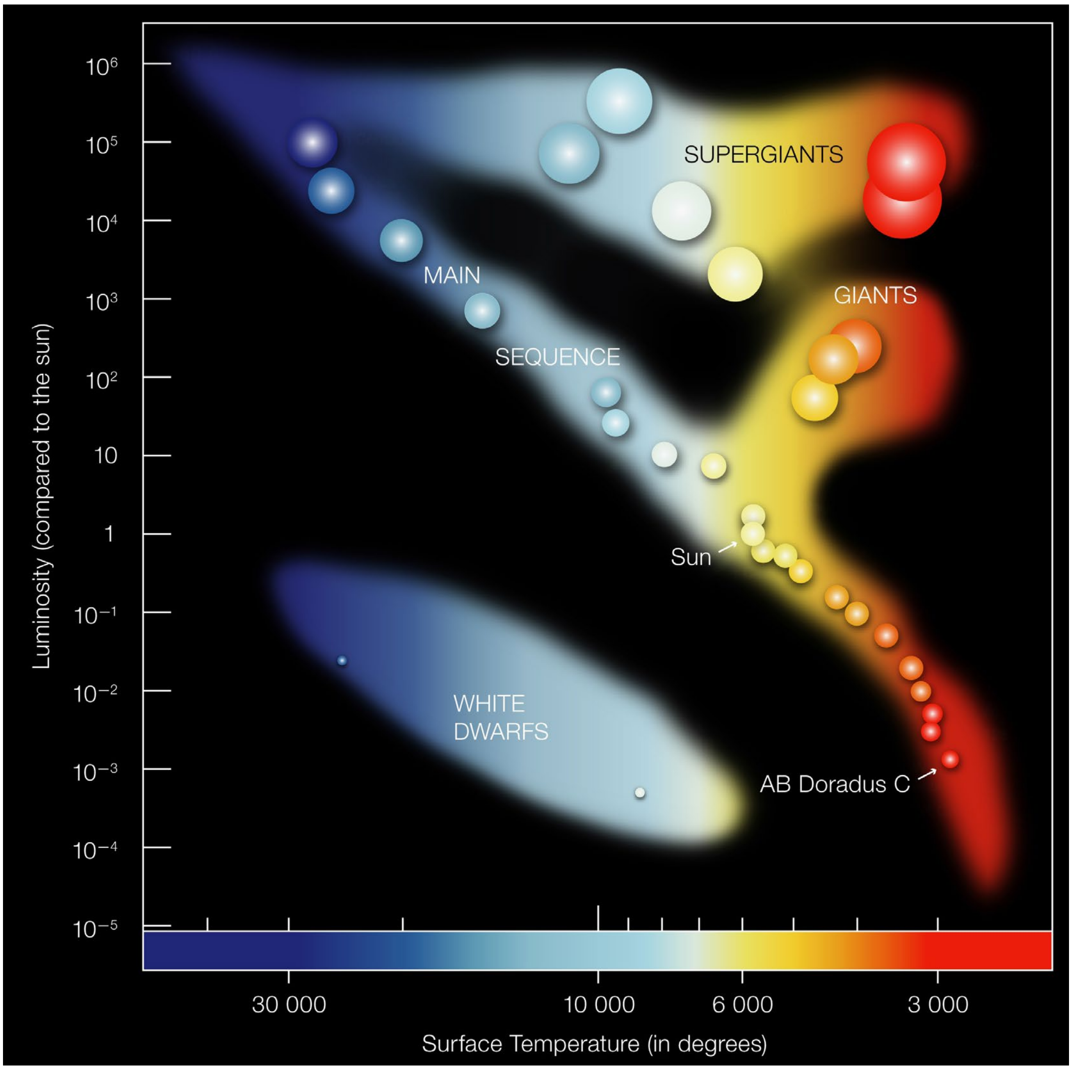


Figure 8.1: The Hertzsprung-Russell Diagram, where regions of specific evo- lutionary stages are drawn. Image courtesy of ESO -European Southern Observatory.

8.1 THE HERTZSPRUNG-RUSSELL DIAGRAM

The most important tool, by far, in understanding stellar evolution is the Hertzsprung-Russell diagram (from here on referred to as the HR-diagram). It is basically a graph, where along the x-axis the surface temperature (T_{eff}) is plotted and along the y-axis the luminosity (L), which is a distance independent measure for brightness, see Chapter 7.3.
When stars are plotted in this diagram, where each star is just represented by a small dot, a coherent picture emerges, which enables us to group stars into different categories. Although stars come in many different sizes (i.e. with many different masses), they share some common traits when they are in the same evolutionary stages.

The HR-diagram is often also called the colour-magnitude diagram, because T_{eff} is what determines the colour of a star, and the magnitude is closely related to the luminosity. As a star evolves it will move around in the HR- diagram, because its outer physical properties changes as its interior fusion go through different stages. In figure 8.1 a HR-diagram is depicted, where the relative sizes of the stars are illustrated as well as their colours.

In the bottom left corner of the HR-diagram we find the White Dwarfs, which are small but hot. Their surface temperatures bring them all the way to the left in the diagram (blue is hotter than red), and because they are relatively small their combined luminosity is rather faint. In the top right corner we find the red giants and supergiants, which are bright but have a relatively low surface temperature. The diagonal line stretching from the top left corner to the bottom right corner is the Main Sequence (MS in shorthand), stars spend most of their lives in this region of the HR-diagram. The brightest stars, located in the top of the diagram, are also the rarest stars as they exhaust their fuel faster than the fainter stars located at the bottom of the diagram. By finding a star's position in the HR-diagram we can determine its evolutionary stage and its mass and size.

When astronomers look at an isolated group of stars, called a globular cluster, where all the stars were formed from the same cloud, the HR-diagram can be used to determine the absolute age of the cluster. This is a consequence of the fact that the larger stars burn up faster. The main sequence will not stretch the full diagonal as the larger stars have turned into giants and supergiants. Instead the main sequence will in the top veer off to the right, where the stars now are populating the giant region. The brightest star still to be on the main sequence is defined as the turn-off point. From the turn-off point the age of the cluster can be determined. The method is akin to looking at a candle in order to find out how long time it has been burning, which is easy if you know the initial length of the candle and the burning speed.

8.2 PROTOSTARS

Protostars are pre main sequence stars, i.e. stars that are just being formed. A dense molecular cloud, or a region of it can contract due to its own gravity. If the mass of the cloud is large enough, more specifically the potential energy must be twice the kinetic energy, the cloud has reached a critical mass necessary for the collapse. This criterion was first discovered by Sir Jean in 1902 and is called the Jean's limit.

Protostars starts to shine as they contract, because the release of potential energy is converted into kinetic energy and radiation. However, they are difficult to observe, because they are often situated deep inside large dense clouds.

Protostars are located to the right in the HR-diagram, where the less massive protostars are less bright (towards the bottom) and the very massive protostars are more luminous (towards the top). This is because the more massive protostars have a larger energy release because of their larger masses. Initially the radiation can escape the protostar freely, because the density is low. As the protostar contracts and gets denser this changes, and a larger fraction of the released energy is converted into heat instead of being radiated away. As the temperature rises, so does the pressure, which works against the gravitational pull inwards. Consequently the collapse slows down in the denser centre, but keeps being in free fall further out.

The collapse stops when the protostar has reached equilibrium, where the gravitational force inward is balanced by the pressure outward. Protostars in equilibrium are found on the Hayashi Track, which is a luminosity-temperature relation that protostars follow. This relation can conveniently be drawn in the HR-diagram, where it is an almost vertical line to the far right. All protostars, regardless of mass, will at some point be on this line. During the initial collapse the protostars were to the right of this line, a region also called the 'forbidden zone' because no stable star with the physical properties corresponding to that region can exist. Protostars on the Hayashi track are fully convective, which means they are efficiently mixing the material throughout the whole star.

The collapse of a solar mass protostar onto the Hayashi track takes around 100,000 years. The protostars are still contracting and thereby reducing their radius and luminosity, which means they move downwards on the Hayashi track. Inside the protostars the temperature rises and a radiative zone around the cores forms causing the star to increase in surface temperature without much change to the luminosity.

As soon as the radiative zone is in effect the protostar leaves the Hayashi track in the HR-diagram. The protostar moves to the left towards the main sequence along an almost horizontal line (constant luminosity) while its surface temperature increases. Hydrogen burning starts while the protostar is still on the Hayashi track, albeit in reactions that demand a lower temperature than the pp-chain does. When the protostar reaches a core temperature of 4 million degrees the pp-fusion-chain switches on and the star is no longer a protostar but a full fledged main sequence star.

Protostars with very low masses never leave the Hayashi track, as their interior never gets hot enough to become radiative. This is because there was not enough mass to provide the necessary energy release and density in the centre. The main sequence intersects with the Hayashi track at a very low luminosity and here is a place for the stars that stay fully convective and also generate energy by hydrogen fusion. The stars that are not massive enough to start fusion will cool and fade. Those stars we call brown dwarfs and they have masses below $0.08M_\odot$, where $M_\odot$ is the mass of the Sun. The upper size of a star is limited by the radiation pressure in the initial formation, which can be so large that the star cannot form.

8.3 MAIN SEQUENCE STARS

The main sequence is what we call the diagonal line in the HR-diagram, which runs from the top left corner to the bottom right corner. Stars in this region of the HR-diagram are burning hydrogen into helium, and they spend the majority of their lifetimes on the main sequence.

Stars on the MS are in a exceptionally stable part of their evolution. The stars do not change in brightness or temperature, only when their fuel is depleted do they change into the next evolutionary stage. This is because the stars are in hydrostatic equilibrium, a condition where there is a balance between two forces at every point in the star.

The force pulling all material towards the centre is gravity, which is balanced by the outward pressure, which essentially is the heat of the star fuelled by fusion. The balance is such, that if we imagine the star to change a little bit, a negative feed-back-loop would bring the star back to its initial conditions. Say a star expanded a bit, that would cause it to cool

off, as the increased surface area lead to a larger luminosity. Cooling off implies that the temperature decreases and the nuclear reactions would decline and with that the pressure. Since it is the pressure that balances out gravity, the star would shrink a bit, restoring the star to its original size.
Similarly, if we pretend to squeeze a star, it would get denser and hotter increasing the nuclear reactions. This would in turn increase the outward pressure, which would expand the star and restore it to its previous equilibrium.

Main sequence stars can be divided into 3 separate groups based on their interior structure.

- The smaller stars with masses, M, between:
  $0.08M_\odot \le M \le 0.26M_\odot$,
- The intermediate size stars with masses, M, between:
  $0.26M_\odot \le M \le 1.5M_\odot$,
- The larger stars, with masses, M, above $1.5M_\odot$

The smallest MS stars are very faint and they are located in the bottom right corner of the HR-diagram. These stars never leave the Hayashi track, as they stay fully convective throughout their entire hydrogen burning lives. The convection causes the stars to be well mixed, which implies that all the hydrogen present is available for the nuclear fusion. When all the hydrogen has been converted into helium these stars will fade and contract into helium white dwarfs.

The intermediate sized stars are brighter and hotter than the smaller stars. The Sun is a typical MS star in its size group. The interior energy transportation is radiative (via photons), and it is convective in the envelope. This division means than this type of star is not mixed throughout. The hydrogen fusion changes the chemical composition in the core, where there over time will be more and more helium in the core, whereas the surface will not change much from its original composition. Intermediate sized stars burn hydrogen into helium via the pp-chain (described in chapter 6.1).

Stars with masses above $1.5M_\odot$ are even brighter and hotter. They have a convective core and a radiative envelope, exactly opposite to the structure of the intermediate sized stars. These hot stars are warm enough in their centres to burn hydrogen into helium via the CNO cycle. The CNO cycle is named after the three catalyst elements in the reaction, namely Carbon, Nitrogen and Oxygen. The nuclei synthesis in the CNO cycle go through the following reactions:

$$
\begin{aligned}
{}^{12}C + {}^{1}H &\longrightarrow {}^{13}N + \gamma \\
{}^{13}N &\longrightarrow {}^{13}C + e^{+} + \nu_e + \gamma \\
{}^{13}C + {}^{1}H &\longrightarrow {}^{14}N + \gamma \\
{}^{14}N + {}^{1}H &\longrightarrow {}^{15}O + \gamma \\
{}^{15}O &\longrightarrow {}^{15}N + e^{+} + \nu_e + \gamma \\
{}^{15}N + {}^{1}H &\longrightarrow {}^{12}C + {}^{4}He + \gamma,
\end{aligned}
$$

where γ is a photon, e^+ a positron, and ν_e an electron neutrino.

The CNO cycle is very temperature sensitive and will only take place in the core, where the temperature has to be more than 20 million degrees. The hotter a star is in its centre the more effective the CNO cycle becomes. Really large stars with very hot centres burn off their hydrogen at a faster rate than smaller stars.

A main sequence star is stable as long as its internal energy production stays the same. Because stars form with a huge supply of hydrogen, stars remain stable on the main sequence for a long time. While on the MS the core of a star slowly gets denser because helium has a larger mean molecular weight than hydrogen. As the core gets denser the star gets slightly brighter, which moves it upwards along the MS. The time that a star spends on the main sequence is determined by its mass. Smaller, i.e. cooler, stars have slower fusion reactions rates and therefore will remain on the main sequence longer.

The main sequence is not a thin line, it has a certain width. This is because stars with different chemical compositions and otherwise same masses and ages, will have slightly different brightnesses and surface temperatures. In all, when looking at stars, we are more likely to find them in their MS evolutionary stage than at any other point in the HR-diagram, because they simply spend so much more time in this stage than in any other stage. A star like the Sun (an intermediate sized star) will burn hydrogen in its centre on the MS for around 10,000 million years (10×10^9). The Sun has been a MS star for about 4.5 billion years (4.5×10^9), and is thus almost halfway through its MS phase.

8.4 THE RED GIANT PHASE

The red giant phase is, essentially, where the star is not producing energy in the core but in a shell around the core. This is an evolutionary path that only stars with masses below $5M_\odot$ take, as heavier stars will have denser and hotter centres making fusion there possible throughout their lives. In the Red Giant phase the star expands, which leads us to name it a giant. In its expansion the outer layer cools and the surface appears redder. In the following the path from the main sequence to the Red Giant Branch in the HR-diagram is explained, see Figure 8.2 for an illustration.

Stars with masses below $0.26M_\odot$ are convective throughout their entire MS phase, making all the hydrogen in the star available for fusion. When the hydrogen is exhausted these stars will not have a hydrogen shell burning phase, as there is no hydrogen left in the star at all. These stars, too light to initiate helium burning, will fade and cool off, ending their lives as white dwarfs consisting of predominantly of helium.

When there is no more hydrogen in the core the star will move to the right in the HR-diagram. This stage marks the end of its MS life. When the hydrogen burning ceases in the core the star is no longer in hydrodynamical equilibrium, as the outward pressure was driven by the energy release in the nucleosynthesis. This causes gravity to win and the star collapses, thereby converting potential energy into kinetic energy, which then heats the star up. In the collapse the star also becomes denser, both in the core and outside the core. The collapse continues until the pressure and density in the core are large enough to halt the collapse. This pushes the temperature outside the core up to the hydrogen burning ignition point, and fusion then takes place in a thick shell around the helium core.

At this stage the star has an inner (dormant) helium core surrounded by a hydrogen burning shell and finally an envelope, which is also the outer layer of the star. The energy production in the shell supplies enough energy that the star is no longer in a state of collapse. The luminosity generated in the thick shell exceeds what was produced by the core on the MS, and as a result the star gets brighter. Not all of the energy is radiated away, a fraction of it goes into slowly expanding the envelope. This causes the surface temperature to fall and the star becomes redder. This implies that the star moves almost horizontally to the right in the HR-diagram. The star's evolutionary path to the right after the MS phase is slow and gradual for intermediate mass stars and fast for higher mass stars. The position of the star in the HR-diagram is now on the sub-giant branch (SGB).

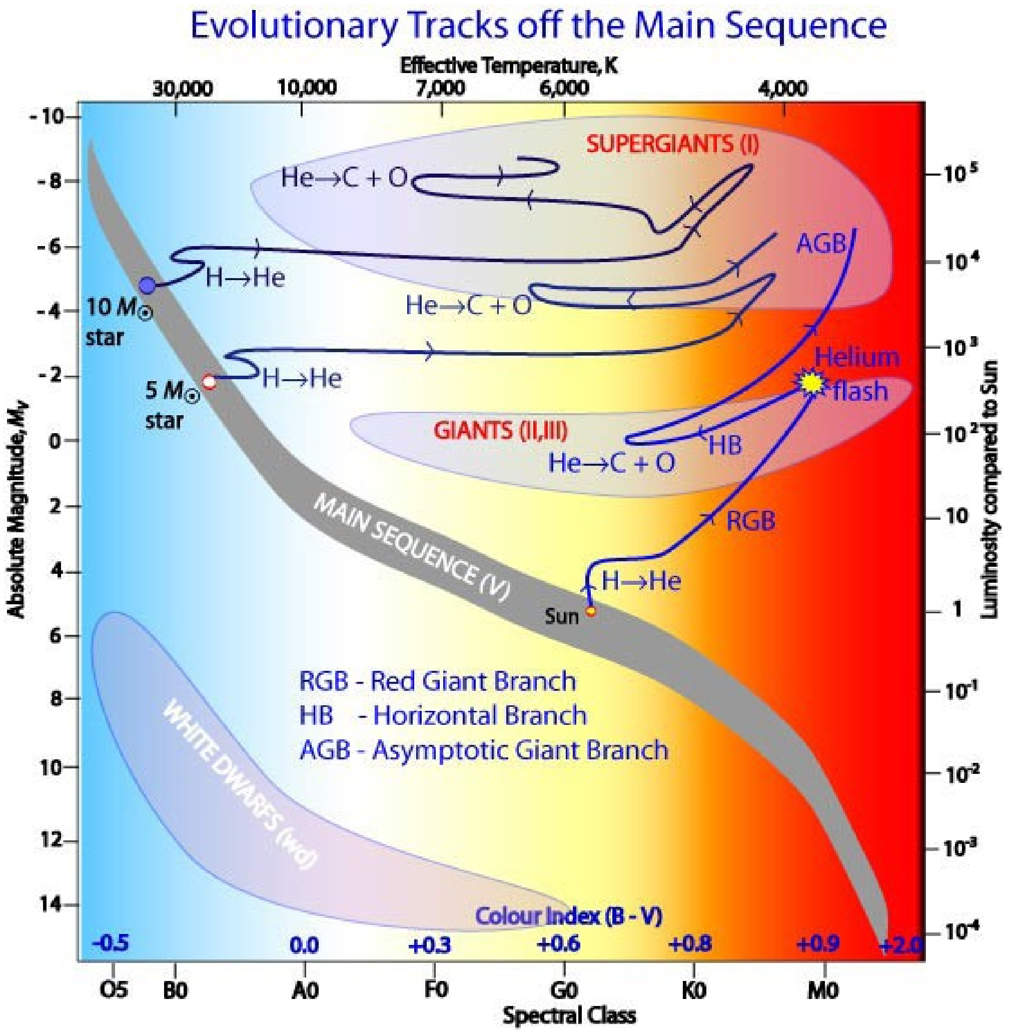


Figure 8.2: The Hertzsprung-Russell Diagram where three evolutionary tracks post MS are drawn. Image courtesy of CSIRO Australia.

During hydrogen shell burning the mass of the helium core will increase as the hydrogen burning shell contributes newly synthesised helium to it.

The sub-giant phase of evolution ends when the mass of the core has become too large for it to support the layers above it. The core then becomes degenerate, as the pressure from degenerate electrons supports its outward pressure. Read more extensively about degeneracy in Chapter 9.2. This outward pressure keeps the star stable, while it continues to burn hydrogen in a shell around the core. The shell has, during the collapse, also become denser and hotter, and although the shell is now thinner, it generates energy more efficiently. This forces the envelope of the star to expand as it absorbs most of the energy released by the shell. As a result the surface temperature and luminosity of the star decreases.

The core continues to contract and the energy production of the narrowing hydrogen burning shell increases still further. Since the envelope is saturated, it can absorb no more energy, the energy is radiated away, which increases the luminosity and the radius expands further. The surface of the star cools and the star reaches the Hayashi track. Between the MS and reaching the Hayashi track the star's exact movement in the HR-diagram depends on its mass. After reaching the Hayashi track the star will move almost vertically upwards along the Hayashi track as its brightness increases. The star will then have reached the evolutionary stage that we call Red Giant, and the star will be located on the Red Giant Branch (RGB).

Stars with masses below $0.5M_{\odot}$ are not heavy enough, and thereby not hot enough, for helium fusion to start. For these stars the RGB phase will be their last. Their Red Giant phase is relatively short and when the hydrogen in their hydrogen burning shells is exhausted they will fade and cool. The stellar remnant is thus a helium white dwarf. The evolutionary track in the HR-diagram cross the main sequence and ends in the lower left hand corner, where the white dwarfs reside.

8.5 HORIZONTAL BRANCH

Stars positioned on the horizontal branch (HB) in the HR-diagram are fus- ing helium in their cores. All stars with masses above $0.5M_\odot$ will at some point be on the HB. Lower mass stars (below $5M_\odot$), who had gone through a red giant phase will move on to the HB from the right, whereas larger stars (above $5M_\odot$) move directly onto the HB from the MS, i.e. from the left. Larger stars are generally brighter and they tend to be on a HB that is located further up (brighter) in the HR-diagram.

Stars, with $M \leq 5M_\odot$, move to the left along the HB (after the Red Giant phase), when they start fusing helium into carbon. For stars with masses above 0.5M the core will eventually become dense and hot enough (1.3×10^8K) for the triple alpha process to begin. This reaction has not taken place at lower densities and temperatures, because the helium atom nuclei (essentially α-particles) normally repel each other. At high enough temperature and density this hurdle can be overcome.

The triple alpha reaction consists of two processes:

$$^4\text{He} + {}^4\text{He} \longleftrightarrow {}^8\text{Be} \tag{8.1}$$

$$^4\text{He} + {}^8\text{Be} \longrightarrow {}^{12}\text{C} + \gamma \tag{8.2}$$

The first of which goes both ways, as Be is highly unstable and will decay into two helium nuclei really fast. The second reaction thus happens when the helium nuclei can interact with Be the instant it was formed. Because the two reactions have to happen so fast after each other, they are often bundled together and just referred to as the triple alpha reaction. During helium fusion some of the newly synthesised carbon nuclei will react with helium and form oxygen:

$$^{12}\text{C} + {}^4\text{He} \longrightarrow {}^{16}\text{O} + \gamma. \tag{8.3}$$

For stars with masses below $2M_\odot$ the onset of the triple alpha reaction in the core is rather violent. These stars have degenerate cores, which means the core temperature can rise to the trigger temperature throughout. This feature is a result of electron degeneracy having a very weak dependence on temperature. We will go into detail about degeneracy in Chapter 9.2. When the triple alpha reaction sets in, it does so in the entire core simultaneously. The energy released is tremendous, around 10^{11} times the luminosity of the Sun ($10^{11}L_\odot$), but also very brief -just a few seconds. Most of the energy never reaches the surface of the star, as the upper layers absorb it, often also causing mass loss from the star. This brief core helium burning is called 'helium flash'. The helium flash lifts the degeneracy of the core, and it can again behave as an ideal gas. i.e. expand and cool. The expanded core is still hot enough to burn helium and it will continue to do so.

Stars with masses above $2M_\odot$ have a less violent onset of helium fusion. The cores of massive stars reach 10^8 K before they are dense enough to become degenerate, which means that helium fusion begins more smoothly. When the star is fusing helium in its core it contracts and it is no longer a red giant.

The onset of core helium fusion in a low mass star marks the evolutionary phase where the star moves to the left along the horizontal branch (HB) in the HR-diagram. The energy released leads to a higher surface temperature, but because the star has contracted the surface area is smaller and the luminosity remains roughly the same. That translates into an evolutionary track in the HR-diagram that is horizontal (constant brightness) and to the left (hotter surface means bluer colour). It might take the star several contractions/ expansions to settle on the HB. Stars like our Sun will have enough helium to stay on the HB for 100 million years.

8.6 THE ASYMPTOTIC GIANT BRANCH

When the all the helium in the core has been converted into carbon and oxygen the fusion ceases. The outward pressure, which was fuelled by fusion, is no longer supported and the star contracts. The contraction of the star causes the temperature to increase in the whole star as potential energy is converted into kinetic energy.

Eventually, the layer just above the core is hot enough for helium to ignite. The Star now has a helium burning shell just above its quiet core. Helium fuses into carbon via the triple reaction listed above. Energy from the helium burning heats up the above layer, which then ignites hydrogen burning in a shell there. A helium burning shell surrounded by a hydrogen burning shell is the hallmark of a star on the asymptotic giant branch (AGB).

Like on the red giant branch, the stars on the AGB have no energy production in their centres. Instead, the energy supporting the star from gravitational collapse comes from helium and hydrogen burning in two separate shells. The lower mass limit for an AGB star is $0.5M_\odot$, which is the limiting mass needed to get a temperature hot enough to ignite helium. The upper mass limit for an AGB star is $7M_\odot$, above which the core becomes hot enough for nuclear fusion of carbon and oxygen to ignite. AGB stars thus have insufficient gravitational energy to generate the temperatures needed to have fusion in their cores ever again.

It turns out that the two fusion burning shells close to each other is unstable as the layers might mix, which leads to thermal pulses rather than a steady shell burning. Like stars on the red giant branch, stars on the AGB experience mass loss. The outer layer is already puffed up in size, which translates to a relative low density, and the radiation pressure is then able to blow part of the outermost layer away. Mass loss rates are given in solar masses per year, and for AGB stars it is about $10^{-6}M_\odot/yr$. However, the thermal pulsations can increase this rate up to about $10^{-4}M_\odot/yr$ as the star evolves up the AGB.

The ignition temperature for helium is higher than for hydrogen, which implies that the inner layer is hotter than the outer layer. The hydrogen burning shell is therefore hotter near the helium burning shell, where the hydrogen fusion rate is correspondingly greater. The hydrogen burning shell in general is hotter than it was in the RGB phase, and therefore also burning hydrogen at a higher rate with a larger energy output. An AGB star is at least 10 times brighter than a RGB star.

The two burning shells and the quiet core translates in the HR-diagram into a movement from the HB to the right (redder) and up (increase in brightness). The asymptotic giant branch lies just to the right and above the red giant branch. The AGB phase marks the largest and brightest the stars in this mass range will ever be. Where, for comparison, a solar mass star in the AGB phase will have a luminosity about 10,000 times that it had on the main sequence. A brightness this great is normally reserved for higher mass stars ($M > 7M_\odot$), and it is only in these late stages that lower mass stars ($M < 7M_\odot$) come within reach of that.

At the end of the AGB phase the star is not heavy enough to get to the temperatures necessary for further nucleosynthesis. Instead the star will expel the outer layers in the final thermal pulses. These layers are later observed as spectacular nebulas, see figure 8.3. Some nebulae show intricate geometrical structures, which are caused by the influence of a companion star. Spherical nebulas (like the image) are from single stars. In the centre the hot core contracts and as it cools it becomes a carbon-oxygen white dwarf. The white dwarf ends up in the lower left corner of the HR-diagram with the other white dwarfs. This is also the ultimate fate of our Sun.

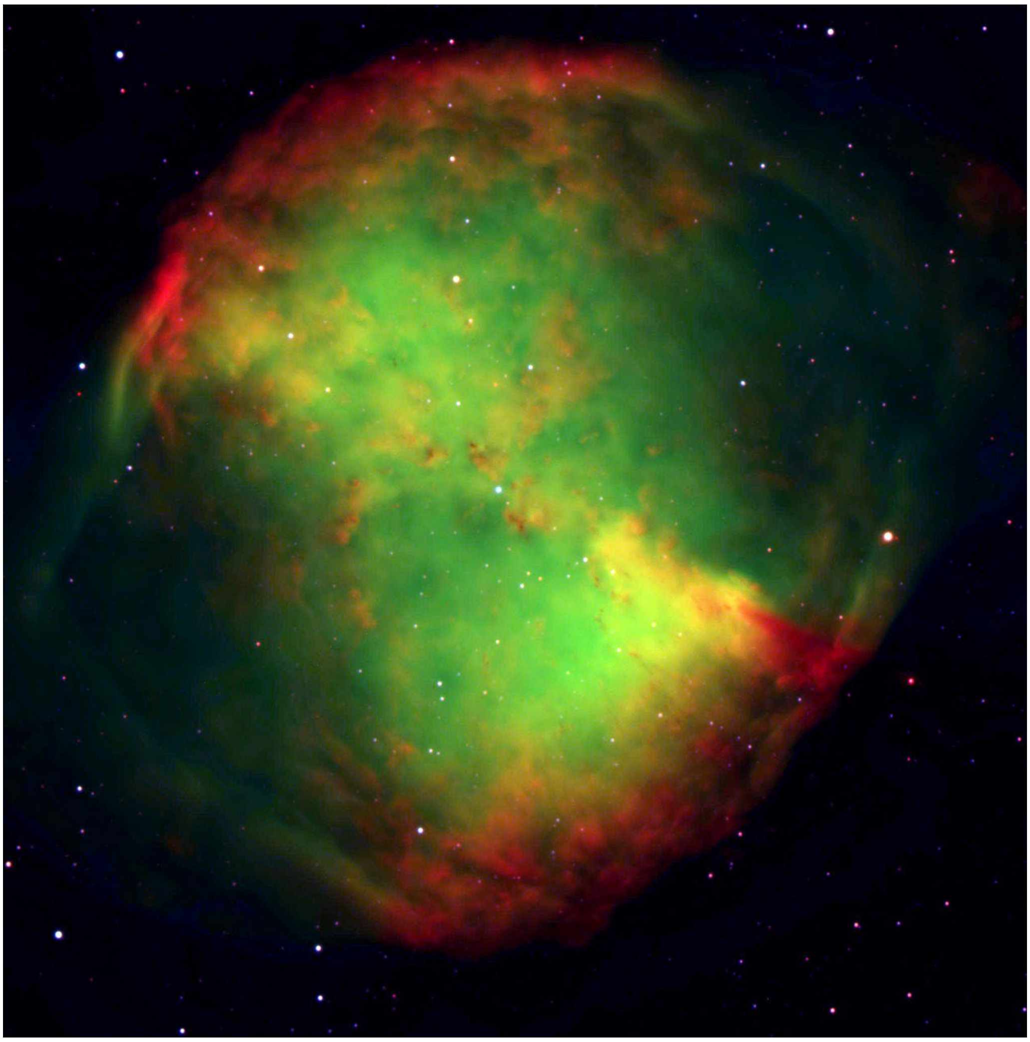


Figure 8.3: The Dumbbell Nebula is a typical planetary nebular (nothing to do with planets), i.e. the remnant of an AGB star. The white dot in the centre is the remaining white dwarf. Image courtesy of ESO.

8.7 SUPER AGB STARS

Stars with masses above $7M_\odot$ and below $9.25M_\odot$ (maybe all the way up to $12M_\odot$, all depending on metallicity and mass loss history) will go through an AGB phase, where the core is dormant and the main energy production takes place in shells. These stars are named Super AGB stars (SAGB) because of their larger masses, different cores, and potentially different death.

Super AGB stars have cores that consist of oxygen and neon (ONe) rather than carbon and oxygen. These stars have been fusing carbon in the core:

$$
\begin{aligned}
{}^{12}\mathrm{C} + {}^{12}\mathrm{C} &\longrightarrow {}^{24}\mathrm{Mg} + \gamma && (8.4)\\
&\longrightarrow {}^{20}\mathrm{Na} + {}^{1}\mathrm{H} && (8.5)\\
&\longrightarrow {}^{23}\mathrm{Ne} + {}^{4}\mathrm{He} && (8.6)\\
&\longrightarrow {}^{23}\mathrm{Mg} + \mathrm{n} && (8.7)\\
&\longrightarrow {}^{16}\mathrm{O} + 2{}^{4}\mathrm{He} && (8.8)
\end{aligned}
$$

After the star has exhausted all its carbon the core stops nuclear fusion, but the layers above can continue shell burning. On the outside these stars look like bright AGB stars, but in their centres the have a different core.

The SAGB phase means that the star increases in size and has a cooler surface (and appears redder). Two separate layers are producing energy via nucleosynthesis. The innermost layer is fusing helium to carbon and the outermost hydrogen to helium. Like their lighter cousins, the super AGB stars will have thermal pulses and significant mass loss.

If such a star loses its envelope during thermal pulsing, the core will remain as a white dwarf. The core will cool and contract and end up with all the other white dwarfs in the lower left corner in the HR-diagram. For the heavier SAGB stars the core can increase in mass during the shell burning phase. This will eventually push the mass of the core to the Chandrasekhar mass, at which point the core will collapse. In this collapse energy is released and the star explodes as a supernova. These Super AGB stars might be the lowest mass of a progenitor star to explode as a supernova. In Chapter 9.1.3 we will go in to more detail about the Chandrasekhar mass limit and supernova.

8.8 MASSIVE STARS

Stars with masses above $7M_\odot$ are called massive stars. The smallest of them go through a super AGB phase, see previous section. The stars with masses above $9.25M_\odot$ have cores large enough to go through many different fusion processes until there is no more energy to be gained.

On the main sequence massive stars burn hydrogen into helium via the CNO chain, described in section 8.3. The CNO fusion rate increases with temperature. Because there is a direct correlation between stellar mass and core temperature, massive stars burn really bright and fast. They will exhaust their hydrogen in around a million years, a thousand times faster than a solar type star, and that is even though massive stars have at least 10 times more mass to begin with.

When the hydrogen has been exhausted in the core, the core collapses and thereby increases in temperature. These stars never become red giants, where the core is dormant. Instead the temperature increase from the collapse is enough to ignite helium burning. The core is hot and dense enough for helium burning to start smoothly. The core is now around 10^8 degrees and it is burning helium in the triple alpha process, described in formulas 8.1 & 8.2.

Massive stars in this phase are on the horizontal branch in the HR-diagram, albeit a HB up in the supergiant region, as these stars are very large and very bright, see Figure 8.2. Energy from the core helium fusion raises the temperature of the surrounding hydrogen layer, and thereby ignites hydrogen fusion there. The outer layer expands causing the surface temperature of the star to drop, which means the star moves towards the right in the HR-diagram. The movement is horizontal, because the brightness stays the same, as the decrease in temperature is countered by a larger surface area.

Once helium is exhausted in the core, the outward pressure is no longer supported and the core collapses, thereby converting potential energy into kinetic energy, i.e. heat. The mass is large enough that the resulting temperature reaches 5×10^8 K, where fusion of carbon can take place. Carbon fusion is described above in formula 8.8. While the core is burning carbon, the layer above is burning helium, and the layer above that is burning hydrogen.

As the dominant element in the core changes, so does the density and temperature needed for the next nucleosynthesis step. Heavier elements have larger nuclei and a larger positive charge. The positive nuclei repulse each other and the density (closer proximity) and temperature (speed/energy of individual nuclei) needed, for a fusion reaction to take place, increase accordingly. The collapse of the core after the exhaustion of each element provides exactly this. The mass of the star is so great, that gravity can squeeze the core making it denser, and the release of potential energy becomes the source of temperature increase.

Once carbon is exhausted in the core, the core collapses and heats up to about 2×10^9 K, which is hot enough for oxygen to burn:

$$^{16}O + ^{16}O \longrightarrow \; ^{32}S + \gamma \tag{8.9}$$
$$\longrightarrow \; ^{31}P + ^{1}H \tag{8.10}$$
$$\longrightarrow \; ^{28}Si + ^{4}He \tag{8.11}$$
$$\longrightarrow \; ^{31}S + n \tag{8.12}$$
$$\longrightarrow \; ^{24}Mg + 2^{4}He \tag{8.13}$$

While oxygen is burning in the core there is a shell around it burning carbon. Around this carbon burning shell is a helium burning shell, which again is surrounded by a hydrogen burning shell. This onion like structure signifies massive stars in their late stages.

Once oxygen is exhausted in the core, the core collapses and heats up to about 7×10^9 K, which is hot enough for silicon to burn:

$$^{28}S + ^{28}S \longrightarrow \; ^{56}Ni + \gamma \tag{8.14}$$
$$^{56}Ni \longrightarrow \; ^{56}Fe + 2e^{+} + 2\nu_e \tag{8.15}$$

This is the last nucleosynthesis reaction that takes place in the stellar core. Elements heavier than iron cannot be fused together with the release of energy, to fuse heavier elements costs energy.

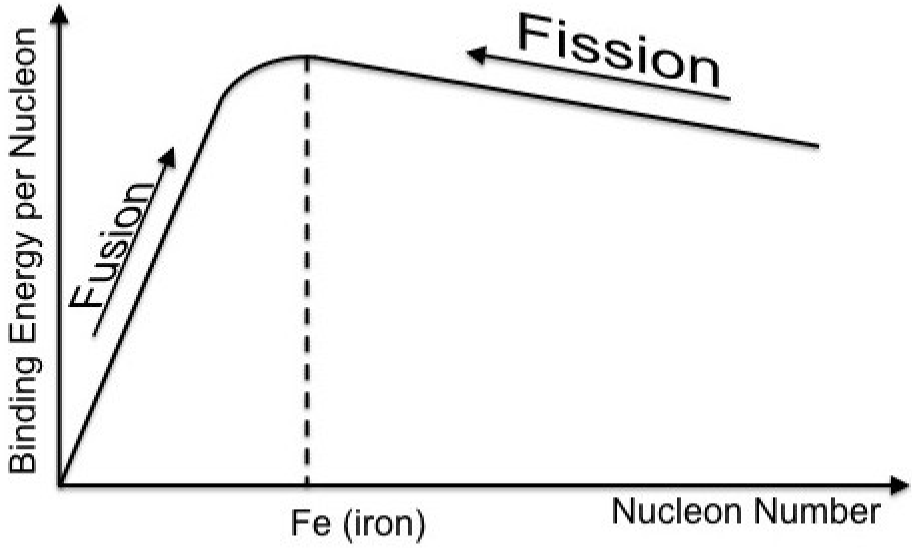


Figure 8.4: A diagram of the binding energy per nucleon vs. the nucleon number. Gaining energy by melding lighter nuclei together is called fusion. Gaining energy by splitting larger nuclei into smaller nuclei is called fission.

The energy released in nucleosynthesis comes from the difference in binding energy for the nuclei, see figure 8.4. As long as two or more nuclei can be fused together into one nucleus with a higher binding energy, then there is energy to be gained. The particles in the newly formed nucleus are tighter bound than they were before in the lighter nuclei. Or in other words, we have to spend more energy to break the new nuclei apart, than if we had tried to break the lighter nuclei apart. This energy difference is what is released in nuclear fusion. Of all the elements, iron has the highest binding energy. Nuclei heavier than iron have smaller binding energies, i.e. their nuclei are not as tightly bound as iron. Actually, if we break a nucleon heavier than iron apart, we will be releasing energy as the newly formed nuclei have larger binding energy.

When the core has exhausted all its silicon, there is no energy to drive the outward pressure and the star will undergo core collapse. The increase in temperature and density will not lead to another nucleosynthesis, and the star will become a core-collapse supernova, which will be discussed further in Chapter 9.

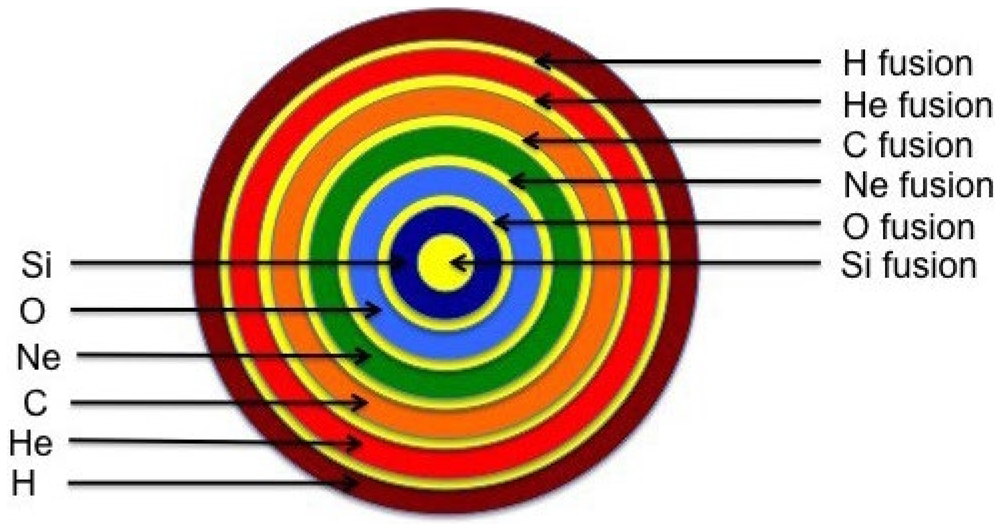


Figure 8.5: Onion like structure for highly evolved massive stars, where each shell is burning a different nuclear fuel.

Highly evolved massive stars have an inner structure that resembles an onion. When helium burning started in the centre it heated the above layer up, so that hydrogen fusion could take place. Above the hydrogen burning layer is a layer of hydrogen not hot enough to burn. When the core had exhausted its helium and started burning carbon it became hot enough to ignite helium burning in the layer above. Above the helium burning layer is a layer of helium not hot enough to burn, and above that the layers from before remained, i.e. a hydrogen burning layer and a passive hydrogen layer. At the end the massive star has a structure of shells upon shells of different nuclear burning separated by passive layers, see figure 8.5. The star is coldest at the surface and hottest in the core. As we go deeper into the star the temperature rises making fusion possible of steadily heavier nuclei.

Massive stars burn bright and fast compared to lower mass stars. Hydrogen burning is faster because the CNO chain burn more efficiently as the temperature is higher. The more massive a star is the higher its core temperature is. Fusion of elements heavier than helium releases less energy. This can also be deduced from figure 8.4, where we see that the energy gained in fusion of successive heavier elements is less and less.

The star needs the same energy output rate for the pressure to balance out gravity. This means that the fusion reactions have to happen at a higher rate as the nuclei fuel increases in size. That means that the star burns through its fuel faster and faster as the fusion element gets larger and larger. Massive stars spend around a few 10^6 years on the MS and around a couple of 10^5 years on the HB. Carbon burning will last a few 10^2 years, oxygen burning half a year and finally silicon burning just one day.
Not only do massive stars have relative short lives, they are rare to begin with. This makes it difficult to predict their evolutionary tracks in the HR- diagram, as we don't have a large number of massive stars in all the different stages.

8.9 SUMMARY

Stellar evolution depends mainly on mass and the HR-diagram is a great tool to understanding the evolutionary paths that stars of different masses take.

Stars in a double star system will evolve as singular stars, if the distance between the two stars is too great for mass transfer to take place. Close binary stars, also called contact binaries, will evolve according to their masses, however, the mass of a member can abruptly decrease or increase, which can cause the star to jump around in the HR-diagram.

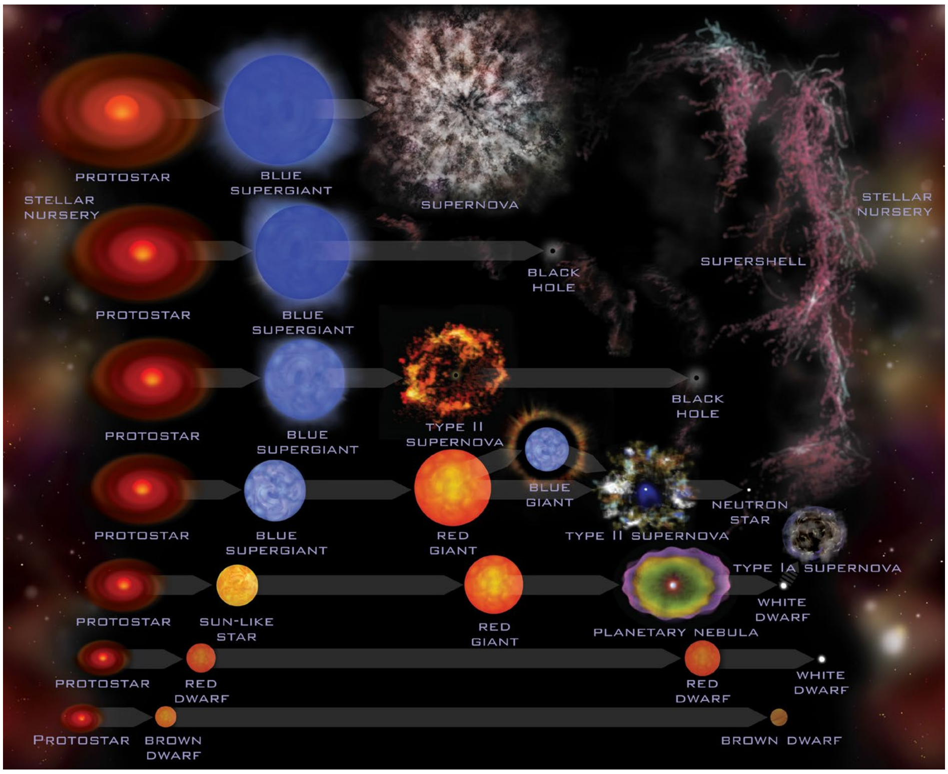


Figure 8.6: This graphic gives a summary of our best current understanding of the evolution of stars, showing their birth, middle age and eventual demise. The lowest mass stars are shown at the bottom and the highest mass stars at the top. Image courtesy of NASA.

Figure 8.6 is a graphic illustration (by NASA) of the different evolutionary paths stars take depending on their masses. The bottom ones are the smallest stars and the top ones the most massive stars.

The different evolutionary stages for stars of all masses are summarised in Table 8.1 below.

Bibliography and Links

Iben has written the best and most solid papers about stellar evolution. Most of this chapter is based on his papers on stellar evolution. For an in depth understanding of this topic I would recommend reading his papers and in particular:

Iben, 1965, 'Stellar Evolution. I. The Approach to the Main Sequence', The Astrophysical Journal, Vol.141, pp. 993.

Iben, 1967, 'Stellar Evolution within and off the Main Sequence', Annual Review of Astronomy and Astrophysics, Vol. 5, pp.571-626.

The helium core ignition value of 0.5 solar masses has been adopted from Hans O. U. Fynbo et al.'s work:
Fynbo et al., 2005, 'Revised rates for the stellar triple-alpha process from measurement of 12C nuclear resonances', Nature, Vol.433, pp.136-139.

The following papers have been used to distinguish between evolutionary paths of different sized stars as shown in Table 8.1.

Burrows et al., 1993, 'An expanded set of brown dwarf and very low mass star models', The Astrophysical Journal, Vol.406, pp.158.

Garcia-Berro & Iben, 1994, 'On the formation and evolution of super-asymptotic giant branch stars with cores processed by carbon burning', The Astrophysical Journal, Vol. 434, pp. 306-318.

Poelarends et al., 2008, 'The Supernova Channel of Super-AGB Stars', The Astrophysical Journal, Vol.675, pp.614.

Siess, 2006, 'Evolution of massive AGB stars. I. Carbon burning phase', Astronomy and Astrophysics, Vol. 448, pp.717.

	Initial mass of star							
Fusion	$M \leq 0.08 M_{Sun}$	$0.08 M_{Sun} \leq M \leq 0.26 M_{Sun}$	$0.26 M_{Sun} \leq M \leq 0.5 M_{Sun}$	$0.5 M_{Sun} \leq M \leq 2 M_{Sun}$	$2 M_{Sun} \leq M \leq 5 M_{Sun}$	$5 M_{Sun} \leq M \leq 7 M_{Sun}$	$7 M_{Sun} \leq M \leq 9 M_{Sun}$	$9 M_{Sun} \leq M$
Core H	-	Main Sequence						
Shell H	-	-	Red Giant Branch			-	-	-
Core He	-	-	-	He Flash	Horizontal Branch			
Core He	-	-	-	Horizontal Branch				
Shell He	-	-	-	AGB			-	-
Core C	-	-	-	-	-	-	Upper HB	
Shell He	-	-	-	-	-	-	Super-AGB	-
Demise	-	-	Planetary Nebula				PN/CC-SN	CC-SN
Remnant	BD	He WD	He WD	CO WD			WD/NS/BH	NS/BH

AGB: Asymptotic Giant Brach

BD: Brown Dwarf

BH: Black Hole

CC-SN: Core-collapse Supernova

HB: Horizontal Branch

NS: Neutron Star

PN:Planetary Nebula

WD: White Dwarf

Table 8.1: The different evolutionary phases for stars of all masses.

Ritossa et al., 1999, 'On the Evolution of Stars that Form Electron-degenerate Cores Processed by Carbon Burning. V. Shell Convection Sustained by Helium Burning, Transient Neon Burning, Dredge-out, Urca Cooling, and Other Properties of an 11 Msolar Population I Model Star', The Astrophysical Journal, Vol.515, pp.381.

9 SUPERNOVAE & REMNANTS

Your goals for this chapter are to learn about:

- The different types of supernovae.
- Supernova remnants and the origin of the elements.
- Degenerate Stars, i.e. White Dwarfs and Neutron Stars.
- Black Holes and the Schwarzschild Radius.

When Tycho Brahe observed a supernova in 1572 he named it a new star, hence the name nova, which is new in latin. In China the 1054 supernova was named 'a visiting star', which is fitting as supernovae are transient objects, i.e. only briefly observable compared with the permanence of stars. We now know that a supernova is associated with the demise of a star, where either a stellar remnant is left behind, or the explosion completely disrupts the star and no central object remains.

A massive star will end its life as a core-collapse supernova, which will lead to the formation of either a neutron star or a black hole. A less massive star will end its life as a white dwarf. White Dwarfs (WD), Neutron Stars (NS) and Black Holes (BH) are grouped together as stellar remnants or compact objects. White dwarfs are supported by the pressure of electron degeneracy and neutron stars by neutron degeneracy. Therefore these types of objects are called degenerate stars.

The 4 possible endpoints to stellar evolution can be summarised as:

- Cataclysmic disruption of the star in a supernova explosion. No stellar remnant remains.
- A White Dwarf is formed if the mass does not exceed the Chan- drasekhar mass MCh $\approx 1.2 - 1.4 M_\odot$.
- A Neutron Star is formed causing a supernova explosion. The mass may not exceed the Oppenheimer-Volkoff mass MOV $\approx 1.5 - 2 M_\odot$.
- A Black Hole is formed in a core-collapse, no upper mass limit. We will only see a supernova explosion if the stellar mass is below a certain limit, otherwise the black hole swallows the would-be supernova energy and light.

9.1 SUPERNOVAE

A supernovae (SN) marks the cataclysmic end of a star's life. They are observed as an intense increase in brightness, which easily can outshine the host galaxy. Supernovae are named after the year they were discovered, where the first SN in the year 2016 is named 2016a, the second 2016b and so on.

Supernovae are divided into two types based on their spectrum. Type I are those without any hydrogen lines in their spectra and Type II are those with hydrogen lines. We distinguish between two physically different drivers of the supernova explosion: 1) the collapse of the core and 2) explosive thermonuclear burning. Both mechanisms can release huge amounts of energy in a short time. Confusingly, the type division doesn't reflect the physics, as type Ia are powered by explosive thermonuclear burning and subsequent disruption of the star, and all other types are core collapse driven supernovae.

Observations of supernovae also include constructing a light-curve. Light curves are simply the brightness of a SN plotted as a function of time. Observing supernovae implies many observations over time to monitor the brightness change. Supernova light curves have a steep rise to maximum followed by a slower decline. Generally, core collapse supernovae have a less bright maximum brightness, but the combined energy in photons released over time is far greater. Overall, about two thirds of SN are core-collapse supernovae (type II and type Ib/c), and just below one third are thermonuclear disruptions of a WD (type Ia), see references for Smartt et al., 2009.

9.1.1 CORE-COLLAPSE SUPERNOVA

A Core Collapse Supernova (CC-SN) is the regular endpoint of singular massive stars. When energy production in the core ceases, the core collapses into a neutron star or a black hole. As the above layers fall upon the newly formed NS they bounce back against the now hard surface. The bounced material causes a shock wave spreading outwards. Upon the formation of a neutron star there is a great release of neutrinos, a small fraction of which deposit energy in the stalled shock giving it energy to move outwards. The resulting supernova consists of material moving outwards, while sweeping up the upper layers of the star. The temperature and densities reach conditions where nucleosynthesis ignite. The surplus of energy and neutrons means that elements heavier than iron are also formed.

Below is a detailed description of a CC-SN, please keep in mind that while some events causes other events, a lot of the different parts of a SN happens simultaneously.

The Core Collapse

The star had, up until this point, been producing energy in its core via nucleosynthesis of lighter elements into heavier elements. The structure of the star is one of a highly evolved massive star, with onion layered shells and a nickel-iron core. After the production of iron in the core there is no more energy to be gained by nucleosynthesis and fusion stops. As the energy production in the core was the source of the outward pressure, gravity now takes complete control of the core and it collapses. The core is so massive (above MCh = $1.4M_{\odot}$), that the collapse does not halt until the density of a neutron star is reached. For very massive stars the core is so heavy, that the degenerate pressure from the neutrons is not enough to halt a further collapse to a black hole.

Photodisintegration

As the core collapsed potential energy was converted into kinetic energy and heat. The temperature in the core reaches 1010 K, which means the photons are energetic enough to break the iron into protons, neutrons, α- particles, and photons. This process is called

Photodisintegration. Iron has the highest binding energy so breaking it apart costs energy, which means that the outward pressure drops further. The core collapse has now reached a speed of 70,000 km/s (23% of the speed of light) in the outer core.

Neutrinos

The matter in the core is now pressed so much together that the electrons and protons to combine and thereby form neutrons and neutrinos:

$$p + e^- \rightarrow n + \nu \tag{9.1}$$

Neutrinos interact weakly with matter, which means they can escape from the core, and thereby carry away energy, which further accelerates the col- lapse. The collapse continues until the density has reached ρNS = 10^{17}kg/m^3. At this stage the outward pressure of degenerate neutrons is in equilibrium with the gravitational pull towards the centre. A NS has been formed.

The first detectable sign that a CC-SN is happening is a large burst of neutrinos from the formation of the NS. Exactly that is what was observed, when the CC-SN 1987A marked the end of a massive star. To this date, this is the only object, other than the Sun, from which we have observed neutrinos. SN 1987A was located in one of our nearby neighbour galaxies, and there has not been any CC-SN that close since then, see Figure 9.1.

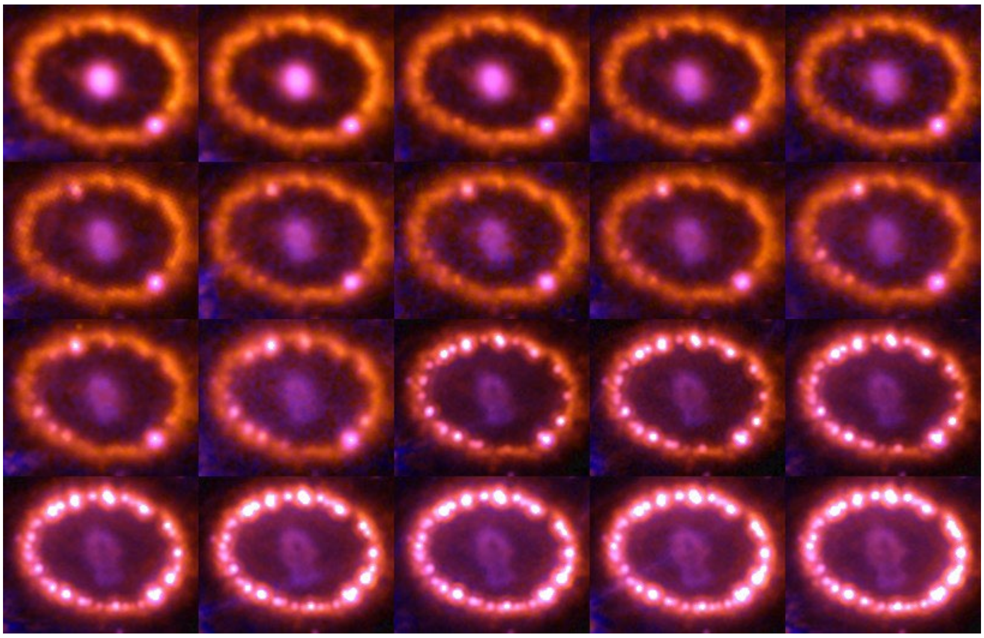


Figure 9.1: A mosaic of SN1987A, the closest SN to date. The ring, ionised by the SN, is made from material lost in the star's evolution. The ring starts to glow more fiercely as the material from the SN is reaching it. Also the inner ejecta from the explosion is slowly expanding, shrouding what is believed to be a NS. Image courtesy of SAINTS Team/Peter Challis/NASA.

Bounce & Shock Wave

The surrounding layers of the star are still falling freely towards the centre of the star, now occupied by the newly formed NS. As the layers above crash into the NS, they bounce off this compact object, which is in a state where its matter cannot be squeezed further together. The in falling material, however, can become denser, which it does upon impact, while it also gains a net velocity outwards in the star. The bounce causes the formation of a massive shockwave which propagates throughout the star. The shock wave is caused by material travelling at great speed colliding into material travelling in the opposite direction. When the speeds involved supersede the speed of sound in that material it causes a shockwave. As the shock wave passes through the onion layered star, unfused elements will reach temperatures and densities high enough for fusion to ignite.

Shock Break Out

The shock stalls, but is reinvigorated by the neutrinos released from the newly formed neutron star. Only about 1% of the neutrinos deposit their energy in the envelope, but calculations show that it is still a huge amount of energy (10^{44} J). The neutrino energy helps drive the dispersion of the outer layers in what we see as the supernova explosion.

As the shock breaks out through the surface of the star, we see light from the supernova for the first time. Fusion in the newly compressed layers also provides an energy release that drives the explosion and dispersion of the outer star. The energy release is enough to eject the outer envelope of the star and briefly let the star outshine its host galaxy.

Supernova Nucleosynthesis

A tremendous amount of energy is released in the core collapse, which allows elements heavier than iron to be formed. The source of the neutrons for this process is the photodisintegration that ripped the iron nuclei apart in the core.

Most of the nuclei heavier than iron are formed by neutron capture. The material in the shocked regions will absorb the neutrons, which, because they don't have an electric charge, can penetrate into the nucleus. There are two kind of neutron capture processes, the s-process and the r-process. The s-process (s for slow) is one where the flux of neutrons is small enough for the receiving element to undergo a β-decay before it receives another neutron. A β-decay is when a neutron in the nucleus is transformed into a proton under the emission of an electron or a positron and associated neutrinos.
The r-process (r for rapid) is one where the flux of neutrons is so large (above $10^{22}/cm^3$) that the receiving elements absorbs many neutrons before it decays. The s-process also takes place in stars, whereas the r-process demands conditions only met in supernovae, where a neutron star is being formed.

A supernova explosion is the only formation channel for some of the heavier elements. It is spectacular to consider that the heavy elements that we handle every day, were made in a core-collapse supernova explosion.

While the collapse of the core and the subsequent supernova explosion only lasted seconds (part of the processes only fractions of a second), the energy released was enormous. Most of the energy released is actually in the form of neutrinos (10^{46} J), with only 1% in the form of visible light (10^{44} J).

Sizes of Exploding Stars

When it comes to stars, the bigger they are, the brighter they also are. With supernovae this is reversed. Very large stars will, at the end of their lives, collapse to form black holes. As their cores rapidly collapse there is no time for the above layers to escape, all light and energy is swallowed up by the newly formed BH and we see no supernova. Faint core collapse supernovae (CC-SN) are caused by stars a bit smaller, as some energy and matter managed to escape the potential well.

Since stars can lose a lot of mass in their evolution it is helpful to only discuss the mass of the remaining core at the time of explosion. Doing so we can conclude that the brightest CC-SN are those with the smallest core masses. However, the stellar core has to be above the Chandrasekhar mass to be eligible for a core collapse in the first place. This discussion is summarised nicely by Eldridge & Tout, 2004:

- $M_{Core} > 15M_\odot$, no display.
- $15M_\odot > M_{Core} > 8M_\odot$, faint SN.
- $8M_\odot > M_{Core} > 5M_\odot$, possibly faint SN.
- $M_{Core} < 5MG$, bright SN.

Three main factors are deciding, which supernovae stars end up as. The initial mass of the star, its metallicity, and its mass loss (or gain in binary systems).

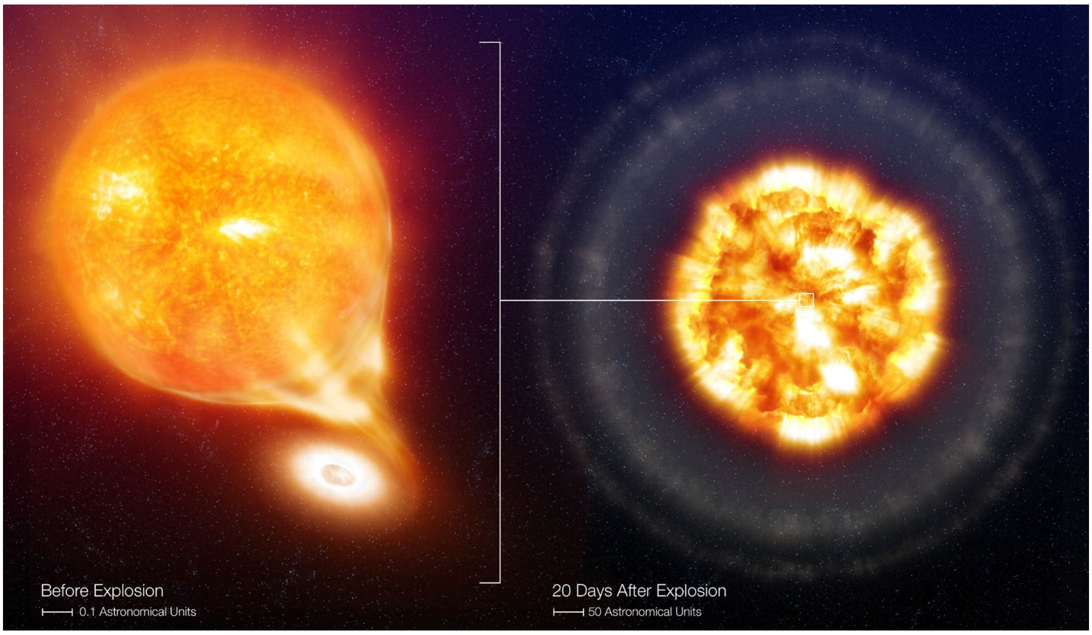


Figure 9.2: An artists illustration of a WD in a close binary system accreting mass from its companion. The right panel is a zoom on the WD as it explodes. Illustration courtesy of ESO.

9.1.2 THERMONUCLEAR EXPLOSION

Supernovae type Ia are unique amongst SN in that they are not core-collapse events. Instead, the SN is a complete destruction of a white dwarf. The progenitor is a low mass star ($0.5M_\odot < M < 7M_\odot$) that has evolved all the way to the white dwarf stage. Instead of fading away like other WDs, these stars are in a close binary system, where they are close enough to accrete mass from their companion, see Figure 9.2. As the WD approaches the limiting mass for a WD (M_{Ch}) it also approaches the conditions for the onset of carbon burning. With the increase in mass an increase in pressure and density follows, which raises the temperature of the core.

At some point a flame front is born, powered by carbon fusion. The ignition mechanism is still not clear. Oxygen fusion is initiated shortly thereafter, but this fuel is not consumed as completely as carbon. With the nucleosynthesis burning in the WD the temperature increases. This does not lead to an increase in pressure or an expansion of the WD, as it would in a star. While the nuclear cores are partaking in fusion, the electrons are still supporting the WD against gravity. The degenerate electron pressure is only weakly influenced by temperature, so the temperature can continue to increase, while the density increases. The result is a thermonuclear runaway reaction.

The runaway thermonuclear reaction is a carbon-oxygen flash, much like the helium flash that marks the onset of helium core burning in low mass stars. However, in the case of type Ia supernovae there is no stellar envelope to conceal the energy. Most of the carbon will be converted into iron before the electron degeneracy is lifted. The thermonuclear fusion causes an energy release of the order of 10^{44} J, which is easily enough to completely disrupt the WD. The energy released in the explosion also causes an extreme increase in luminosity, which make these supernovae brighter than their CC- SN cousins. A contributing factor to this difference is that the CC-SNe have a large amount of neutrinos that carry away energy.

In the fusion processes during the explosion large amounts of unstable isotopes are formed, most dominant amongst them radioactive nickel. Nickel decays into cobalt, which in turn decays into iron, which is stable. During the decays photons are emitted, which power the exponentially decaying light curves.

This type Ia category of supernovae produces consistent peak luminosity because of the uniform mass of white dwarfs that explode via the accretion mechanism. The stability of this value allows these explosions to be used to measure the distance to their host galaxies.

9.1.3 TYPE I

Type I supernovae have no hydrogen lines in their spectra. Typically the progenitor star, i.e. the star it was before it exploded, was either a white dwarf or a Wolf-Rayet star, a type of massive star with extreme stellar winds. Both of those types of stars have completely lost their outer hydrogen envelopes, which explain the lack of hydrogen lines.

Type I supernovae are divided into the types Ia, Ib, and Ic. SNe type Ia are not core-collapse supernovae, but rather the complete destruction of a star close to the Chandrasekhar mass limit. SNe type Ib and type Ic are core-collapse supernovae, and because they are very similar, they are in the following bundled together.

Type Ia

Supernovae type Ia are the violent destruction of a WD pushed over the Chandrasekhar mass limit, MCh. The resulting runaway thermonuclear explosion is described above in section 9.1.2. These events are extremely bright as all of the material in the star is used as fuel. Since the material in the SN originates in a WD, the lack of hydrogen lines in the spectra is easily explained. In the runaway thermonuclear explosion large amounts of radioactive nickel is ejected, which decays into cobalt and iron. This decay can be seen in the light curves for Ia supernovae, where after maximum they follow an exponential decay.

There is a subset of SN type Ia, where the progenitor is not a WD in a binary system but a single star. In this scenario the envelope of an old AGB star has not been removed and the core is almost heavy enough, and thereby hot enough, to ignite carbon. The core would accrete mass from the envelope and eventually approach the Chandrasekhar mass, which would ignite the carbon there. The onset of carbon burning would be a carbon flash, much like the helium flash for the degenerate helium cores of RGB stars. The result is a thermonuclear supernova explosion, i.e. a type Ia, which completely disintegrates the star. The supernova would look like a type Ia with a type II mask.

Type Ib/c

Type Ib/c supernovae are essentially core-collapse supernovae, and thus more related to the type II supernovae. The progenitor is a massive star that has lost its hydrogen envelope at some point in its evolution. A likely progenitor candidate is a Wolf-Rayet star, a type of star that has extreme winds that blow the outer layers away. Another progenitor scenario is a star in a binary system, where the companion star accreted the outer envelope of the progenitor. This would also lead to a massive star progenitor devoid of hydrogen. Where type Ib supernovae have no hydrogen in their spectra, type Ic supernovae also have no helium.

9.1.4 TYPE II

Type II supernovae have hydrogen lines in their spectra. They are regular core collapse supernova (CC-SN) of massive stars. Type II supernovae have fainter maximum brightnesses and a more irregular light curve decay than their type I cousins. The irregularities of the light curves in the late stages are caused by material ejected in the SN interacting with mass lost during their evolutionary phases. Figure 9.1 shows SN1987a's interactions with previously lost material.

Type II supernovae are divided into several sub categories, of which the most important ones are described below.

Type IIP

A supernova type IIP shows hydrogen in its spectra and has a light curve that goes through a plateau phase. The progenitor must have retained a large part of its original hydrogen as the light curve plateau is powered by a moving hydrogen ionisation front. The stellar type that typically becomes a type IIP supernova is a red supergiant ($8.5 - 16.5 M_\odot$) with an extended hydrogen rich envelope.

Type IIL

Type IIL supernovae have light curves which decay linearly. The progenitor scenario is much like that of type IIP, the difference being the smaller amount of hydrogen in type IIL.

Type IIn

Supernovae type IIn are signified by narrow emission lines in their spectra and slowly declining light curves. These objects have unusually dense circumstellar gas, with which the ejecta interact. The progenitor star is thought to be some of the largest we see explode as supernovae.

9.1.5 SUPERNOVA REMNANTS

Massive stars end their evolution in a supernova explosion. The collapse of the stellar core leads to violent ejection of the outer layers, which remain as an expanding gas cloud with a speed around 10,000-20,000 km/s. This cloud is called a supernova remnant (SNR). When the light from the supernova explosion has faded and the remaining material has cooled off, there is no energy source to keep it shining. The material will expand and eventually mix with the interstellar medium.

When the stellar remnant in the core is a fast spinning magnetic neutron star, then there is a rich source of photons from synchrotron radiation (see Chapter 5.4.1). This is beautifully exemplified in the Crab nebula Figure 5.6, where the hydrogen has been ionised by the energetic UV photons from the NS. The blue and green optical light comes directly from synchrotron radiation.

As a comparison it is interesting to look at planetary nebulae, as they are not remnants of supernova explosions, but remnants left by stars with smaller masses. From Chapter 8.6we know that a planetary nebula is the expelled envelope of a star in its last evolutionary phase. The envelope forms an expanding shell around the stellar remnant (WD) -the core of the original star. Planetary nebula are often spheric symmetric and the expansion is around 20-30 km/s, so much slower than the expansion in a SNR.

9.2 WHITE DWARFS

A white dwarf has a larger density than a regular star. To give an idea about how dense a white dwarf is, it is helpful to imagine the sun squeezed into the size of Earth. This is physically possible, because it is supported by electron degeneracy, which allows it to pack matter closer together than when the matter is in the state of plasma -the normal state of matter in a star. White dwarfs are therefore also often called degenerate stars. Although WD are relatively small they have very hot surfaces.

White dwarfs are formed from the cores of highly evolved (old) stars. These stars have often lost their envelopes either after a Red Giant phase or in a violent AGB pulsation. The remaining stellar core is not heavy enough to start the next stage of fusion. For helium WD the next stage would be he- lium burning and for CO white dwarfs it would be carbon-oxygen burning. Regardless of how far the core had managed to evolve the situation remains the same. Without an energy source to drive the outward pressure the core collapses under its own gravity.

A stellar core in which outward pressure comes from the temperature (i.e. movement on a molecular level) alone, can only support itself and the above layers up to a certain point. When the mass exceeds this limit the core becomes degenerate. The collapse halts and a new equilibrium is reached when the density is high enough for quantum effects to play a part. That is when the pressure from a degenerate gas opposes the gravity. For a white dwarf it is the electrons that become degenerate.

Electrons are fermions and therefore they obey Pauli's exclusion principle, which states that the electrons cannot occupy the same quantum state. Electrons come with spin, either up or down, which means that each energy level available in the core can be occupied by two electrons, one of each spin. The quantum energy states then fill up, starting with the lowest level. The electrons in the highest levels are so energetic, that their movements act as a high outward pressure.

The Pauli exclusion principle is also the reason behind why there is room for two, and only two, electrons in the lowest energy state for atoms. There is an universal law that everything will try to occupy the lowest possible energy state. Say you let go of the pencil in your hand it will not just hover in the air, it will fall to its lowest state. In principle all the electrons would like to be in the lowest state in the atoms, but Pauli's exclusion principle prohibits them from all occupying the lowest level, there is only 'space' for one of each kind of electron, spin up and spin down.
Greatly simplified, electron degeneracy is a quantum mechanical effect that restricts how closely matter can be compacted.

White dwarfs cannot be more massive than the Chandrasekhar mass, MCh = $1.4M_{\odot}$. Electron degeneracy can only support a star with a mass up to $M \leq M_{Ch}$, above it and the force of gravity is so great that the electrons are pushed to react with the protons to form neutrons. If that happens you no longer have a white dwarf but a neutron star.

9.3 NEUTRON STARS

Neutron stars are even denser than white dwarfs. If the mass of the core of the exploding star is larger than $1.4M_{\odot}$, then the core will contract beyond the WD density until it reaches the NS density.

When the pressure of the degenerate electron gas fails to withstand the gravitational collapse, the matter is pressed even further together, which forces the electrons and protons to combine and thereby form neutrons. The collapse is at this stage still ongoing. As the number of neutrons in the nuclei grows the binding energy decreases. When the density reaches 4×10^{14}kg/m^3 the neutrons start to leak out of the nuclei. The collapse continues until the density has reached ρNS = 10^{17}kg/m^3, where matter is no longer ordered into nuclei. This is because ρNS is actually the same as the density of atomic nuclei.

The matter is now a neutron porridge with about 0.5% protons and elec- trons. The neutrons act combined as a neutron gas, which is degenerate because neutrons are fermions (like electrons) and they have to obey the Pauli exclusion principle, where two neutrons cannot occupy the same energy state. The neutron gas has an outward pressure, which is large enough to withstand the pull of gravity towards the centre. The neutron star is in equilibrium.

The radius of a neutron star is around 10 km. Furthermore, the radius of a degenerate star (also WD) is inversely proportional to the cubic root of the mass. That is:

$$R_{NS} \propto \frac{1}{M^{1/3}} \tag{9.2}$$

This implies that the larger the mass, the smaller the radius. Such a relation is directly opposite to what we have come to expect from regular stars.

The upper mass limit for a NS is called the Oppenheimer-Volkoff limit. The upper limit is not well defined as the calculations involve interaction probabilities, which are currently not accurately known. Estimations have been made and they put the limit in the range $M_{OV} \approx 1.5 - 2M_{\odot}$, although for some models the mass could be as high as $5M_{\odot}$. A stellar core above M_{OV} undergoing collapse will form a black hole and not a neutron star.

Unlike normal stars NS have a well defined hard surface. The upper crust is a metallic solid with the density growing rapidly inwards. Most of the star is a neutron superfluid (superfluid is a liquid with no viscosity (super smooth), i.e. with no friction between constituents of the fluid). The core may consist of rare matter, like a quark-gluon soup, but it is really guess work.

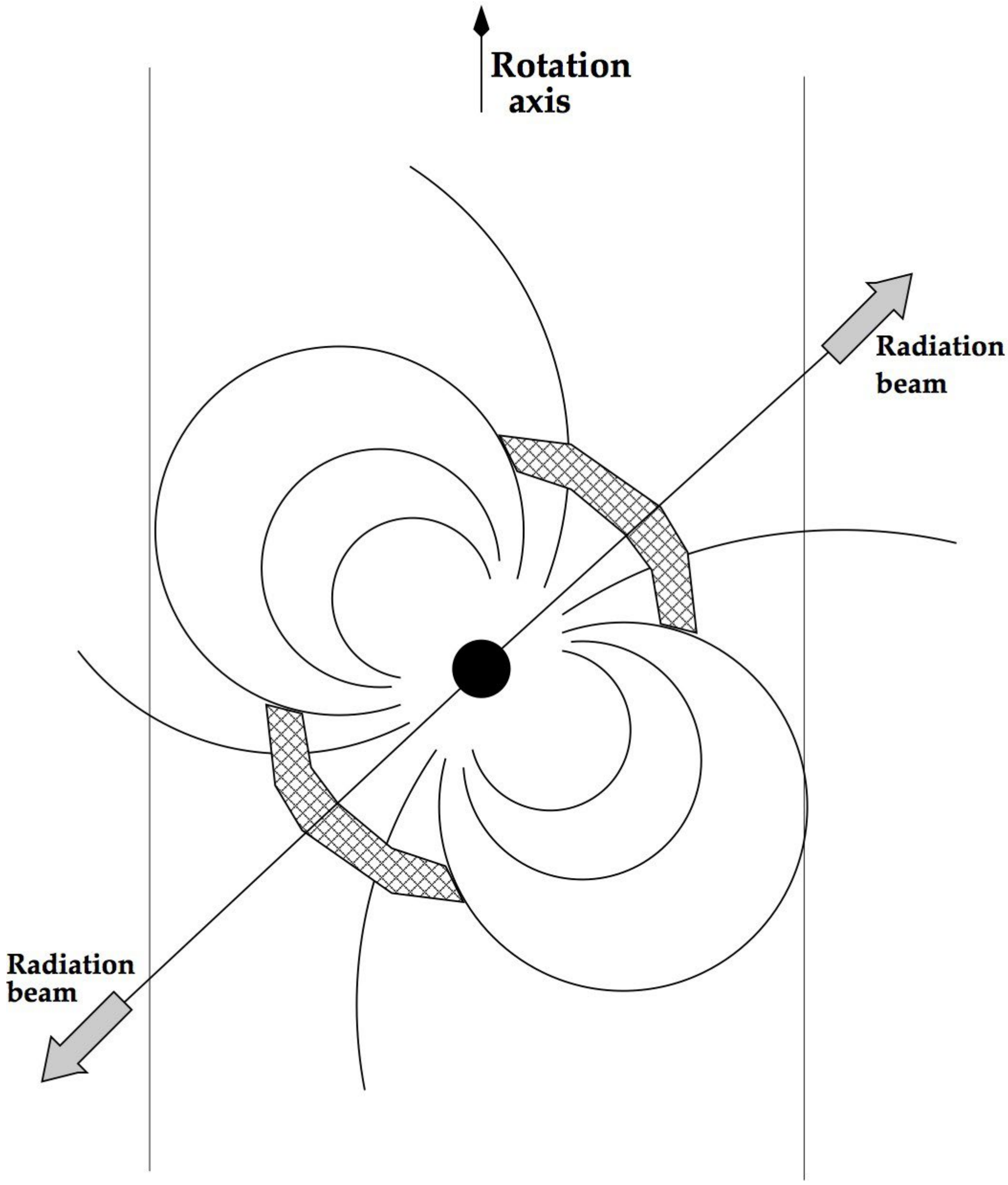


Figure 9.3: The basic pulsar model. The black disk in the middle represents a neutron star and the curved lines the magnetic field. The meshed areas are where the emission originates. The vertical lines marks the light cylinder where a point co-rotating with the neutron star would move at the speed of light. Illustration is from Karina Kjær's Master Thesis: 'The Rotational Evolution of Radio Pulsars', University of Copenhagen 2003.

The neutron star will inherit two major properties from its progenitor star, namely the rotation and the magnetic field. Neutron stars rotate very rapidly, this is a consequence of the universal law of conservation of angular momentum. The mass is now inside a much smaller radius, but its angular momentum has to stay the same. This is resolved by spinning faster. The concept of conservation of angular momentum is familiar to us, when we look at a figure skater, who spins faster when the arms are pulled closer to the body and spins slower when the arms are stretched out. Often the spin of a NS is several hundred times per second.

In the collapse the magnetic field of the progenitor star is compressed, which result in a very strong field.

9.3.1 RADIO PULSARS

The existence of neutron stars was theorised in the 1930s, shortly after the discovery of the neutron. However, they were not discovered until 1967, where a new type of pulsating radio emitting object was discovered. This type of object was named 'pulsar' (pulsating radio source) and identified as being a neutron star. Today more than 1,000 pulsars have been detected.

Pulsars emit very regular radio pulses. The shape of the pulse is sharp, meaning that the intensity rises fast and falls off again fast. That is consistent with a rotating object sending out a cone of light, which only passes us quickly. A maritime comparison would be that of a lighthouse, which also helps to illustrate that you would only see a pulsar if the beam passes your position. Because of the short time between pulses, the object emitting it must be rotating very fast. If it was a regular star rotating this fast, then the surface of the star would be moving faster than the speed of light. Therefore, the object must be very small. The neutron star was the obvious candidate.

One of the least understood aspects of radio pulsars is the emission mechanism. According to the magnetic dipole model, the pulsar is a fast spinning neutron star with a strong magnetic field, which axis is inclined to the axis of rotation, see figure 9.3. Radiation is emitted along the magnetic axis. When the magnetic axis is not aligned with the axis of rotation, the emission will sweep across our line of sight, once every rotation of the NS. We then detect the emission as pulses, it is briefly there, and then it is not there. The reoccurrence of the emission is so precise that you can set a clock by it. The energy of the emission is tapped from the rotational energy of the NS, which means that pulsars are generally slowing down.

Pulsars also shine because of their temperature, but because their radii are so small the luminosity from them is small. The Crab Pulsar, the remnant from the 1054 SN in the Crab constellation, is one of the few to be observed at visible wavelengths.
Another emission coming from pulsars is synchrotron radiation. That emission originates from charged particles trapped in the strong magnetic field of the neutron star. Here the electrons or protons are forced to move along the magnetic field lines, which cause them to emit synchrotron radiation, see Chapter 5.4.1 for more detail.

9.3.2 BINARY SYSTEMS

Many stars are actually part of a double star system, also often referred to as a binary system. For binary systems the proximity of the member stars determine if the stars will influence each other's evolution.

When a neutron star has a companion star that is a giant, there will be a transfer of mass from the giant to the NS. Upon reaching the surface of the NS, the material will burn violently and fast, and we observe the system as a X-ray source. Some systems show periodic variations of the X-ray emission, causing the NS to be classified as a 'X-ray pulsar'.

Double neutron star systems have also been found. The most famous one is the system PSR 1913+16 (the naming is actually its coordinates). This system consists of one radio pulsar and one radio quiet NS. The two neutron stars interact only via gravity, i.e. there is no mass transfer. The orbital values for the system are known to great precision, because of the impeccable timing of the pulsar in the system. Detailed observations of this system have revealed that the system is loosing orbital energy, the orbital period it slowing down with 89 10^{-9}s every orbit. The energy lost corresponds exactly to the amount of energy one would expect to be radiated away as gravitational waves. Albeit not a direct observation, it was the first indirect evidence of gravitational waves.

9.4 BLACK HOLES

A stellar remnant with a mass larger than MOV will go on contracting beyond the density of a neutron star. For such a remnant, not even the degenerate neutron gas can oppose the gravitational force. In fact nothing can stop the collapse, and the remnant contracts to form a black hole.

In December 1915, just a few months after Einstein published his famous equations, Karl Schwarzschild derived a general relativistic expression for the gravitational field around a non-rotating spherical, electrically neutral black hole. The definition of a black hole is a region in space-time that cannot communicate with the external universe. In black holes the gravitational force dominates completely, so much so that all the mass occupies only a single point with an infinitely high density. Such a construction, mathematically, is also called a singularity, and it means that Einstein's equations collapse. Black holes have extremely strong gravitational fields, because all the mass is concentrated at one point.

The gravitational field near the object is so severe that nothing can escape to the outside world, not even light. A black hole has a boundary region, inside of which the escape velocity is faster than the speed of light. Since nothing is faster than the speed of light, nothing can escape this region. The boundary region has a radius, which can be calculated. Because of Karl Schwarzschild's early contributions on the topic the radius is named after him, RS. The Schwarzschild radius can be calculated by using the expression for escape velocity 1.12 found in Chapter 1.5.3:

$$V_{escape} = \sqrt{\frac{2G \times M}{R}}. \tag{9.3}$$

By substituting V_{escape} with the speed of light c, and R = R_S we get:

$$R_S = \frac{2G \times M_{BH}}{c^2}, \tag{9.4}$$

where M_{BH} is the mass of the black hole and G is the gravitational constant.

The Schwarzschild radius for a $1M_{\odot}$ is 3 km. When we compare that with a NS, where the radius is 10 km and the mass 1.4MG, we get a sense of how much more matter has to be compressed in order to get a black hole. We would not expect to find any BH with R_S = 3km, as the stars that end up as BHs have larger cores, and the smallest BHs will have masses just over the M_{OV} limit.

Because not even light can escape the object, the object will appear black. No emission at all wavelengths. This also implies that, whatever is inside a black hole, we have no means of finding it out. It is as if there is hole in our space-time, which is now removed from our universe. The state of matter inside is unknown to us, and the concept of a singularity remains somewhat abstract. Because there is no coupling between quantum field theory and general theory of gravitation, i.e. because there is no quantum theory of gravitation, we cannot even begin to understand, what takes place inside a black hole.

There are only three parameters which can be known for a black hole: Mass, charge, and rotation. It is unlikely that a BH will have a charge, but rota- tion is common in stars and probably also in BHs.

A common misconception about BHs is that they mercilessly suck up everything. It is true, that a BH will accrete anything, but only if it comes inside its sphere of influence. If the Sun turned into a black hole today, we would merrily continue to orbit around it in the same orbit as we have always done. The mass of the Sun has not changed, and that is all that matters for our orbit around it. That the Sun is now a BH will influence our climate and our survival, as Earth would freeze over, but Earth would not be sucked into it.

To have a BH with a RS similar to the distance between the Sun and Earth (1 AU), the mass of the BH would have to be:

$$M = \frac{R_S c^2}{2G} = 5 \times 10^7 M_\odot \tag{9.5}$$

That means that the Sun would have to have sucked up 50 million Sun sized stars in order to have a boundary all the way out to Earth. Black holes are often portrayed as dangerous sink holes, but we should keep in mind that they are rather small and that space is quite big.

Black holes are hard to observe, because they do not emit any light. However, if they are in a binary system, where they are accreting mass from a companion star, we will be able to detect emission from the mass being transferred. Mass flowing from a star onto a black hole, will assemble in a disk around the black hole called an accretion disk. The disk will be visible, often in X-rays, as the material heats. The innermost rim of the accretion disk is where matter is actually falling into the black hole.

Supermassive black holes (SMBH), often sitting at the centres of galaxies, are typically indirectly observed by looking at the orbits of stars in the vicinity. The orbital speed of stars and the physical size of the central object hints at a high density object without any emission. SMBHs have masses around $10^6 - 10^9 M_\odot$.

9.5 GRAVITATIONAL WAVES

Gravitational waves (GW) are a natural consequence of Einstein's field equations. Where Newton explained gravity by placing a gravitational potential in a space time, Einstein's equations both construct space time and define a gravitational potential in this space time. In the General Theory of Relativity (GTR) mass influences space time, by telling it how it must curve. If you look at the general formula for GTR, you then have mass and geometry on one side of the equation and on the other side a stress-energy tensor. The mass gives values to the stress energy tensor and the distribution of the stress energy determines the curvature of space. That means, that there is stress in space time, when there is mass. The gravitational field can then be defined as a distribution of mass, energy, and momentum.

If the distribution of mass in a system changes, the curvature around the system will change. This change in curvature of space-time will propagate outwards from the system as gravitational waves. Although a GW is massless and propagates at the speed of light, it still carries information and energy. Any system causing GWs will be losing energy and momentum.

Gravitational waves are described as ripples in the curvature of space-time, which propagate with the speed of light. Although the image of water ripples is a good 2-dimensional analogy, we have to remember that the ripples are in the space time itself, and not some material on the surface of it. A gravitational wave carries information about the change in the gravitational field, due to changes in the distribution of mass, energy, and momentum caused by the moment of the object(s). GW cause fluctuations in the metric (space) they pass. The fluctuations are perpendicular to the direction of propagation. The fluctuations are a stretching or a compression of matter, but only in a direction, which is perpendicular to the direction of the wave. A GW wave will only influence matter, so they cannot influence massless particles (e.g. photons).

With GW we distinguish between indirect and direct measurements. Indirect measurements are those, where we can see that a system must be emitting GW, and the loss in energy (angular momentum) closely fits the expected loss due to GW radiation. GWs have been indirectly measured since the discovery of PSR 1913+16.

Direct measurement was until very recently something out of our grasp. Despite a genius thought out detector, involving light, mirrors and long sealed off chambers, GWs had proven to be illusive. The LIGO detector, consisting of two parts (one in Hanford, WA, and one in Livingston, LA), detected in September 2015 the first gravitational wave GW150914 (naming based on the date). This is the first wave to be directly detected, and the signal originated in a BH-BH merger. The peer reviewed publication and press release followed in February 2016, marking it the largest discovery in recent time. We now know for sure that gravitational waves exist.

Bibliography and Links

For the advanced reader I recommend Eldrige and Tout's work, which not only show which progenitor star leads to which type of core-collapse SN, they also investigate the dependence of metallicity in the star and its mass loss history.
Eldridge & Tout, 2004, The progenitors of core-collapse supernovae', Monthly Notices of the Royal Astronomical Society, Vol.353, pp.87.

For an observational approach, Smartt et al. investigated a 10 year sample of CC-SN and their progenitors, their work published in this paper: Smartt et al., 2009, 'The death of massive stars I. Observational constraints on the progenitors of Type II-P supernovae', Monthly Notices of the Royal Astronomical Society, Vol.395, pp.1409-1437.

Further reading on supernovae:
Arnett, W. D., 1980, 'Analytic solutions for light curves of supernovae of Type II', The Astrophysical Journal, Vol.237, pp.541

Chevalier, 1976, 'The hydrodynamics of Type II Supernovae', The Astro- physical Journal, Vol.207, pp.872:

For the advanced reader I recommend the book 'Black Holes, White Dwarfs, and Neutron Stars' by Sharpiro & Teukolsky. Fundamental descriptions of WD, BH, and NS are based on this book.

The number of pulsars catalogued in 2010 was 1135 according to: Ridley, J. P. & Lorimer, D. R., 'Isolated pulsar spin evolution on the diagram', 2010, Monthly Notices of the Royal Astronomical Society, Vol.404, pp.1081.

The Australia Telescope National Facility (ATNF) is a great source of all pulsars and their observational values:
http://www.atnf.csiro.au/people/pulsar/psrcat/

For more details on pulsar emission see models described in:

1) Lyutikov et al., 1999, 'On the nature of pulsar radio emission', Monthly Notices of the Royal Astronomical Society, Vol.305, pp.338.
2) Wright, 2003, 'An empirical model for the radio emission from pulsars', Monthly Notices of the Royal Astronomical Society, Vol. 344, pp.1041.

The first observation of a gravitational wave is published here:
By the LIGO Scientific Collaboration and the Virgo Collaboration, 2016, 'Observation of Gravitational Waves from a Binary Black Hole Merger', Physical Review Letters, Vol.116, Issue 6, id.061102.
Also available for free here:
http://arxiv.org/abs/1602.03837

Abbott et al., 2016, 'Astrophysical Implications of the Binary Black Hole Merger GW150914', The Astrophysical Journal Letters, Vol.818:L22.
http://iopscience.iop.org/article/10.3847/2041-8205/818/2/L22/pdf

For an overview of Gravitational waves from binary compact objects: Belczynski, K. et al., 2001, 'A Comprehensive Study of Binary Compact Objects as Gravitational Wave Sources: Evolutionary Channels, Rates, and Physical Properties', The Astrophysical Journal, Vol.572, Issue 1, pp. 407- 431.
Available for free here:
http://arxiv.org/abs/astro-ph/0111452

POSTFACE

The intention of this work is that everybody will have access to knowledge about astronomy. The level is aimed at first year university students, but I hope than anyone who finished high school will be able to read it and learn. Should there be passages that could use clarification, then I would appreciate feedback. The same goes for typos and grammar mistakes.

I can be contacted at: Astronomy.bookboon@gmail.com

I would like to thank my friends and family for support and encouragement while writing this. Special thanks to Dr. Helen Roberts and Dr. Yuri Beletsky.

In this second edition I would like to additionally thank Mr Domenico Romano from NW Italy for making me aware of an error in the previous version of this book.