

CfE Advanced Higher Physics – Unit 1 – Astrophysics

Astrophysics

Historical Introduction

The development of what we know about the Earth, Solar System and Universe is a fascinating study in its own right. From earliest times Man has wondered at and speculated over the ‘Nature of the Heavens’. It is hardly surprising that most people (until around 1500 A.D.) thought that the Sun revolved around the Earth because that is what it seems to do! Similarly most people were sure that the Earth was flat until there was definite proof from sailors who had ventured round the world and not fallen off!

It may prove useful therefore to give a brief historical introduction so that we may set this topic in perspective. For the interested student, you are referred to a most readable account of Gravitation which appears in “Physics for the Inquiring Mind” by Eric M Rogers - chapters 12 to 23 (pages 207 to 340) published by Princeton University Press (1960). These pages include astronomy, evidence for a round Earth, evidence for a spinning earth, explanations for many gravitational effects like tides, non-spherical shape of the Earth, precession and the variation of ‘g’ over the Earth’s surface. There is also a lot of information on the major contributors over the centuries to our knowledge of gravitation. A briefer summary can be found in the Introduction of “A Brief History of Time”, by Stephen Hawking. An even more concise history of gravitation is included as an introduction to these notes.

Claudius Ptolemy (A.D. 120) assumed the Earth was immovable and tried to explain the strange motion of various stars and planets on that basis. In an enormous book, the “Almagest”, he attempted to explain in complex terms the motion of the ‘five wandering stars’ - the planets. He suggested that the universe consisted of nested spheres, one within the other and that each independent spheres motion was responsible for the independent motion of certain objects, with the stars on the outmost sphere, their relative positions “fixed”. This is known as a geocentric (Earth centred) model.

Nicolaus Copernicus (1510) insisted that the Sun and not the Earth was the centre of, not only the solar system, but the universe. He was the first to really challenge Ptolemy. He was the first to suggest that the Earth was just another planet, centred only within the lunar sphere. His great work published in 1543, “On the Revolutions of the Heavenly Spheres”, had far reaching effects on others working in gravitation. This is known as a heliocentric (Sun centred) model.

Tycho Brahe (1580) made very precise and accurate observations of astronomical motions. He did not accept Copernicus’ ideas. His excellent data were interpreted by his student Kepler.

Johannes Kepler (1610) Using Tycho Brahe’s data he derived three general rules (or laws) for the motion of the planets. He could not explain the rules.

Galileo Galilei (1610) was a great experimenter. He invented the telescope and with it made observations which agreed with Copernicus’ ideas. His work caused the first big clash with religious doctrine regarding Earth-centred biblical teaching. His work “Dialogue” was banned and he was imprisoned. (His experiments and scientific method laid the foundations for the study of Mechanics).

Isaac Newton (1680) brought all this together under his theory of Universal Gravitation explaining the moon's motion, the laws of Kepler and the tides. In his mathematical analysis he required calculus - so he invented it as a mathematical tool!

1.4 Gravitation

Consideration of Newton's Hypothesis

It is useful to put yourself in Newton's position and examine the hypothesis he put forward for the variation of gravitational force with distance from the Earth. For this you will need the following data on the Earth/moon system (all available to Newton).

Data on the Earth

"g" at the Earth's surface	= 9.8 m s ⁻²
radius of the Earth, R _E	= 6.4 x 10 ⁶ m
radius of moon's orbit, r _M	= 3.84 x 10 ⁸ m
period, T, of moon's "circular" orbit	= 27.3 days = 2.36 x 10 ⁶ s.
take	$\frac{R_E}{r_M} \approx \frac{1}{60}$

Assumptions made by Newton

- All the mass of the Earth may be considered to be concentrated at the centre of the Earth.
- The gravitational attraction of the Earth is what is responsible for the moon's circular motion round the Earth. Thus the observed central acceleration can be calculated from measurements of the moon's motion:

$$a = \frac{v^2}{r}$$

Hypothesis

Newton asserted that the acceleration due to gravity "g" would quarter if the distance from the centre of the Earth doubles i.e. an inverse square law.

Acceleration due to gravity, $a = g \propto \frac{1}{r^2}$

- Calculate the central acceleration for the Moon: use $a = \frac{v^2}{r}$ or $a = \frac{4\pi^2 R}{T^2} m s^{-2}$.
- Compare with the "diluted" gravity at the radius of the Moon's orbit according to the hypothesis, viz. $\frac{1}{(60)^2} \times 9.8 m s^{-2}$.

Conclusion

The inverse square law applies to gravitation.

Astronomical Data

Planet or satellite	Mass/kg	Density/kg m ⁻³	Radius/m	Grav. accel./m s ⁻²	Escape velocity/m s ⁻¹	Mean dist from Sun/m	Mean dist from Earth/m
Sun	1.99×10^{30}	1.41×10^3	7.0×10^8	274	6.2×10^5	--	1.5×10^{11}
Earth	6.0×10^{24}	5.5×10^3	6.4×10^6	9.8	11.3×10^3	1.5×10^{11}	--
Moon	7.3×10^{22}	3.3×10^3	1.7×10^6	1.6	2.4×10^3	--	3.84×10^8
Mars	6.4×10^{23}	3.9×10^3	3.4×10^6	3.7	5.0×10^3	2.3×10^{11}	--
Venus	4.9×10^{24}	5.3×10^3	6.05×10^6	8.9	10.4×10^3	1.1×10^{11}	--

Inverse Square Law of Gravitation

Newton deduced that this can be explained if there existed a Universal Constant of Gravitational, given the symbol G.

We have already seen that Newton's "hunch" of an inverse square law was correct. It also seems reasonable to assume that the force of gravitation will be dependent on both of the masses involved.

$$F \propto m, F \propto M \text{ and } F \propto \frac{1}{r^2} \quad \text{giving } F \propto \frac{Mm}{r^2}$$

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

where $G = 6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$

Consider the Solar System

$M = M_S$, the mass of the Sun and m is m_p , the mass of any planet.

Force of attraction on a planet is: $F = \frac{GM_S m_p}{r^2}$ (r = distance from Sun to planet)

Now consider the central force if we take the motion of the planet to be circular.

$$\text{Gravitational force } F = \frac{GM_S m_p}{r^2}$$

$$\text{Central Force } F = \frac{m_p v^2}{r} \quad \text{force of gravity supplies this central force.}$$

$$\text{so } \frac{GM_S m_p}{r^2} = \frac{m_p v^2}{r} \quad \text{and} \quad v = \frac{2\pi r}{T}$$

$$\frac{GM_S m_p}{r^2} = \frac{m_p \left(\frac{2\pi r}{T}\right)^2}{r} \quad \text{giving} \quad \frac{GM_S m_p}{r^2} = \frac{m_p}{r} \frac{4\pi^2 r^2}{T^2}$$

$$\text{rearranging} \quad \frac{r^3}{T^2} = \frac{GM_S}{4\pi^2}$$

Kepler had already shown that $\frac{r^3}{T^2}$ was a constant for each planet and, since M_S is a constant, it follows that G must be a constant for **all** the planets in the solar system (i.e. a **universal** constant).

Notes:

- We have assumed circular orbits. In reality, orbits are elliptical.
- Remember that Newton's Third Law always applies. The force of gravity is an action-reaction pair. Thus if your weight is 600 N on the Earth; as well as the Earth pulling you down with a force of 600 N, you also pull the Earth up with a force of 600 N.
- The Gravitational force is very weak compared to the electromagnetic force (around 10^{39} times smaller). Electromagnetic forces only come into play when objects are charged or when charges move. These conditions only tend to occur on a relatively small scale. Large objects like the Earth are taken to be electrically neutral.

"Weighing" the Earth

Obtaining a value for "G" allows us to "weigh" the Earth i.e. we can find its mass.

Consider the Earth, mass M_E , and an object, mass m , on its surface. The gravitational force of attraction can be given by **two** equations:

$$F = \frac{GM_E m}{R_E^2} \quad F = W = mg$$

where R_E is the separation of the two masses, i.e. the radius of the earth.

thus $mg = \frac{GM_E m}{R_E^2}$ $M_E = \frac{gR_E^2}{G} = \frac{9.8 \times (6.40 \times 10^6)^2}{6.67 \times 10^{-11}} = 6.02 \times 10^{24} \text{ kg}$

The Gravitational Field

In earlier work on gravity we restricted the study of gravity to small height variations near the earth's surface where the force of gravity could be considered constant.

Thus $F_{\text{grav}} = m g$

Also $E_p = m g h$ where $g = \text{constant} (9.8 \text{ N kg}^{-1})$

When considering the Earth-Moon System or the Solar System we cannot restrict our discussions to small distance variations. When we consider force and energy changes on a large scale we have to take into account the variation of force with distance.

Definition of Gravitational Field at a point.

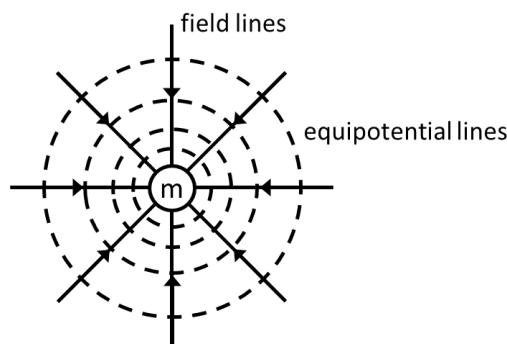
Defined as the force experienced per unit mass in a gravitational field. i.e. $g = \frac{F}{m} (\text{N kg}^{-1})$

The concept of a field was not used in Newton's time. Fields were introduced by Faraday in his work on electromagnetism and only later applied to gravitation.

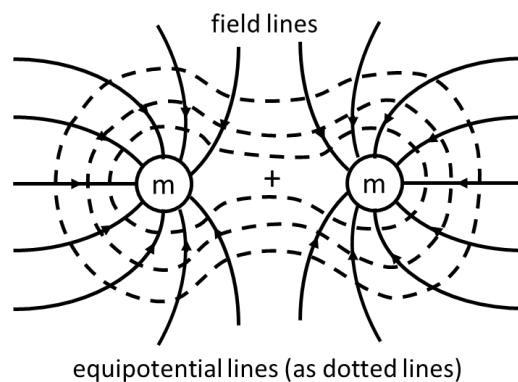
Note that g and F above are both vectors and whenever forces or fields are added this must be done by appropriate vector addition (taking into account direction as well as magnitude!).

Field Patterns (and Equipotential Lines)

(i) An Isolated 'Point' Mass



(ii) Two Equal 'Point' Masses



- Notes: (1) equipotential lines are always at **right angles** to field lines.
 (2) the cross represents the centre of mass of the system
 (3) viewed from far enough away, any gravitational field will begin to look like that of a point mass.

Variation of g with height above the Earth (and inside the Earth)

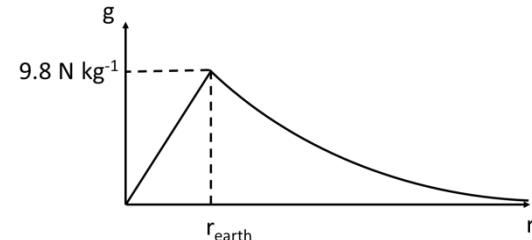
An object of mass m is on the surface, radius r , of the Earth (mass M). We now know that the weight of the mass can be expressed using Universal Gravitation.

Thus

$$mg = \frac{GMm}{r^2} \quad (r = \text{radius of Earth in this case})$$

$$g = \frac{GM}{r^2} \quad (\text{note that } g \propto \frac{1}{r^2} \text{ above the Earth's surface})$$

However the density of the Earth is **not** uniform and this causes an unusual variation of g with radii **inside** the Earth. For a uniform density we would see the following graph for g at radii both within and without the earth



Variation of "g" over the Earth's Surface

The greatest value for "g" at sea level is found at the poles and the smallest value is found at the equator. This is caused by the rotation of the earth.

Masses at the equator experience the maximum spin of the earth. These masses are in circular motion with a period of 24 hours at a radius of 6400 km. Thus, part of a mass's weight has to be used to supply the small central force due to this circular motion. This causes the measured value of "g" to be smaller.

Calculation of central acceleration at the equator:

$$a = \frac{v^2}{r} \quad \text{and} \quad v = \frac{2\pi r}{T} \quad \text{give} \quad a = \frac{\left(\frac{2\pi r}{T}\right)^2}{r} \quad a = \frac{4\pi^2 r}{T^2} = \frac{4\pi^2 \times 6.4 \times 10^6}{(24 \times 60 \times 60)^2} = 0.034 \text{ m s}^{-2}$$

Observed values for "g": at poles = 9.832 m s^{-2}
 at equator = 9.780 m s^{-2}
 difference is 0.052 m s^{-2}

Most of the difference has been accounted for. The remaining 0.018 m s^{-2} is due to the non-spherical shape of the Earth. The equatorial radius exceeds the polar radius by 21 km. This flattening at the poles has been caused by the centrifuge effect on the liquid Earth as it cools. The Earth is 4600 million years old and is still cooling down. The poles, nearer the centre of the Earth than the equator, experience a greater pull.

In Scotland "g" lies between these two extremes at around 9.81 or 9.82 m s^{-2} . Locally "g" varies depending on the underlying rocks/sediments. Geologists use this fact to take *gravimetric surveys* before drilling. The shape of underlying strata can often be deduced from the variation of "g" over the area being surveyed. Obviously very accurate means of measuring "g" are required.

Satellites in Circular Orbit

This is a very important application of gravitation. The central force required to keep the satellite in orbit is provided by the force of gravity.

$$\text{Thus: } \frac{mv^2}{r} = \frac{GMm}{r^2}$$

$$v = \sqrt{\frac{GM}{r}} \quad \text{and} \quad v = \frac{2\pi r}{T}$$

$$\sqrt{\frac{GM}{r}} = \frac{2\pi r}{T} \rightarrow T = 2\pi r \sqrt{\frac{r}{GM}} = 2\pi \sqrt{\frac{r^3}{GM}}$$

Thus a satellite orbiting the Earth at radius, r , has an orbit period, $T = 2\pi \sqrt{\frac{r^3}{GM}}$.

This is Kepler's 3rd Law of planetary motion (or any circular orbital motion, for that matter).

Energy and Satellite Motion

Consider a satellite of mass (m) a distance (r) from the centre of the parent planet of mass M where $M \gg m$.

$$\text{Rearranging: } \frac{mv^2}{r} = \frac{GMm}{r^2} \rightarrow \frac{1}{2}mv^2 = \frac{GMm}{2r}$$

We find that the kinetic energy of a satellite, E_K , is given by $E_K = \frac{GMm}{2r}$

Note that E_K is always positive.

The gravitational potential energy of the satellite in this system is $E_P = -\frac{GMm}{r}$

Note that E_P is always negative.

Thus the total energy is

$$E_{\text{total}} = E_K + E_P$$

$$E_{\text{total}} = \frac{GMm}{2r} + \left(-\frac{GMm}{r} \right)$$

$$E_{\text{total}} = -\frac{GMm}{2r}$$

Care has to be taken when calculating the energy required to move satellites from one orbit to another to remember to include **both** changes in gravitational potential energy and changes in kinetic energy.

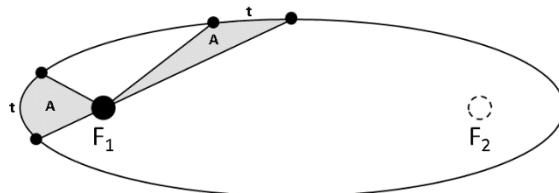
Consequences of Gravitational Fields

The notes which follow are included as **illustrations** of the previous theory.

Kepler's Laws

Applied to the Solar System these laws are as follows:

- The planets move in elliptical orbits with the Sun at one focus,
- The radius vector drawn from the sun to a planet sweeps out equal areas (A) in equal times (t).
- The square of the orbital period of a planet is proportional to the cube of the semi-major axis of the orbit $T^2 \propto r^3$.



Tides

The two tides per day that we observe are caused by the **unequal attractions** of the Moon (and Sun) for masses at different sides of the Earth. In addition the rotation of the Earth and position of the Moon also has an effect on tidal patterns.

The Sun causes two tides per day and the Moon causes two tides every 25 hours. When these tides are in phase (i.e. acting together) **spring** tides are produced. When these tides are out of phase **neap** tides are produced. Spring tides are therefore larger than neap tides. The tidal humps are held 'stationary' by the attraction of the Moon and the earth rotates beneath them. Note that, due to tidal friction and inertia, there is a time lag for tides i.e. the tide is not directly 'below' the Moon. In most places tides arrive around 6 hours "late".

$$\text{Gravitational Potential} \quad V = \frac{\text{work done}}{\text{mass}}$$

We define the gravitational potential (V) at a point in a gravitational field to be the work done by external forces in moving a unit mass, m , from infinity to that point.

We define the theoretical zero of gravitational potential for an isolated point mass to be at **infinity**. (Sometimes it is convenient to treat the surface of the Earth as the practical zero of potential. This is only valid when we are dealing with **differences** in potential.)

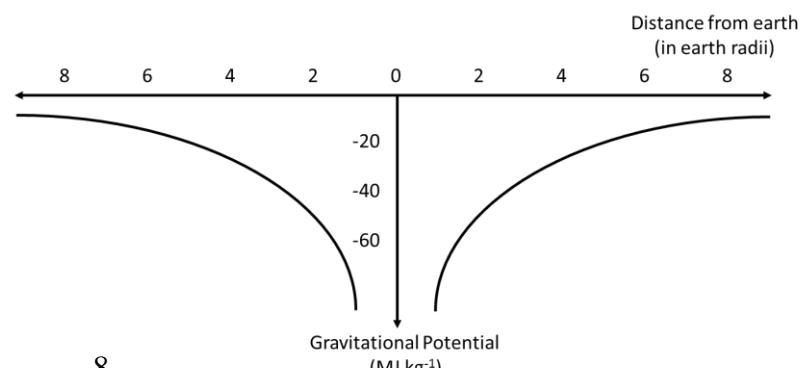
Gravitational Potential at a distance

This is given by the equation below, the units of gravitational potential are J kg^{-1}

$$V = -\frac{GM}{r}$$

The Gravitational Potential 'Well' of the Earth

This graph gives an indication of how satellites are 'trapped' in the Earth's field. Imagine a 3 dimensional,

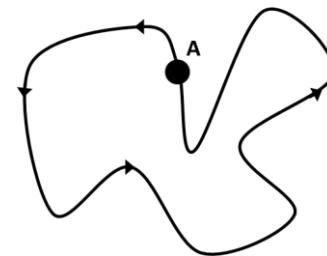


frictionless, curved funnel with a marble rolling around inside. Always “attracted” to the centre, yet never falling in since no energy is lost to the friction. This is a stable orbit.

Conservative field

The gravitational force is known as a conservative force because the work done by the force on a particle that moves through any round trip is zero i.e. energy is conserved. For example if a ball is thrown vertically upwards, it will, if we assume air resistance to be negligible, return to the thrower's hand with the same kinetic energy that it had when it left the hand.

An unusual consequence of this situation can be illustrated by considering the following path taken in moving mass on a round trip from point A in the Earth's gravitational field. If we assume that the only force acting is the force of gravity and that this acts vertically downward, work is done **only** when the mass is moving vertically, i.e. only vertical components of the displacement need be considered.



Thus for the path shown below the work done is zero.

By this argument a non-conservative force is one which causes the energy of the system to change e.g. friction causes a decrease in the kinetic energy. Air resistance or surface friction can become significant and friction is therefore labelled as a non-conservative force.

Escape Velocity

The escape velocity for a mass escaping to infinity from a point in a gravitational field is the **minimum velocity** the mass must have which would allow it to escape the gravitational field i.e. from a point at radius r to infinity.

At the surface of a planet the gravitational potential is given by: $V = -\frac{GM}{r}$.

The **potential energy** of mass m is given by $V \times m$ (from the definition of gravitational potential).

$$E_p = -\frac{GMm}{r}$$

The potential energy of the mass at infinity is zero. Therefore to escape completely from the sphere the mass must be given energy equivalent in size to $\frac{GMm}{r}$.

To escape completely, the mass must just reach infinity where its E_k reaches zero

(Note that the condition for this is that at all points; $E_k + E_p = 0$).

$$\text{at the surface of the planet} \quad \frac{1}{2}mv_e^2 - \frac{GMm}{r} = 0 \quad m \text{ cancels}$$

$$v_e^2 = \frac{2GM}{r}$$

$$v_e = \sqrt{\frac{2GM}{r}} \quad \text{or greater to escape}$$

For the Earth the escape velocity is approximately 11 km s^{-1} .

No regard is given here for the presence of an atmosphere and so no energy is lost to air resistance.

Atmospheric Consequences:

$$\begin{aligned} v_{\text{rms}} \text{ of H}_2 \text{ molecules} &= 1.9 \text{ km s}^{-1} & (\text{at } 0^\circ\text{C}) \\ v_{\text{rms}} \text{ of O}_2 \text{ molecules} &= 0.5 \text{ km s}^{-1} & (\text{at } 0^\circ\text{C}) \end{aligned}$$

When we consider that this is the r.m.s. of a range of molecular speeds for hydrogen molecules, and that a small number of all H_2 molecules will have a velocity greater than v_e , it is not surprising to find that the rate of loss of hydrogen from the Earth's atmosphere to outer space is considerable. In fact there is very little hydrogen remaining in the atmosphere. Oxygen molecules on the other hand simply have too small a velocity to escape the pull of the Earth. The Moon has no atmosphere because the escape velocity (2.4 km s^{-1}) is so small that gaseous molecules will have enough energy to escape from the moon.

Black Holes and Photons in a Gravitational Field

A dense star with a sufficiently large mass and small radius could have an escape velocity greater than $3 \times 10^8 \text{ m s}^{-1}$. This means that light emitted from its surface could not escape - hence the name **black hole**.

The physics of the black hole cannot be explained using Newton's Theory. The correct theory was described by Albert Einstein in his General Theory of Relativity (1915). We will look into this more later. Another physicist called Karl Schwarzschild calculated the radius of a spherical mass from which light cannot escape. It is given by:

$$v_e = \sqrt{\frac{2GM}{r}} \rightarrow c = \sqrt{\frac{2GM}{r}} \rightarrow c^2 = \frac{2GM}{r} \rightarrow r = \frac{2GM}{c^2}$$

Photons are affected by a gravitational field although is there a gravitational force acting on the photon if it has no mass? Photons passing a massive star **are deflected** by that star and stellar objects 'behind' the star may appear at a very slightly different position because of the change in direction of the photon's path. Again, more on this later.

Work done by a photon in a gravitational field and Gravitational Redshift

If a small rocket is fired vertically upwards from the surface of a planet, the velocity of the rocket decreases as the initial kinetic energy is changed to gravitational potential energy. Eventually the rocket comes to rest, retraces its path downwards and reaches an observer near to the launch pad.

Now consider what happens when a photon is emitted from the surface of a star. The energy of the photon, $h\nu$, decreases as it travels to positions of greater gravitational potential energy (remember that this is negative, so we are moving towards zero hence greater) but **the velocity of the photon remains the same**. Observers at different heights will observe the frequency and hence the wavelength of the photon changing, i.e. blue light emitted from the surface would be observed as red light at a distance from a sufficiently massive, high density star. (N.B. this is known as the Gravitational Redshift **not** the redshift caused by receding stellar objects in the expanding universe).

If the mass and density of the body are greater than certain critical values, the frequency of the photon would decrease to zero at a finite distance from the surface and the photon will not be observed at greater distances.

It may be of interest to you to know that the Sun is not massive enough to become a black hole. The critical mass is around 3 times the mass of our Sun.

1.5 General Relativity

Special relativity deals with the observation that the speed of light is a constant for all observers, independent of their (inertial) frame of reference. This is covered in CfE Higher Physics in more detail.

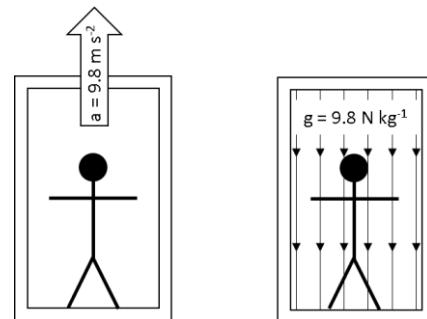
General relativity leads to the interpretation that mass curves space-time and what we perceive as gravity arises from the curvature of space-time. It also deals with the similar effects and consequences of accelerating (non-inertial) reference frames.

As the mass of an object increases, the distortion of space-time will also increase. An example of this is a dense neutron star that has a large mass for its size – it will cause a huge distortion of space-time. Collapsing stars can become very small and occupy a very small space, which in turn causes a point of immense curvature that leads to an *event horizon*.

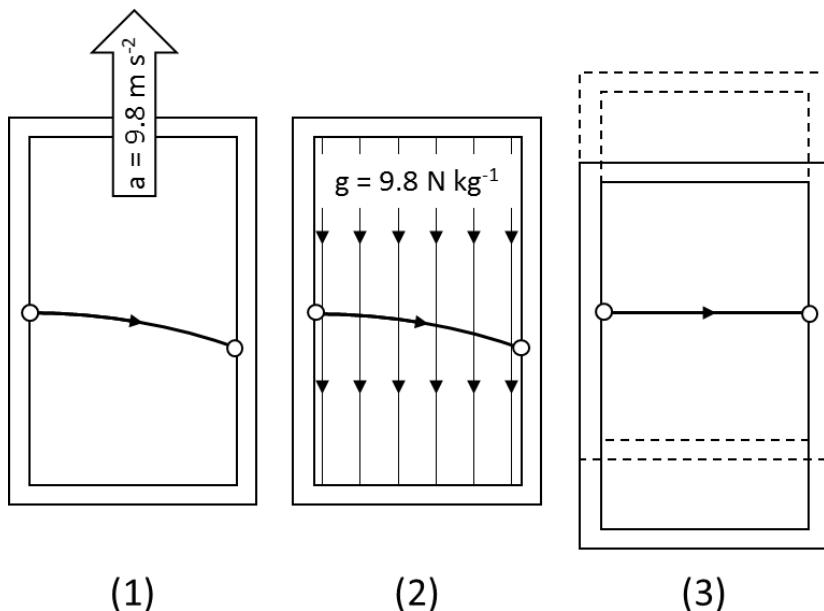
Equivalence Principle

When we carry out Physics experiments in a uniform gravitational field we make the same observations as we do when carrying out the same experiments in a non-inertial (accelerating) reference frame. This effect is known as the Equivalence Principle. The effects of constant acceleration (in a straight line) are the same as those of a uniform gravitational field.

This can be demonstrated by thinking about an observer on board a rocket, without windows. The rocket accelerates and the observer experiences a pseudo-force, which mimics a gravitational force. If the acceleration was 9.8 m s^{-2} then the observer would not know whether they were, in fact, on board a rocket or in a room without windows on the surface of the Earth.



Another example is that of a projectile (1) in a gravitational field and (2) in an accelerating spacecraft.



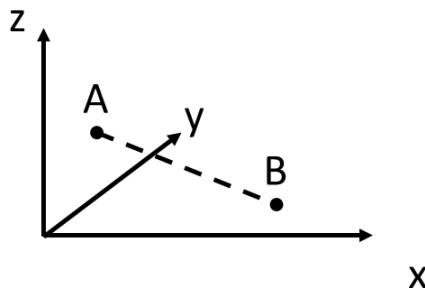
In both instances the projectile follows a curved motion within the experiments frame of reference. Outside the frame of reference, we may not necessarily observe the same motion. For the spacecraft we see a ball move in a straight line while the spacecraft moves (3)!

The same rules apply to light. The motion of light in an accelerating frame of reference appears curved. The motion of light appears curved in a gravitational field.

Space-time

The concept of space-time acknowledges that any particle (or event) is located (or occurs) at not just a specific spatial coordinate but also at a specific time. It literally has to be at the right place and at the right time! Space-time assigns 3 spatial dimensions, x, y and z, as well as a fourth dimension for time, t.

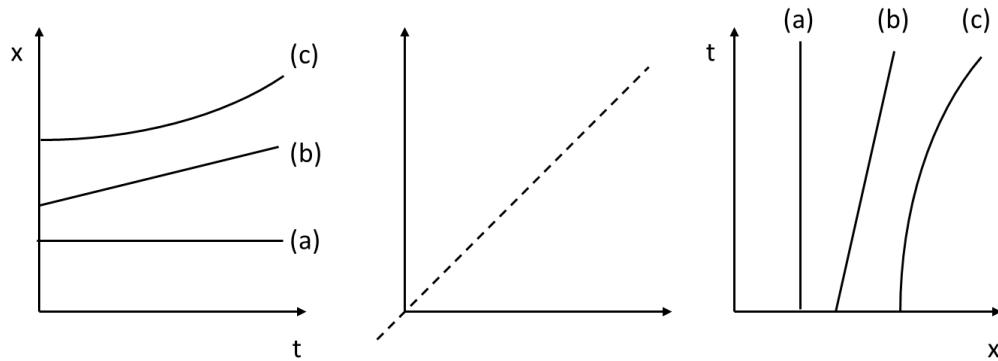
In Newtonian mechanics we use Euclidean space, consisting of three mutually perpendicular directions, often denoted by x, y and z.



A length interval (e.g. AB) is $ds^2 = dx^2 + dy^2 + dz^2$.

The time is stated separately since in Newtonian mechanics there is an absolute background time. It is considered to be the same everywhere.

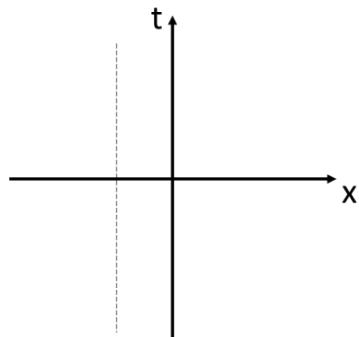
It is impossible to draw four dimensions on paper, but one or more of the spatial dimensions can be represented in a suppressed form with one dimension of space and one of time. The space-time diagram can represent the position of a particle at any given time (here in one spatial dimension). The rules are the same as for a position-time graph but the axes are flipped.



The left hand diagram is a position-time graph, where (a), (b) and (c) represent constant position, constant speed and constant acceleration, respectively. The space-time diagram on the right uses the same principles. Again (a), (b) and (c) represent constant position, constant speed and constant acceleration, respectively.

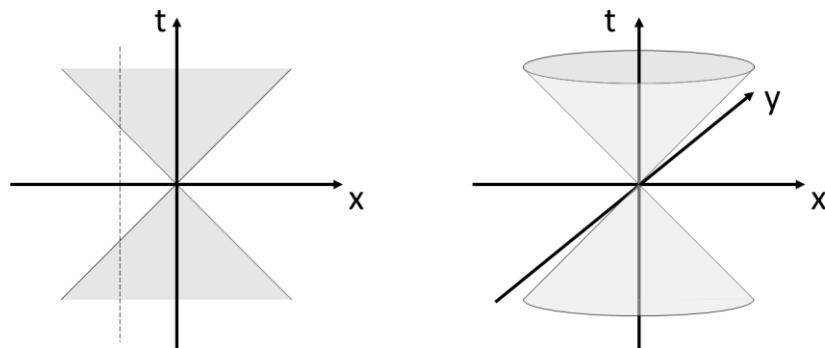
Past, present and future

The following space-time diagram has 4 quadrants shown. This means it can show a particle or event in the past, present and future, in a positive and negative direction. The dotted line may show the behaviour of a stationary object just behind you. Its position is constant and so it is not moving. It was behind you, it is behind you and it will be behind you!



Past, present, future!

Space-time diagrams and light



In the space-diagram on the left light is emitted simultaneously from the origin in all directions (in 2 dimensions this can only be in the positive and negative directions). The beams will travel on the lines $t = x$ and $t = -x$, represented by the diagonal lines. For an event E at the origin the region between the diagonal lines above the x-axis, shaded grey, defines all events in the future that could be **affected by E**. Similarly, the region between the diagonal lines below the x-axis, also shaded, defines past events **affecting E**. The diagram on the right extrapolates this into a second special dimension, y. The shape formed is a cone, this is known as the light cone, and the consequences for events is the same as for the triangle in 1 dimension. 3 dimensions cannot be represented on two dimensional paper, not if we want to include time (which is essential for it to remain a space-time diagram...)

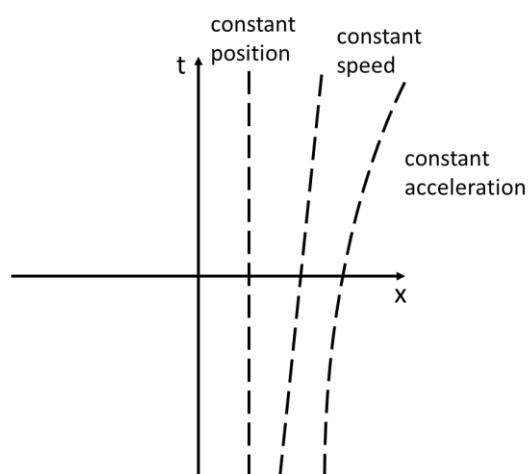
These two shaded regions are referred to as **time-like**.

The regions to the left and right are called **space-like** and events in these regions are unrelated to our event E since nothing (not even neutrinos, according to relativity) can travel faster than the speed of light.

Again the dashed vertical line represents the **world line** of a stationary particle (behind you). Here the value of x remains the same as the time moves forward.

This space-time diagram has 3 world lines on it. Another name for a world line is a geodesic, meaning the path followed by a particle that is acted on by no unbalanced forces apart from maybe gravity. A geodesic path is the shortest distance between two events in space-time.

As can be seen from the diagram that constant velocity has a geodesic that is a straight line at an angle while the geodesic for a stationary particle is a straight line with constant position (x). Acceleration and motion in a gravitational field can be represented by a curve with changing gradient.



Principle of covariance

Physics is described by equations that put all space-time co-ordinates on an equal footing.

This means that the regular flat space-time of special relativity is no longer sufficient and we need to adopt a more complex vision of space-time to help explain this. This is called curved space-time. The amount of matter ‘tells’ space-time how to curve. The more concentrated the matter the more space-time curves.

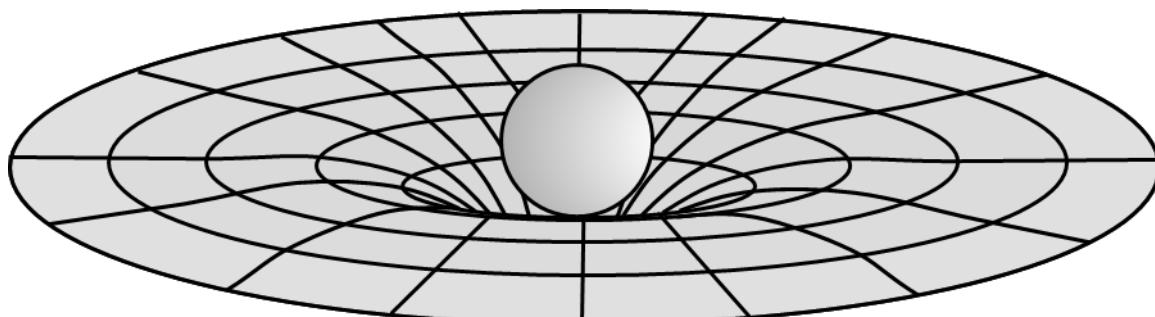
Note

Mathematics was not sufficiently refined in 1917 for a geometric, co-ordinate-independent formulation of physics. Both demands were described by Einstein as general covariance. As Einstein developed his general theory of relativity he had to refine the accepted notion of the space-time continuum into a more precise mathematical framework. Covariance essentially infers that the laws of physics must take the same form in all co-ordinate systems. In other words, all space-time co-ordinates are treated the same by the laws of physics, in the form of Einstein’s field equations.

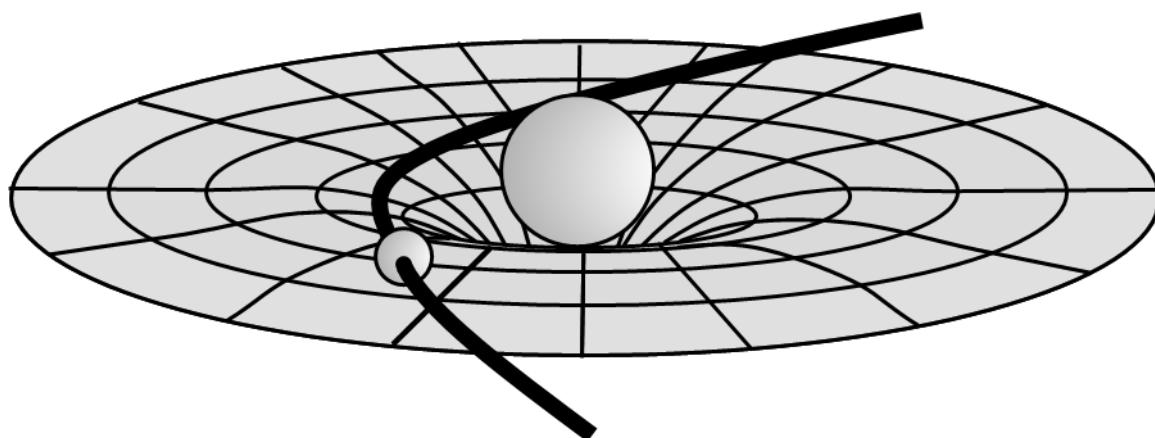
Rubber sheet analogy

Think of space as a stretched rubber sheet. When something heavy is placed on the sheet, the sheet dips. The diagram illustrates this.

Matter tells space how to curve:



Space tells matter how to move:

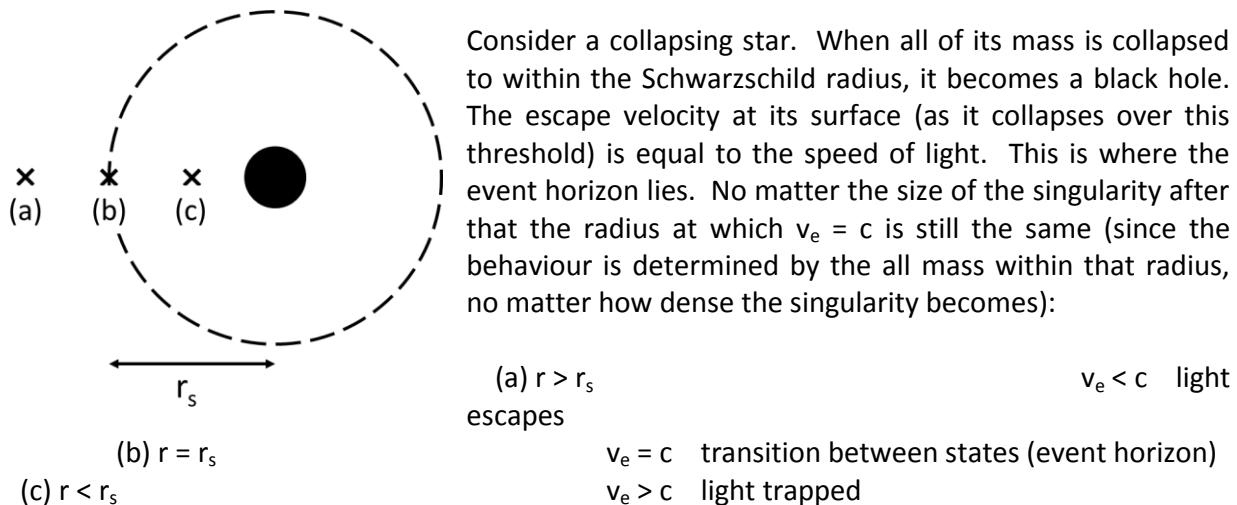


The heavier the object the deeper the resulting gravitational well. Matter tells space how to curve. Once we accept the curvature of space, it is easy to see that smaller objects will move along the straightest possible line that they can in that curved space. This even affects light,

which travels the shortest distance between 2 points. Since space-time is curved the light follows a curved path. From our point of view we see the light “bend”.

Schwarzschild radius and the event horizon

Karl Schwarzschild was a German soldier and physicist. He provided a solution to Einstein's equation on general relativity while at the Russian front. Sadly he died during the war due to a condition that he acquired while a soldier. He derived expressions for the geometry of space-time around stars. He found that as the mass became more dense in a small volume its gravity 'crushes' itself into a phenomenon known as a *black hole*. Up to a certain distance from a black hole everything (including light) is pulled back into the black hole. The minimum distance from which light can just escape is termed the *event horizon* (this occurs at the *Schwarzschild radius* and the two are often referred to as being the same thing).



As such the event horizon is like a one-way valve: it is possible to go from outside the horizon to inside, but impossible to complete the reverse manoeuvre.

$$\text{Schwarzschild radius } r_s = \frac{2GM}{c^2}$$

G is the universal gravitational constant

M is mass of the body

c is the velocity of light in a vacuum.

For the Sun:

$$\begin{aligned} M_{\text{Sun}} &= 2.0 \times 10^{30} \text{ kg} \\ G &= 6.67 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2} \\ c &= 3.0 \times 10^8 \text{ m s}^{-1} \end{aligned}$$

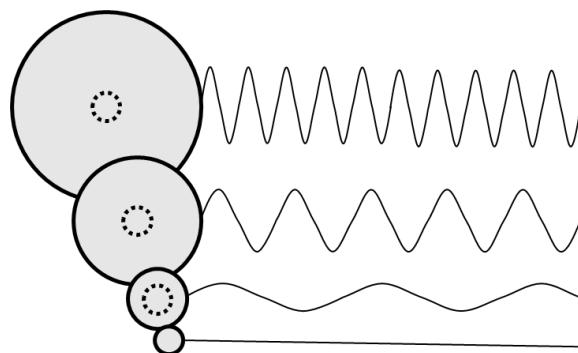
$$r_s = \frac{2GM}{c^2} = \frac{2 \times (6.67 \times 10^{-11}) \times (2.0 \times 10^{30})}{(3 \times 10^8)^2} = 2964 \text{ m}$$

Remember, however that this would not be possible for the Sun! The reasoning behind this lies in how closely the matter that comprises the Sun can be arranged. There is a limit to the packing density of the neutrons within even a neutron star. The neutrons may not overlap. M must be at least about $3M_{\text{sum}}$ for the mass to be sufficient to achieve a black hole.

Gravitational fields and Time dilation

Einstein said ‘An atom absorbs or emits light of a frequency which is dependent on the potential of the gravitational field in which it is situated.’

In a strong gravitational field light appears to be red shifted, its waves ‘stretched’. The light still has to travel at c . A simple way to look at this is that light emitted at the event horizon is so stretched out that it is flat. In the diagram below the dotted line represents the point at which the star will become a black hole (Schwarzschild radius).



For a collapsing star, mass is compressed into smaller volumes and the gravitational field strength at the surface increases. If each wave represents the tick of a clock then, for a black hole a clock seen by a distant observer will appear to stop ticking at the event horizon.

Clocks run slow when in a stronger gravitational field.

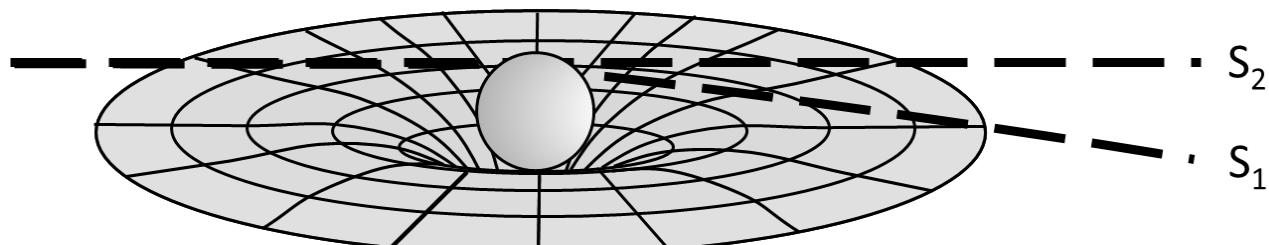
Global positioning systems

To achieve the required accuracy, global positioning systems (GPS) use the principles of general relativity to correct satellites’ atomic clocks, since they run slightly faster, being “higher up” in the Earth’s gravitational field.

Evidence for general relativity

As has been previously stated that general relativity indicates that a beam of light will experience curvature similar to that experienced by a massive object in a gravitational field. In order to see light being ‘bent’ by the Sun we need to view this effect during an eclipse.

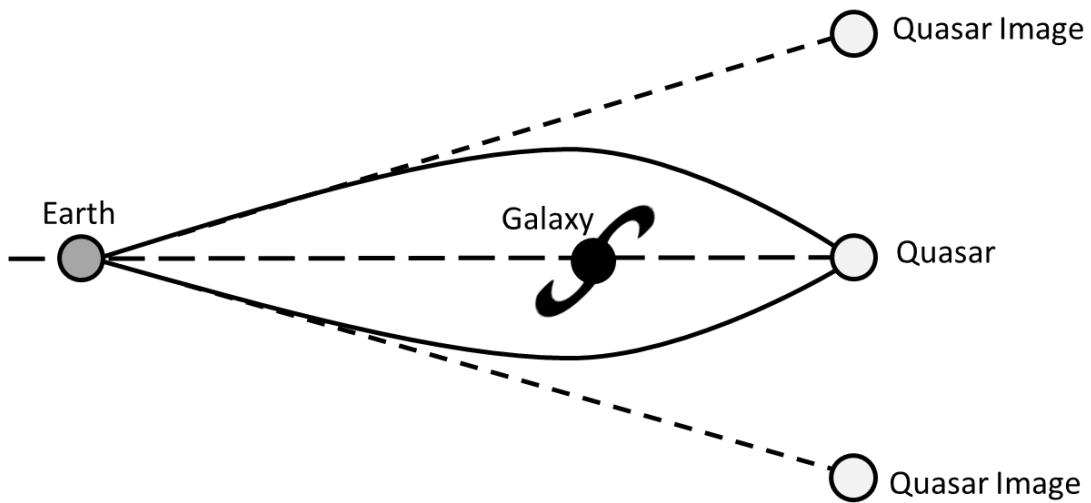
This is done by measuring the position of a star during an eclipse (S_1 – curved path due to curvature of space-time caused by the presence of the Sun) and again when the Sun is in another part of the sky (S_2 – not deviated, since the Sun is not nearby). The angle between S_1 and S_2 can then be compared to the theoretical value.



The change in direction of a path of light leads to the observed phenomenon known as gravitational lensing.

Gravitational lensing

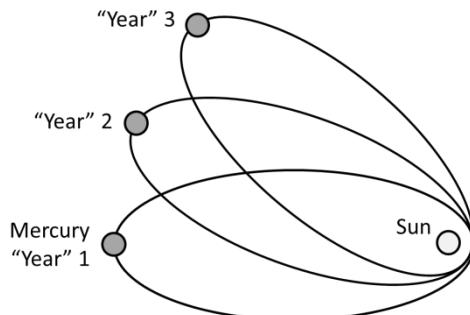
This effect can occur when light from a distant star follows a curved path around a massive object en route to Earth. On Earth we might observe a circle or arc of light, although the original light comes from a single source, for example light from a distant quasar (quasi-stellar radio source).



If light from a distant quasar passes close to a massive galaxy the galaxy will bend the light and images of the quasar will be seen in a circle around the galaxy. This can happen if the Earth, quasar and galaxy are on the same direct line axis. If the earth, quasar and galaxy are not in a straight line then images of the quasar at different distances will be viewed. This phenomenon can enable the distance to a quasar to be calculated if the distance to the galaxy is known. Other information such as the mass of the galaxy can also be calculated.

The precession of Mercury's orbit

In our solar system the planets orbit our central star, the Sun. According to Newton the orbit of a single planet around the Sun will take the form of a constant ellipse. However, in reality, the effect of other planets will cause the orbit to change slightly (or precess) with each rotation. The diagram below shows this effect, but greatly exaggerated.



Observations by the end of the 19th century concurred with calculations using Newtonian methods for all the planets with the exception of Mercury, the closest planet to the Sun. This observed precession agrees with the general relativity principle that the space-time near to the Sun is curved and this changes the predicted orbits of the planets by a tiny amount. For most planets this is virtually negligible. However, in the case of Mercury the precession value is approximately 1/180th of a degree per century. A relativity correction term has to be used to adjust the calculations for Mercury.

The next section begins to deal with Stellar Physics. This is a very large topic which could form the basis for several University courses. As a result it is difficult to describe a small section of the field without providing relevant context. These notes will most likely give you more information than you need in order to provide this context. It will be up to you (with the help of your teacher) to determine what is essential and what is supportive.

1.6 Stellar Physics

Stellar physics is the study of stars throughout their birth, life cycle and death. It aims to understand the processes which determine a star's behaviour appearance and, ultimately, its fate.

Throughout the observable Universe there are estimated to be at least 100 billion (10^{11}) galaxies, each consisting of perhaps 10^{11} stars, meaning that there are in the order of 10^{22} stars in the observable Universe. Naked-eye observations from Earth reveal just a few thousand of these and even less if you have to deal with light pollution.

Our closest star is, of course, the Sun, approximately 150 million kilometres from Earth. After that the next closest is Alpha Centauri at 4.4 light years (42 trillion km). The most distant stars are beyond 13 billion light years (10^{23} km) away. This means that travelling at the speed of light of 3.0×10^8 m s⁻¹, light from these stars would take 13 billion years to arrive on Earth. As the Universe is believed to be 13.7 billion years old, some of this light set off soon after the Universe was created at the Big Bang.

Astronomers classify stars according to their mass, luminosity and colour, and in this part of the course we will explore the processes, properties and life cycle of the major stellar classes.

Properties of stars

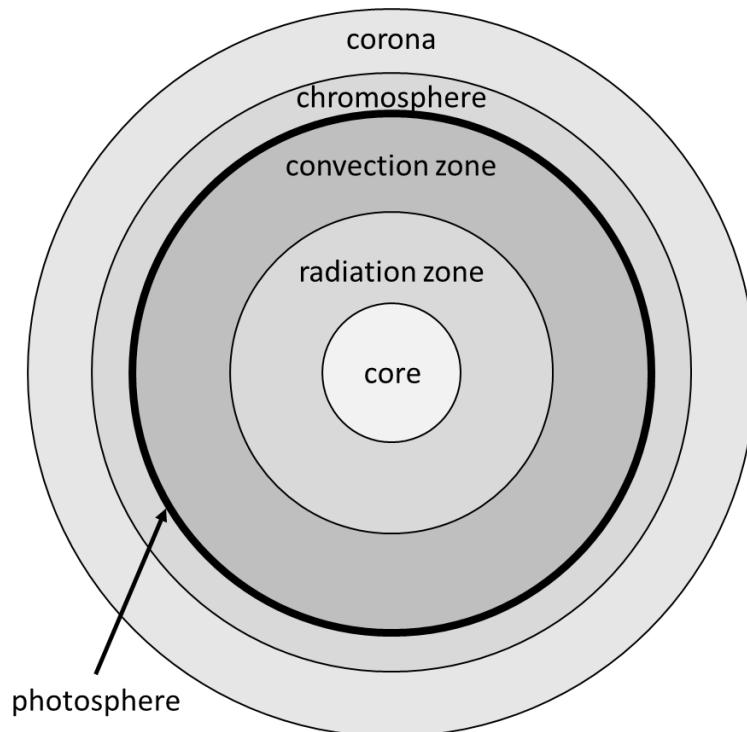
Until the beginning of the 20th century, the process by which our Sun produced heat and light was not well understood. Early theories, developed before the size and distance of the Sun was known, involved chemical production, just like burning wood or coal.

However, by the mid-19th century, when the Sun's size and distance were more accurately estimated, it became clear that the Sun simply couldn't sustain its power output from any sort of chemical process, as all of its available fuel would have been used up in a few thousand years. The next suggestion was that heat was produced by the force of gravity producing extremely high pressures in the core of the Sun in much the same way that pressurising a gas at constant volume causes its temperature to rise. It was estimated that the Sun could produce heat and light for perhaps 25 million years if this were the source of the energy.

At around this time it became apparent from geological studies that the Earth (and therefore the Sun) were very much older than 25 million years so a new theory was required. This new theory emerged in the early years of the 20th century with Einstein's famous $E = mc^2$ equation establishing an equivalence between mass and energy. Here at last was a mechanism which could explain a large, sustained energy output by converting mass to energy.

We now know that an active star produces heat by the process of nuclear fusion. At each stage of a fusion reaction a small amount of mass is converted to energy. The process involves fusing two protons together to produce a deuterium nucleus, a positron, a neutrino and energy. The positron is annihilated by an electron, producing further energy in the form of gamma rays. The deuterium nucleus then combines with a further proton to produce a helium-3 nucleus, gamma rays and energy. Two helium-3 nuclei then combine to produce a single helium-4 nucleus, two protons and energy. The energy released is the *binding energy*, the energy that would be

required to overcome the strong nuclear force and disassemble the nucleus again. The whole process is summarised on the next page.

The Structure of the Sun

The solar zones

Around 99% of the heat production from fusion takes place in the core (although even here this is at a surprisingly low rate per unit volume, around 300 W m^{-3} , which is considerably less than the heat produced by a human body. Stars have such a huge power output because of their enormous size).

Moving outwards from the core of the Sun there is first the radiation zone, where heat transfers radiatively outwards, followed by the convection zone, where pressures are low enough to allow heat transfer by convection. The ‘surface’ of the Sun (i.e. the boundary between the plasma and gas) is the photosphere, above which is the ‘atmosphere’ consisting first of the chromosphere then the corona. The typical temperatures and properties of each layer are given in the table below.

Temperature of solar regions

Region	Typical temperature (K)	Key Property of Region
Core	15×10^6	Fusion process
Radiation zone	10×10^6	Heat transferred by radiation
Convection zone	10×10^6 to 6000	Heat transferred by convection
Photosphere	5800	Effective “surface”
Chromosphere	10000	Inner atmosphere
Corona	1×10^6	Outer atmosphere

Although the corona is at around one million kelvin, the surface temperature we perceive from Earth is that of the photosphere because this is a relatively dense region compared to the corona.

Since most of the energy is released in the core in the form of photons, these must radiate and convect outwards through the extremely dense inner layers of the Sun. This high density means that a photon takes hundreds of thousands of years to reach the surface of the Sun, then just a little over 8 minutes to reach the Earth.

So far we have considered data relating just to our Sun since that is the star about which we know the most. From here on we will consider stars in general.

Size of stars

There is a large variation amongst stars and each star is placed into a ‘spectral class’ (which we will look at later) partly depending on the star’s size. Our Sun, a main-sequence star (again we will deal with this later), has a radius, R_{Sun} , of approximately 696,000 km (109 times that of Earth). White dwarfs (small, dense stars nearing the end of their lives) may have a radius of around $0.01 R_{\text{Sun}}$, whilst at the other extreme supergiants may typically be $500 R_{\text{Sun}}$.

Surface temperature

The surface temperature of a star is, in general, the temperature of the photosphere and depends on many factors, not least the amount of energy the star is producing and the radius of the star.

When viewing the night sky it is apparent that stars are not a uniform white colour. There are red stars, such as Betelgeuse in the Orion constellation, and blue ones, like Spica in Virgo. A dramatic contrast in star colours can be seen by observing the double star Albireo in Cygnus using a small telescope. Here the contrast between an orange and blue-green star is striking.

The reason why stars appear in a wide range of colours is directly related to their surface temperature. The hotter the star the more blue or white it appears. The surface temperature is shown to be directly linked to the irradiance (power per unit area). The energy radiated by the hot surface of a star can be calculated using a modified Stefan–Boltzmann law.

$$\text{power per unit area} = \sigma T^4$$

power per unit area (W m^{-2})

σ is the Stefan–Boltzmann constant ($5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$)

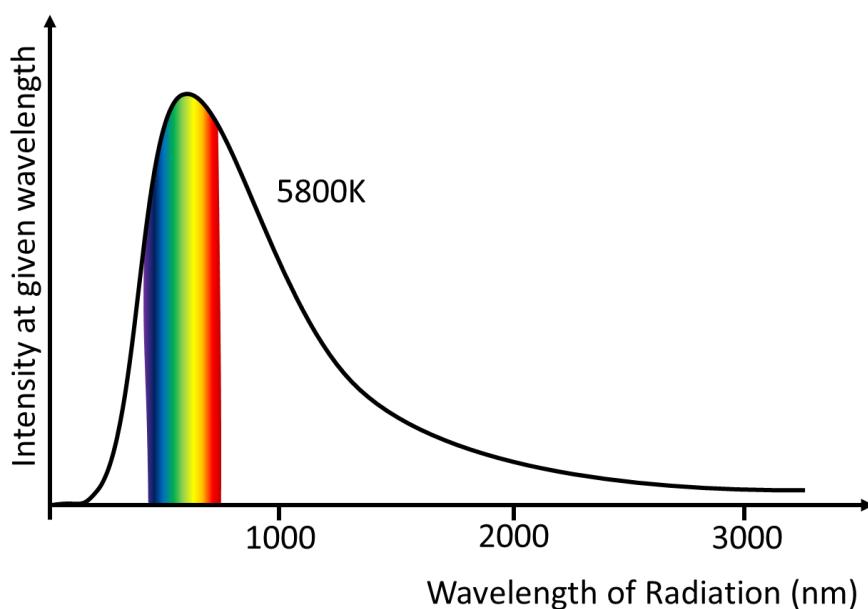
T is the absolute temperature (K).

For example, the power per unit area of our Sun would be calculated as follows:

$$\text{power per unit area} = \sigma T^4$$

$$\text{power per unit area} = (5.67 \times 10^{-8}) \times (5800)^4$$

$$\text{power per unit area} = 64 \times 10^6 \text{ W m}^{-2}$$

Black-body radiation from the Sun

As the power per unit area is proportional to T^4 , this means that as temperature rises the quantity of radiation emitted per unit area increases very rapidly. However, radiation is not emitted at a single wavelength but over a characteristic waveband, as shown above. This diagram is called a black-body radiation diagram. Black-body radiation is the heat radiation emitted by an ideal object and has a characteristic shape related to its temperature. Many astronomical objects radiate with a spectrum that closely approximates to a black body. The above diagram shows this black-body pattern for our Sun, with a surface temperature of 5800 K. The visible part of the spectrum has been plotted to put the graph in context and to demonstrate the relatively small proportion of total energy emitted in the visible waveband.

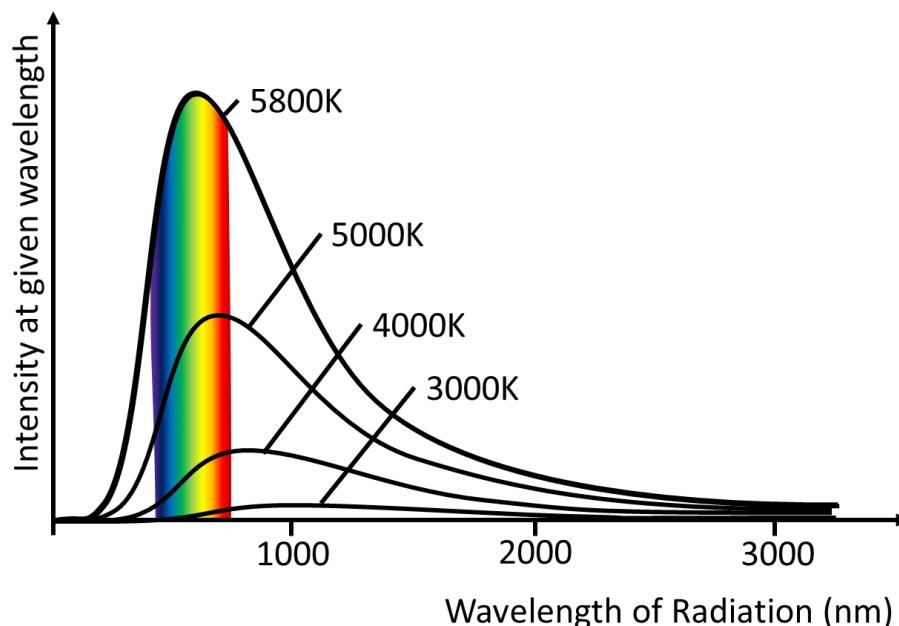
The maximum intensity in the diagram occurs at a wavelength of:

$$\lambda_{\max} = \frac{2.9 \times 10^6}{T} \quad (\text{Wien's law})$$

Which gives a value of 500 nm for the light from our Sun, a value towards the middle of the visible spectrum (400–700 nm). Evolution has resulted in humans having eyesight most sensitive to light with the greatest energy per unit wavelength.

Maximum Wavelength and Surface Temperature

If the emission spectra of stars of different temperatures are plotted the emission curves show the same general shape but different values for their peak radiation. By measuring λ_{\max} from a particular star we can therefore calculate its surface temperature. An alternative approach using just the visible part of the spectrum is to measure the ratio of red to violet intensity and use that to predict λ_{\max} by applying the Planck radiation law.

Mass

Our Sun has a mass, M_{Sun} , of 1.989×10^{30} kg, or roughly 330,000 times that of the Earth. This is often referred to as 1 solar mass. Other stars range in size from roughly 0.08 solar masses to 150 solar masses. The most numerous stars are the smallest, with only a (relatively) few high mass stars in existence. No stars with masses outside this range have been found. Why might this be?

Stars with a mass greater than 150 solar masses produce so much energy that, in addition to the thermal pressure which normally opposes the inward gravitational pressure, an outward pressure caused by photons (and known as radiation pressure) overcomes the gravitational pressure and causes the extra mass to be driven off into space.

Stars with a mass less than 0.08 solar masses are unable to achieve a core temperature high enough to sustain fusion. This arises because of another outward pressure, called degeneracy pressure, which occurs in stars of this size. This pressure is explained in terms of quantum mechanics, which we won't cover here, but the outcome is that this outward pressure prevents the star from collapsing and the core temperature and pressure are never enough to sustain nuclear fusion.

Small stars which fail to initiate fusion become brown dwarfs and slowly dissipate their stored thermal energy and cool down. To give an idea of scale, a star of 0.08 solar masses is approximately 80 times the mass of Jupiter, so brown dwarfs occupy a middle ground between the largest planets and the smallest stars.

Brightness, luminosity and magnitude.

Brightness – an ambiguous term, usually meaning the energy per unit area received from or emitted by a celestial object – its flux. The unit for brightness is W m^{-2} . This value will be at its maximum on the stars surface and will decrease by an inverse square law when observed at greater distances. This power per unit area is directly related to the stars surface temperature.

$$\text{power per unit area} = \sigma T^4$$

σ is the Stefan-Boltzmann constant $5.670 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$

T is the surface temperature of the star (K)

For the Sun:

$$\text{power per unit area} = \sigma T^4$$

$$\text{power per unit area} = 5.670 \times 10^{-8} \times (5800)^4$$

$$\text{power per unit area} = 64 \times 10^6 \text{ W m}^{-2}$$

Luminosity – luminosity is the total rate at which energy is radiated by a celestial body over all wavelengths. The Sun's luminosity is about $3.9 \times 10^{26} \text{ W}$. It is calculated using the power per unit area of a star and the surface area of that star:

$$L = 4\pi r^2 P \quad \text{or} \quad L = 4\pi r^2 \sigma T^4$$

L is the luminosity of the star (W)

P is the power per unit area (W m^{-2}) (be careful not to confuse this with conventional power)

r is the radius of the star (m).

For example, we can calculate the luminosity of the Sun:

$$L = 4\pi r^2 \sigma T^4$$

$$L = 4 \times \pi \times (6.96 \times 10^8)^2 \times 64 \times 10^6$$

$$L = 3.9 \times 10^{26} \text{ W}$$

The Sun has a luminosity of $3.9 \times 10^{26} \text{ W}$. The range of stellar luminosities found throughout the universe is from around one thousandths of this value to 10 thousand times this value.

$$\sim 4 \times 10^{23} \text{ to } 4 \times 10^{31} \text{ W}$$

Because the range of luminosities covers eight to nine orders of magnitude, the luminosity can also be described using the magnitude scale. The concept of using a range of magnitudes to measure the brightness of stars was introduced by the Greek astronomer Hipparchus, in the second century BC. His scale had the brightest stars as magnitude 1 whilst the faintest visible to the naked eye were magnitude 6.

The modern definition of luminosity uses the same principle, but describes what the magnitude of a particular star would be when viewed from a distance of 10 parsecs (a parsec is 32.6 light years). A difference of 5 in magnitude corresponds to a 100-fold difference in luminosity. Also as the magnitude increases, the luminosity decreases. In fact, the most luminous stars have

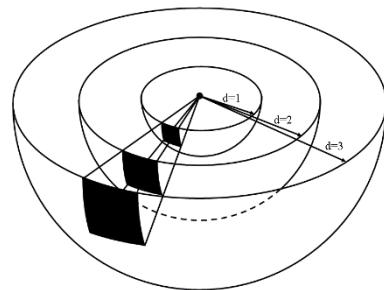
negative magnitudes, e.g. Sirius has an absolute magnitude -3.6 and our Sun $+4.83$. However the apparent magnitude, how bright it **appears**, of Sirius is -1.46 the Sun is -26.7 .

Absolute magnitude – the magnitude a star would appear to have if it at a distance of 10 parsecs from the Sun.

Apparent magnitude – the relative brightness of a star as perceived by an observer on Earth on the magnitude scale.

Apparent brightness

When viewed from Earth, the brightness of a star does not reflect its absolute magnitude, it is quite possible that a high-luminosity star a long way from Earth could appear to be less bright than a low-luminosity star nearby. The brightness of any star when viewed from Earth is its apparent brightness and is a measure of the radiation flux density (W m^{-2}) at the surface of an imaginary sphere with a radius equal to the star–Earth distance.



The radiation flux density is calculated using the inverse square law such that

$$\text{apparent brightness} = \frac{L}{4\pi r^2}$$

L is the luminosity of the star (W)

r is the distance (m) from the star to our observation point, usually Earth.

(notice that if we used the radius of the star itself we would get the power per unit area of the star. This is basically an expanded version of that sphere where the power per unit area diminishes as $\frac{1}{r^2}$)

To apply this to our Sun gives

$$\text{apparent brightness} = \frac{L}{4\pi r^2}$$

$$\text{apparent brightness} = \frac{3.839 \times 10^{26}}{4\pi \times (150 \times 10^9)^2} = 1360 \text{ W m}^{-2}$$

This is the value of the solar constant at the top of Earth's atmosphere. In the UK, measurements of solar radiation at the Earth's surface give a value around 1000 W m^{-2} at midday in midsummer – the difference in values being caused by geometry, atmospheric absorption and scattering.

Just as luminosity has acquired the absolute magnitude scale to make it easier to compare, apparent brightness may also be described by the magnitude scale. This uses the apparent brightness of the star Vega in the constellation Lyra to define an apparent magnitude of 0. Prior to this Polaris was used to define an apparent magnitude of 2, but was subsequently found to be a variable star (i.e. its luminosity is not constant, varying between magnitudes 2.1 and 2.2 every 4 days) and so the definition was dropped in favour of Vega. Just as in the case of absolute

magnitude a difference of 5 in apparent magnitude corresponds to a 100-fold difference in apparent brightness.

From this it will be clear that when a star is described as having a particular magnitude it is important to clarify whether this is an **apparent** or **absolute** magnitude. For example, Sirius has an absolute magnitude of -3.6 but an apparent magnitude of -1.46 . As a guide, the faintest stars detectable by naked eye on a clear night are of around apparent magnitude 5, although the threshold will vary from person to person. The Moon has an apparent magnitude of -10.6 and the Sun -26.72 .

Detecting astronomical objects (mainly for interest)

When recording images of distant stellar objects in order to measure their apparent brightness, people used to use film cameras (indeed some still do). These worked by photons of light reacting with chemically treated rolls of plastic film. Nowadays, digital cameras produce images using a charge-coupled device (CCD) chip positioned directly behind the camera lens (where the film used to be located in a film camera).

A CCD is a doped semiconductor chip based on silicon, containing millions of light-sensitive squares or pixels. A single pixel in a CCD is approximately $10\text{ }\mu\text{m}$ in diameter. When a photon of light falls within the area defined by one of the pixels it is converted into one (or more) electrons and the number of electrons collected will be directly proportional to the intensity of the radiation at each pixel. The CCD measures how much light is arriving at each pixel and converts this information into a number, stored in the memory inside the camera. Each number describes, in terms of brightness and colour, one pixel in the image.

Not all of the photons falling on one pixel will be converted into an electrical impulse. The proportion of photons detected is known as the quantum efficiency (QE) and is wavelength dependent. For example, the QE of the human eye is approximately 20%, photographic film has a QE of around 10%, and the best CCDs can achieve a QE of over 90% at some wavelengths.

Some CCDs detect over a very short waveband to produce several black and white images that are then coloured during image processing. This is how the Hubble Telescope and the Faulkes' Telescopes produce their images. Certainly some of the Hubble images give details that would never be observed by the human eye.

CCDs have several advantages over film: a high quantum efficiency, a broad range of detectable wavelengths and the ability to view bright as well as dim images. CCDs have an advantage over the eye in that their response is generally linear, i.e. if the QE is 100% efficient 100 photons would generate 100 electrons. The human eye does not observe linearly over a large range of intensities and has a logarithmic response. This makes the eye poor at astrophotometry (the determination of the brightness of stars). The human eye and brain also compensate for relative colour and brightness which has a huge impact on the image in front of us and the image that we perceive.

Another of the major advantages of using CCDs in imaging stellar objects is the ability to use software to 'stack' images. Several images of the same object can be taken and easily superimposed to produce depth and detail that would not be observed from a single image.

Requirements when using CCDs include ensuring the detector is properly calibrated and that the absorption of light by the atmosphere is taken into account if CCDs are being used to measure

apparent brightness. One disadvantage relative to film is that the sensors tend to be physically small and hence can only image small areas of the sky at one time, although this limitation can be overcome by using mosaics of sensors.

Stellar classification

Initial attempts to classify stars in the mid-19th century were based on colour and spectral absorption lines. The most commonly used scheme was that devised by Angelo Secchi. It is summarised in the table below.

Secchi stellar classification

Classification type	Colour	Absorption lines	Examples
I	Blue-white.	Strong hydrogen	Sirius, Vega
II	Yellow, orange.	Dense metallic species.	Sun, Arcturus, Capella
III	Orange-red	Titanium oxide	Betelgeuse, Antares
IV	Deep red	Carbon	
V		Strong emission lines	

In the late 19th century a different approach was based on looking at hydrogen absorption line spectra and classifying them alphabetically from A to Q with Class A having the strongest absorption lines. However, it was eventually realised that such a system was slightly flawed and a clearer sequence studying absorption lines of many chemical species was established. This retained the lettering classification but moved some classes around and removed some altogether. Eventually the classification became the one used today and is a result of correcting previous errors or omissions in the classification.

Modern stellar classification

Spectral type	Temperature (K)	Examples	Colour
O	> 30,000	Orion's Belt stars	Blue
B	30,000–10,000	Rigel	Blue-white
A	10,000–7,500	Sirius	White
F	7,500–6000	Polaris	Yellow-white
G	6000–5000	Sun	Yellow
K	5000–3500	Arcturus	Orange
M	<3500	Proxima Centauri, Betelgeuse	Red

The order of the spectral classes has been remembered by generations of astronomy students using the mnemonic '**O**h, **B**e **A** Fine **G**irl/**G**uy, **K**iss **M**e'.

This work was carried out at the Harvard Observatory by (mostly female) 'computers'. They had educated themselves in physics and astronomy, but the social pressures of the time meant they could not become undergraduates or have a formal position at the Observatory so this was the only type of work they could do within this field. Eventually, one of their number, Annie Jump Cannon, who led the work described above, was recognised with an honorary degree from Oxford University almost 20 years after the scheme was adopted.

The spectral types O to M cover quite large temperature ranges, so within each type a numbering system is used, e.g. A0, A1, A2 etc. up to A9, with each number representing one

tenth of the difference between class A and class F. The smaller the number the hotter the star will be. The Sun is a star of spectral type G2 with a surface temperature of 5800 K.

Solar activity and sunspots

Although stellar activity is discussed here in the context of our Sun, similar processes are likely to be at work in all such stars. We saw in the first section that the interior of the Sun has distinct zones or regions, and that the zone closest to the surface is the convective zone, where heat is transferred towards the surface by convection of the plasma. Photographs of the Sun clearly show these areas of convection, with the bright areas being regions where hotter plasma is welling up from below and dark regions those where cooler plasma is sinking.

From time to time, larger dark areas appear, ranging in size from 1000 km to 100,000 km in diameter (with the Earth for comparison being approximately 13,000 km in diameter). These sunspots, which usually occur in pairs, are thought to arise when strong magnetic field lines emerge from the surface of the Sun, curve out into space and then go back into the Sun. Sometimes loops of plasma can become established in this field and give rise to a *solar prominence*. These strong magnetic fields reduce the convective flow of plasma from below and therefore reduce the surface temperature within the sunspot to around 4000 K. Sunspots may last for anything from a few hours to a few months depending on the stability of the magnetic fields.

The regions surrounding sunspots are associated with several aspects of solar ‘weather’. The first of these is the solar flare. The mechanism is believed to be associated with sudden breakdowns of twisted magnetic fields that occur in the vicinity of sunspots. Solar flares last from a few minutes to a few hours and release large quantities of radiation, predominantly X-rays (although radiation is emitted across the entire spectrum), and a stream of fast-moving charged particles.

For reasons which are not yet fully understood, the total number of sunspots present follow a cyclical pattern known as the sunspot cycle. Since sunspots have been observed for several centuries it is possible to trace this cycle back for an extended period and this confirms that sunspot numbers peak (at solar maximum) and trough (at solar minimum) approximately every 11 years but that the total number of sunspots at each peak varies considerably.

The historic data also show that total sunspot number is quite variable from cycle to cycle and that this cyclical behaviour was suspended between about 1645 and 1715. This period is known as the Maunder Minimum and coincided with the onset of the Little Ice Age, although the extent to which the latter was brought on by the former is uncertain (correlation is not necessarily causation). In fact, although the changes in X-ray and ultraviolet emissions are quite pronounced throughout the sunspot cycle, the total energy output of the Sun changes very little. Research into the effects of solar ‘weather’ on Earth’s climate and weather is still at a relatively early stage.

Interestingly, the Sun’s magnetic field reverses completely at each solar maximum so that it takes 22 years for the magnetic field to go from one state to another then back to the original state. This is known as the solar cycle.

Earlier we saw that the temperatures of the Sun’s chromosphere and corona are much higher than that of the photosphere, although the surface temperature that we observe from Earth is that of the photosphere. This arises because the densities of the chromosphere and corona are very small. The reasons for the occurrence of these high temperatures are not fully understood,

but it is thought that they are produced by energy transferred via the movement of loops of magnetic field lines which are, in turn, caused by strong currents within the convection zone.

The high temperatures of the chromosphere and corona cause them to emit X-ray radiation. X-ray images of the Sun reveal that the hottest areas of the corona are related to areas with sunspots and that the polar regions of the Sun have relatively cool areas of corona. These areas are known as coronal holes and are characterised by having magnetic field lines that do not loop back into the photosphere, but trail out into space, allowing streams of charged particles to escape along them. This stream of particles travels outwards from the Sun at around 500 km s^{-1} and is known as the solar wind.

Coronal mass ejections (CMEs) are a more violent form of solar flare or storm. Massive quantities (far in excess of the quantity associated with the solar wind) of charged particles are ejected from the Sun. These pulses of particles have very strong magnetic fields, which can, if they happen to be ejected in the direction of Earth, cause geomagnetic storms.

During a geomagnetic storm, the effects on Earth are potentially serious. Bright auroras (northern and southern lights) may be observed but the main problem is that the strong magnetic fields cause currents to be induced in conductors. Thus electrical and electronic devices may malfunction or cease to work entirely, affecting communications and power distribution. If a CME is observed to be heading towards Earth, then we have some warning of the event as it takes around 3 days for the CME to reach us. CMEs are potentially very dangerous to those aboard the International Space Station.

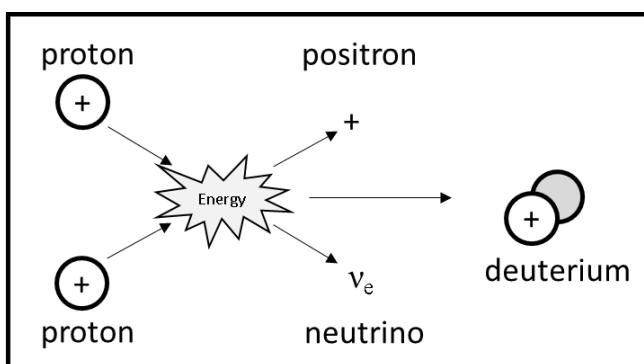
In recent years there have been several CME warnings, the most serious outcomes being:

2003 a series of CMEs close together destroyed a satellite, disrupted others and caused problems with aircraft navigation systems

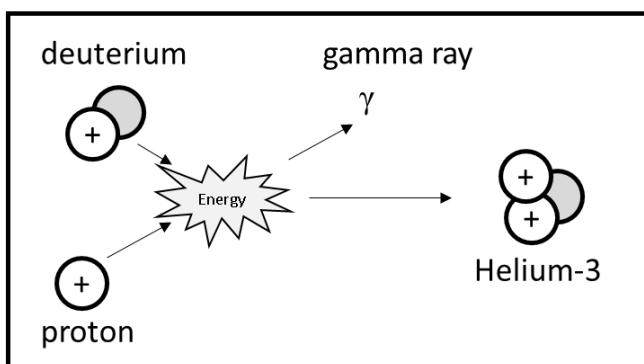
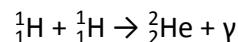
1989 parts of the Canadian power distribution grid failed for 9 hours and communications were lost with over 2000 military satellites. The total cost of this event was estimated to be around \$100 million, emphasising how damaging a CME can be. Our almost total reliance on electronic and electrical devices nowadays mean that any threat of a CME hitting Earth must be taken seriously. A sufficiently large event could induce currents large enough to melt the cores of large numbers of national power grid transformers, causing serious disruption and potentially taking a long time to repair.

Fusion in the Sun

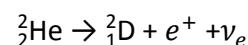
The Proton - Proton Chain - PPI

Step 1

This is actually a 2 stage process in which two proton fuse together to produce a diproton ${}^2_1\text{He}$.



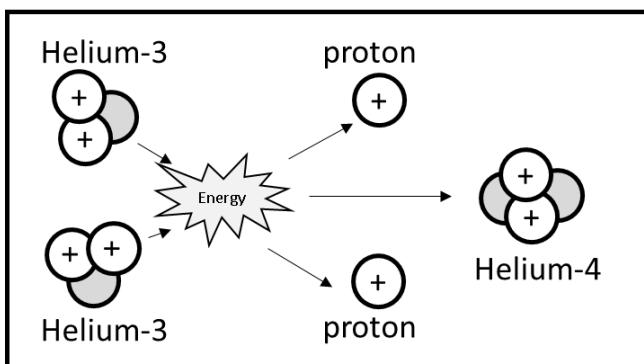
If a rare beta plus decay follows the result is a positron, a neutrino and deuterium ${}^2_1\text{D}$. It is much more likely that the diproton decays back into its original state by proton emission.



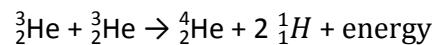
The positron will annihilate should it meet an electron, releasing more energy.

Step 2

The deuterium now fuses with another proton to give Helium-3.

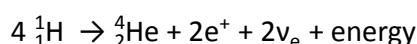
Step 3

The combination of two Helium 3 nuclei results in the final step producing Helium-4.



Step 3 requires 2 instances of steps 1 & 2 having occurred for there to be 2 Helium 3 nuclei available.

This whole sequence is known as the *proton–proton chain reaction* and may be summarised as:

For interest

This chain is dominant in stars with temperatures between 10 and 14 MK.

PPII – Helium 3 + Helium 4 make Beryllium 7, which beta decays into Lithium 7. Lithium 7 + proton make 2 Helium 4. This is dominant at 14 to 23 MK.

P_{III} – Helium 3 + Helium 4 make Beryllium 7, which decays into Boron 8 through proton emission. This is then followed by decays into Beryllium 8, which decays into 2 Helium 4 nuclei. This chain is dominant above 23 MK.

Two protons, being positively charged, will repel each other as they approach. Forcing two protons close enough together so that they will fuse can only occur at extremely high temperature and pressure, great enough to overcome the repulsive electrostatic forces between the nuclei. Conditions in the core of the Sun meet the requirements of nuclear fusion with a temperature of around 15 million kelvin and a pressure of 200 billion atmospheres.

One of the methods by which the proton–proton chain reaction could be confirmed to be the source of the Sun’s energy was to predict the number of neutrinos that could be expected to arrive at the Earth if the reaction was proceeding at a rate which was consistent with the rate of energy production observed. Neutrinos are very difficult to detect as they rarely interact with other particles (in fact as you are reading this many billions are passing unhindered through your body each second), but because so many are produced by the Sun it is possible, with a sufficiently large detector, to measure a fraction of these interactions.

The first solar neutrino detector consisted of an underground reservoir containing 400,000 litres of dry-cleaning fluid. Chlorine nuclei within this fluid occasionally interacted with neutrinos to produce argon, so the quantity of argon accumulated within the reservoir allowed the number of neutrinos arriving at the Earth’s surface to be estimated.

Unfortunately, the quantity of neutrinos detected was only about one-third of the number that theory predicted, suggesting either that the proton–proton chain reaction was *not* the source of the Sun’s energy or that the experiment was for some reason failing to detect two-thirds of the neutrinos. This problem remained unsolved for over 30 years and became known as the solar neutrino problem.

It was known that the proton–proton chain reaction produced only electron neutrinos but the existence of two other types, the muon neutrino and the tau neutrino was also known. (Muons and taus are more massive cousins of the electron.) Until recently neutrino detectors were only able to detect electron neutrinos, but with the introduction of a detector that could detect all three types it rapidly became apparent that the number arriving on Earth did, indeed, match the expected output from the proton–proton chain reaction, but that all three types were present. It is now believed that some of the electron neutrinos are converted to the other two types during their passage from the core to the surface of the Sun. Thus, after three decades of observations, it was possible to be confident that the models which predicted the source of the Sun’s energy were correct.

Modern observations tell us that the Sun converts around 4 million tonnes of its mass to energy each second. Even at this rate, its huge mass means its fuel will last for many billions of years to come.

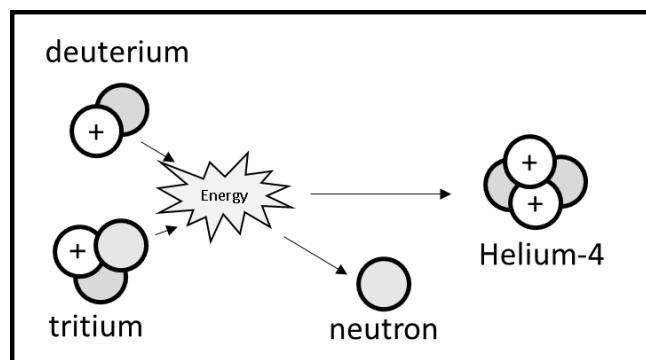
Nuclear fusion is the focus of much applied research on Earth as a way of producing power, allowing our reliance on fossil fuels to be reduced. Progress has been slow because the engineering required to contain plasma at high temperatures and pressures is very challenging.

The nuclei in the Sun actually do not have quite enough energy to completely overcome the repulsive Coulomb forces and rely partly on a process called quantum tunnelling for fusion to

occur. For this reason manmade fusion reactors must produce temperatures of the order of 10–100 times higher than those found in the Sun.

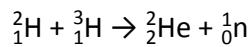
The ITER project based in the south of France hopes by 2019 to produce the first reactor capable of giving out more power (500 MW) than it takes to contain the plasma (50 MW). The diagrams on the next page show one method of creating a fusion reaction on Earth, notice how different this is to that of our Sun.

Possible Earth reactor based fusion reaction

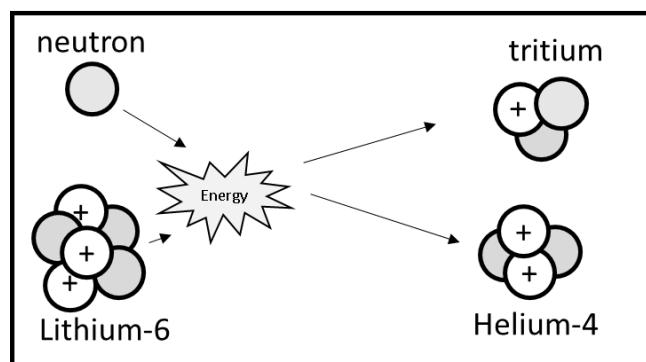


Step 1

In this reaction deuterium and tritium, both forms of heavy hydrogen, fuse to form helium and a neutron.

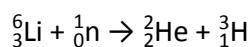


Energy output: **17.6 MeV**



Step 0

To generate the tritium required for the above reaction lithium-6 must be bombarded with neutrons.



Energy output: **4.78 MeV**

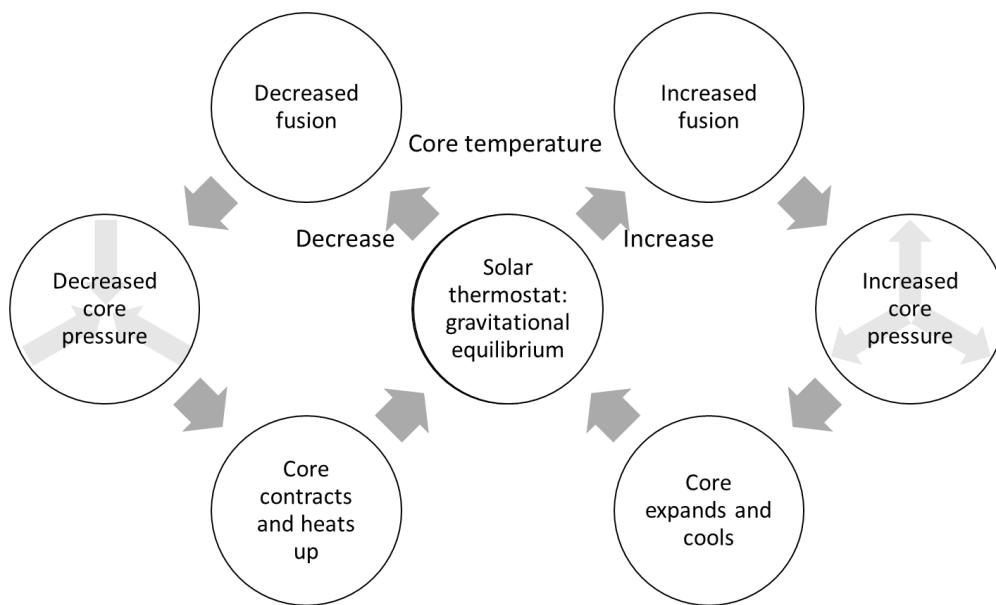
The significant energy output here is from the deuterium-tritium reaction. Tritium can also be produced through similar bombardment of Lithium 7, however this process actually results in an energy “loss” of 2.47MeV.

Stellar Equilibrium

A star remains stable in terms of its size and processes by maintaining an equilibrium between the inward gravitational attraction of its mass and the outward pressure provided by its internal fusion.

Stars maintain a 2 part cycle to balance these two opposing forces.

A cooling within the core will cause the rate of fusion to decrease. This reduces the core pressure and the star contracts. The contraction and increased proximity of the atoms, causes the pressure and core temperature to increase again and the rate of fusion increases as well. This results in an increase in temperature and pressure causing the star to expand. The expansion causes cooling again and so the cycle repeats. In the diagram below, think of this as a figure of 8 path.



Stellar Evolution

We have now explored the structure of stars and explained how they produce energy. In this section we are going to look back to the beginning of a star's life and explore the processes it goes through as it ages.

Star formation

Stars are formed in what are often called interstellar nurseries. These are volumes of space containing (relatively) high densities of gas, although it should be emphasised that the densities are of the order of 300 million molecules per cubic metre. This may sound like a large number but must be considered against the density of air at sea level on Earth, which is of the order of 25×10^{24} molecules per cubic metre.

Viewed from Earth these molecular clouds appear as areas in the sky which block the light from stars beyond them and are called nebulae (although it should be noted that not all objects classed as nebulae are star-forming). One of the best known groups of nebulae are found in the constellation of Orion lying below the belt and forming the sword. They are easily observed with binoculars or a small telescope. Although appearing greenish to the naked eye, the nebulae are in fact a reddish-orange colour. This apparent contradiction arises because the human eye is more sensitive to light in the middle of the visible spectrum. Closer to Orion's belt is the famous Horsehead Nebula, in which a cloud of dark dust lies in front of a faint nebula.

The molecular clouds are formed mostly from hydrogen, with lesser quantities of other molecules such as carbon monoxide, ammonia, water and ethanol. In addition, small quantities of dust are commonly found, consisting of carbon, iron, oxygen and silicon. It should be noted that the first stars formed after the Big Bang arose from clouds containing only hydrogen and helium, and the other elements (up to iron on the periodic table) arose as a result of nuclear fusion in these first stars.

Areas of space away from molecular clouds and stars are called the interstellar medium and contain low densities of molecules and dust. In these areas gravitational attraction between molecules is relatively weak because the molecules are widely spaced apart and this attraction is not enough to overcome the thermal pressure that arises as gravity brings the molecules closer together. Star formation is therefore not possible in the interstellar medium. However, in molecular clouds the molecules are close enough together that, given suitable circumstances, gravitational attraction overcomes the thermal pressure and progressively the molecules compress into a core. The process steadily heats the core until it reaches a temperature where nuclear fusion can be sustained, at which point a star has been formed. The processes which govern energy production and transfer within the star are covered in the section on the properties of stars.

Stellar nucleosynthesis

The process of stellar nucleosynthesis was first postulated by the British astronomer, Fred Hoyle, in the late 1940s when he was trying to understand how elements heavier than hydrogen and helium could have arisen in the Universe. The production of these heavier elements by nuclear fusion in a star of mass $25M_{\text{Sun}}$ is summarised below, which shows that progressively higher temperatures and densities are required to produce the heavier elements.

Summary of nucleosynthesis occurring in a star of mass $25M_{\text{Sun}}$.

Stage	Temperature (K)	Density (g cm ⁻³)	Duration
Hydrogen → helium	4×10^7	5	10^7 years
Helium → carbon	2×10^8	7×10^2	10^6 years
Carbon → neon + magnesium	6×10^8	2×10^5	600 years
Neon → oxygen + magnesium	1.2×10^9	5×10^5	1 year
Oxygen → sulphur + silicon	1.5×10^9	1×10^7	6 months
Silicon → iron	2.7×10^9	3×10^7	1 day

Hoyle's work was interesting not just for what he discovered, but also for the way in which he achieved it. All the reactions above were explainable with the exception of the helium to carbon step, where the pathway required an intermediate reaction product that had never been observed and was not believed to exist. Hoyle adopted what is known as the anthropic principle, reasoning that since he was partly made of carbon-12 and lived in the universe he was investigating, then such a pathway must exist otherwise he could not exist. He predicted the energy level of this intermediate product and then pressed Willy Fowler (an American astrophysicist) to search for this specific product. It was discovered within a few days and thus confirmed the process by which nucleosynthesis occurs.

Beyond iron the energy required for the nucleosynthesis of heavier elements is greater than the energy produced. This would be an inefficient process to be carried out in a star and would cause a decrease in pressure, leading to collapse. To achieve the energy required for the nucleosynthesis of heavier elements a star must explode, this is known as a supernova. During this explosion the massive amounts of energy released can result in the production of these heavier elements. This only occurs in high mass stars.

Thus, all the elements present in our Universe have been created by nucleosynthesis in many generations of stars, either due to the reactions described above or by the formation of heavier elements during a supernova explosion. Everything around us is made of stellar material, including ourselves. As Simon Singh states:

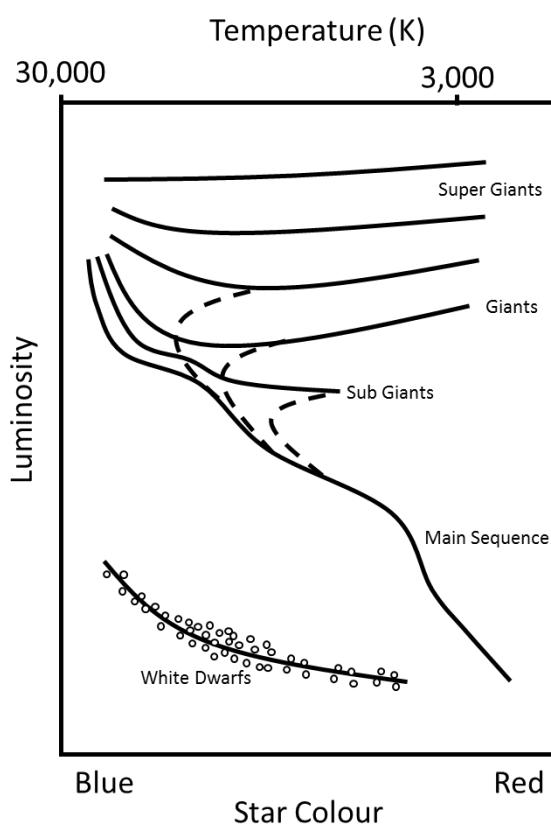
"Romantics might like to think of themselves as being composed of stardust. Cynics might prefer to think of themselves as nuclear waste"

Hertzsprung–Russell diagrams

The characteristics, lifetime and ultimate fate of a particular star depend solely on its mass, which in turn is constrained by the quantity of material available during the star's formation.

We have seen in an earlier section that stellar masses are confined to a range $0.08M_{\text{Sun}}$ to $150M_{\text{Sun}}$. This is because beyond this range gravitational equilibrium is not possible. We will shortly discuss the evolution of low-mass and high-mass stars. However, to understand the processes of evolution it is necessary to understand how mass influences the basic characteristics of a star.

In the early 20th century, a Danish astronomer, Ejnar Hertzsprung, was exploring how a star's spectrum might be related to its apparent magnitude. The graphs he produced of apparent magnitude versus spectral class show distinct patterns and clusters of stars. Working independently, Henry Norris Russell in the USA did a similar thing using absolute magnitude and found the same features as Hertzsprung.



Plots of luminosity versus temperature are now known as Hertzsprung–Russell diagrams (H–R diagrams) and understanding one is vital to explaining the lifetimes of various types of star. Note that both axes are logarithmic and that the x-axis is reversed so that the hottest stars are at the left-hand side of the diagram. Stars are plotted as dots. Here types of star are shown by the lines.

As we saw earlier, the spectral type sequence for classifying stars is O, B, A, F, G, K, M.

There are four main groups of stars to highlight, each group containing stars at different stages of their life cycle: main sequence, giants, supergiants and white dwarfs.

Main sequence stars

Once a star has formed and established hydrogen fusion it will occupy a position somewhere along the main sequence, a band of stars lying from the upper left to the lower right of the H–R diagram. The position within the main sequence is determined by the mass of the star, with the most massive at the upper left and the least massive at the lower right.

The mass of the star also directly influences its surface temperature. High-mass stars have a large gravitational pressure and correspondingly high core temperatures and pressures. This leads to high-mass stars having high surface temperatures in the region of 25,000 K, whilst low mass stars may have surface temperatures as low as 2,000 K. As described earlier, this leads to cool, low-mass stars having a red colouration whilst massive stars appear blue-white.

A further consequence of the very high temperatures and pressures in massive stars is that, despite having much larger reserves of hydrogen than smaller stars, they consume their hydrogen supply at such a rate that they are (relatively speaking) very short lived. Those at the upper left of the diagram have lifetimes of the order of several million years, whilst the smallest, coolest, stars have lifetimes in excess of 300 billion years, or more than 20 times longer than the universe has so far existed. Our own Sun was formed around 4.5 billion years ago and will have a main sequence lifetime of 10 billion years, so it is roughly halfway through its main sequence phase.

Low-mass stars

Subgiants and giants

Towards the end of the life of a low-mass star (such as our Sun), it begins to run out of the hydrogen fuelling the fusion reaction. This causes the stellar pressure to decrease, which in turn causes the gravitational pressure to compress the core, leading to a higher temperature. At this higher temperature helium nuclei begin to fuse to form carbon. The star's luminosity increases but because the star also expands in volume the surface temperature is lower than would be expected if it were still on the main sequence. The star expands and the surface temperature cools, passing through the subgiant region and eventually becoming a red giant. An easily observed example of such a giant star (actually an orange giant) is Aldebaran, the bull's eye in the constellation of Taurus. When this happens to our Sun it will expand past the current radius of the Earth's orbit. The radius of Earth's orbit will by this stage have increased due to loss of solar mass, although tidal interactions with the Sun will tend to counteract this effect, leaving the likelihood that Earth will be engulfed by the Sun somewhat uncertain. Long before this point, however, the oceans will have evaporated due to the Sun's increasing luminosity, making continued human life on Earth unlikely.

Depending on the mass of the original star, once the helium is exhausted a further contraction raises the core temperature again, carbon atoms begin to fuse to produce neon and magnesium, and the star expands further.

This process of core compression followed by star expansion occurs for each of the fusion stages in the nucleosynthesis table, but low to medium mass stars are not able to produce elements heavier than carbon and oxygen.

White dwarfs

As a red giant continues to expand it eventually reaches a size where the outer layers of the star can no longer be retained by gravitational attraction and the star enters the final phase of its lifetime. In terms of the H–R diagram the star moves left and downwards out of the giant region, and the outer layers steadily disperse as a planetary nebula, leaving behind the stellar core. Since fusion no longer occurs in this dead core it steadily cools, moving down the H–R diagram until it becomes an inert white dwarf consisting mostly of carbon and oxygen. The term ‘planetary nebula’ is used for historical reasons and does not imply planet formation.

The death of a Sun-like star, from the point at which the hydrogen runs out until it becomes a white dwarf, is of the order of one billion years, compared with a 10 billion year main-sequence lifetime.

High-mass starsSupergiants

The evolution of a high-mass star is similar to that of a low-mass star in the early stages, when hydrogen is fused to form helium. However, as the hydrogen runs out and helium fusion begins, the process of expansion is much greater than for a low-mass star and the star becomes a supergiant. In addition, the high mass means that fusion proceeds through all the stages listed in the nucleosynthesis table.

Neutron stars, black holes and supernovae

No fusion pathways possible within a star are capable of producing elements more massive than iron (although some stars produce heavier elements via a neutron capture process called the S process) so ultimately the star accumulates iron in its core. This process proceeds up to the point where the mass and density of the core are so large that gravitational forces cause the core to rapidly collapse in on itself. The immense pressure overcomes the electron degeneracy pressure, protons and electrons combine to form neutrons and the core collapses to diameter of perhaps a few kilometres, becoming a **neutron star**. Neutron stars are incredibly dense, the density varying with depth from the surface, but on average being such that a single teaspoon of the material would have a mass comparable with that of the entire human population. Surface gravity is of the order of 10^{11} times greater than that on Earth, meaning that an object released from a height of 1 m would impact the surface at over 1,400,000 m s⁻¹. Due to conservation of angular momentum as the star shrinks, neutron stars spin extremely quickly, having rotational periods ranging between around a millisecond and several seconds.

In some cases, where the mass is large enough, the gravitational force overcomes the degeneracy pressure of the neutrons, after which nothing can stop the collapse. It continues unabated until an object of zero volume and infinite density is created. The gravitational field (or subsequent curvature of space-time) of such an entity is so strong that even light cannot escape. These objects are known as black holes.

The radius within which a given mass must collapse to form a black hole is given by the Schwarzschild radius.

$$r = \frac{2GM}{c^2}$$

G is the universal gravitational constant

M is the mass of the object

c is the speed of light in a vacuum.

For a non-rotating black hole the surface at the Schwarzschild radius forms an *event horizon*, a region from beyond which nothing, including light, can escape to influence the outside universe. Interestingly in the case of a sufficiently massive black hole it would be possible to fall within the event horizon without suffering any ill-effects or necessarily being aware of one's impending doom.

This collapse itself is a sudden and very rapid event, lasting a fraction of a second and releasing enormous amounts of energy. This energy release blows away the outer layers of the star into space and releases a massive burst of radiation known as a supernova, which for a brief moment can outshine an entire galaxy. From Earth, relatively nearby supernovae are seen as a sudden brightening of a star, which lasts from several days to months before gradually fading away. However, the debris from such explosions can remain visible for much longer.

An example of this is found in the Crab Nebula. This nebula arose from a supernova observed in China in 1054 AD and modern observations confirm that there is a neutron star at the centre of the nebula surrounded by clouds of stellar material expanding at thousands of kilometres per second.

The closest red supergiant to Earth at 310 light years is Betelgeuse in the constellation of Orion. It is not possible to know exactly where it is in its life cycle but if it happens to have gone supernova during the 18th century you may yet see the visible evidence arriving on Earth.

The death of a high-mass star, from the point at which the hydrogen runs out until it goes supernova, is of the order of one million years, compared with a five billion year main-sequence lifetime. This is roughly three orders of magnitude shorter than the lifetime of a Sun-like star.

Material ejected during a supernova includes all the elements formed by the fusion processes so that when, ultimately, this material condenses to form a new star it will already have some of the heavier elements present within it. However, the quantity of energy available in a supernova is sufficient to bring about the formation of elements heavier than iron and, together with the S process, supernova nucleosynthesis is responsible for the production of elements up to and including uranium, including many of the elements necessary for human life. The atoms which your body is currently borrowing have existed for billions of years and were once part of a supernova explosion.