Grove Academy

Advanced Higher Physics

Rotational Motion and Astrophysics Summary Notes

Rotational Motion

Kinematic Relationships

Throughout this course calculus techniques will be used. These techniques are very powerful and a knowledge of integration and differentiation will allow a deeper understanding of the nature of physical phenomena.

Kinematics is the study of the motion of points, making no reference to what causes the motion. Displacement, velocity and acceleration will all be addressed here.

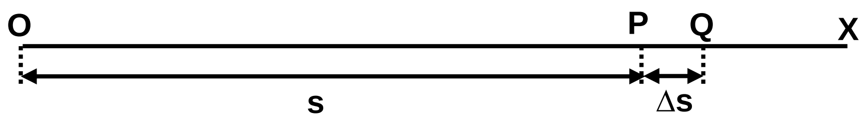
The displacement, **s**, of a particle is the distance **and** direction from the origin to the particle.

Displacement

The displacement of the particle is a function of time:

$$s = f(t)$$

Consider a particle moving along OX.



s is covered in time t.
Δs is covered in time Δt.
ds is covered in time dt (i.e. Δt → 0)

At time t + Δt particle passes Q.

Velocity

Velocity is defined as the rate of change of displacement.

Average velocity: $\bar{v} = \frac{\Delta s}{\Delta t}$

However the **instantaneous** velocity is different, this is defined as:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}$$

$$\boxed{v = \frac{ds}{dt}}$$

Acceleration

Note:
A change in velocity may result from a change in direction (e.g. uniform motion in a circle - see later).

Acceleration is defined as the rate of change of velocity.

Average acceleration: $\bar{a} = \frac{\Delta \mathrm{v}}{\Delta t}$

However the **instantaneous** acceleration is different, this is defined as:

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}$$

$$\boxed{a = \frac{dv}{dt} = \frac{d^2 s}{dt^2}}$$

Mathematical Derivation of Equations of Motion for Uniform Acceleration

Derivation of $v = u + at$ ***(Equation 1)***

$$a = \frac{dv}{dt}$$

$$\int_0^t a\,dt = \int_u^v \frac{dv}{dt}\,dt$$

$$a\int_0^t dt = \int_u^v dv$$

$$a[t]_0^t = [v]_u^v$$

$$at - 0 = v - u$$

$$\boxed{v = u + at}$$

Integrate with respect to time.

Derivation of $s = ut + \frac{1}{2}at^2$ ***(Equation 2)***

$$v = \frac{ds}{dt}$$

$$ds = v.dt$$

$$\int_0^s ds = \int_0^t v\,dt$$

$$\int_0^s ds = \int_0^t (u + at)\,dt$$

$$[s]_0^s = [ut + \frac{at^2}{2}]_0^t$$

$$s - 0 = \left(ut + \frac{1}{2}at^2\right) - 0$$

$$\boxed{s = ut + \frac{1}{2}at^2}$$

Integrate with respect to time.

Remember that:
$$v = u + at$$

Derivation of $v^2 = u^2 + 2as$ ***(Equation 3)***

$$v = u + at$$

$$v^2 = (u + at)^2$$

$$v^2 = u^2 + 2uat + a^2t^2$$

$$v^2 = u^2 + 2a\left(ut + \frac{1}{2}at^2\right)$$

$$\boxed{v^2 = u^2 + 2as}$$

Start with Equation 1.

Square both sides.

Expand the brackets out.

Remember that:
$$s = ut + \frac{1}{2}at^2$$

Variable Acceleration

If acceleration depends on time in a simple way, calculus can be used to solve the motion.
(This is the case when the expression for displacement includes time raised to a power greater than 2.)

Example:

The displacement **s** *of an object in metres is given by the following equation:*

$$\boldsymbol{s = 15 + 20t^2 - 30t^4}$$

(a) *Find an expression for the velocity of the object as a function of time.*

$$v = \frac{ds}{dt}$$

$$v = 40t - 120t^3$$

(b) *Find an expression for the acceleration as a function of time.*

$$a = \frac{dv}{dt}$$

$$a = 40 - 360t$$

(c) *What was the initial displacement of the object?*

$$Initial\ displacement\ is\ the\ value\ of\ s\ when\ t = 0$$

$$s = 15 + 20(0)^3 - 30(0)^4$$

$$s = 15\ m$$

(d) *What was the initial velocity of the object?*

$$Initial\ velocity\ is\ the\ value\ of\ v\ when\ t = 0$$

$$v = 40(0) - 120(0)^3$$

$$v = 0\ ms^{-1}$$

(e) *What was the initial acceleration of the object?*

$$Initial\ acceleration\ is\ the\ value\ of\ a\ when\ t = 0$$

$$a = 40 - 360(0)$$

$$a = 40\ ms^{-2}$$

(f) *At what times is the velocity of the object zero?*

$$0 = 40t - 120t^3$$

$$0 = 40t(1 - 3t^2)$$

$$40t = 0 \qquad \therefore t = 0\ s$$

$$1 - 3t^2 = 0 \qquad 3t^2 = 1 \qquad \therefore\ t = 0.58\ s$$

Graphs of Motion

The slope or gradient of these graphs provides useful information. Also the area under the graph can have a physical significance.

Displacement - time graphs

$$v = \frac{ds}{dt}$$

$$gradient\ of\ the\ slope = instantaneous\ velocity$$
$$area\ under\ the\ slope\ has\ no\ meaning$$

Velocity - time graphs

$$a = \frac{dv}{dt}$$

$$gradient\ of\ the\ slope = instantaneous\ acceleration$$
$$area\ under\ the\ slope = displacement$$

Calculations Involving Uniform Accelerations

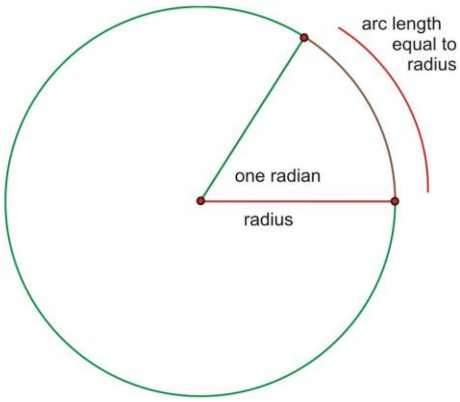
Examples of **uniform** acceleration are:

- vertical motion of a projectile near the Earth's surface, where the acceleration is g = 9.8 ms^{-2} vertically downwards
- rectilinear (i.e. straight line) motion e.g. vehicle accelerating along a road

Angular Motion

The Radian

You have probably come across the measurement of radians in Higher Maths.
One radian is the measurement given to the angle which separates two radii with an arc of length of one radius between them.



As the circumference of a circle is found using:

$$C = 2\pi r$$

$$\frac{C}{r} = 2\pi$$

This means that there are 2π radians in a full circle (360°).

To convert degrees to radians:

$$angle\ in\ radians = \frac{angle\ in\ degrees \times \pi}{180}$$

To convert radians to degrees:

$$angle\ in\ degrees = \frac{angle\ in\ radians \times 180}{\pi}$$

Angular Displacement, θ

The angular displacement of a rotating body is measured in radians.
It is found using the following relationship:

$$\boxed{s = r\theta}$$

where:

s = linear displacement in metres, m
r = radius of rotation in metres, m
θ = angular displacement in radians, rad **(Note: rad not rads for radians!)**

Angular Velocity, ω

The angular velocity of a rotating body is defined as the rate of change of angular displacement.

$$\omega = \frac{d\theta}{dt}$$

where:

ω = angular velocity in radians per second, $rads^{-1}$
θ = angular displacement in radians, rad
t = time in seconds, s

Note:
Sometimes, angular velocities are given in revolutions per minute, rpm, and it is appropriate to convert them to radians:

$$\omega = \frac{d\theta}{dt} = \frac{number\ of\ rpm \times 2\pi}{60}$$

Angular acceleration, α

The angular acceleration of a rotating body is defined as the rate of change of angular velocity.

$$\alpha = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}$$

where:

α = angular acceleration in radians per second per second, $rads^{-2}$
ω = angular velocity in radians per second, $rads^{-1}$
θ = angular displacement in radians, rad
t = time in seconds, s

We assume for this course that α is **constant**.

The derivation of the equations for angular motion are very similar to those for linear motion seen earlier.

Angular Motion Relationships

$$\omega = \omega_0 + \alpha t$$

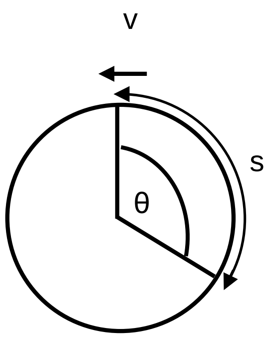
$$\theta = \omega_0 t + \frac{1}{2}\alpha t^2$$

$$\omega^2 = \omega_0^2 + 2\alpha\theta$$

You will note that these **angular** equations have exactly the same form as the **linear** equations. Remember that these equations only apply for **uniform** angular accelerations.

Uniform Motion in a Circle

Consider a particle moving with a uniform speed in a circular path as shown.



$$\omega = \frac{d\theta}{dt}$$

The rotational speed, v, is constant and the angular velocity, ω, is also constant. The period of the motion, T, is the time taken to make one full rotation (2π radians).

> **Note:**
> s is the arc of the circle swept out by the particle in time, t, and $s = r\theta$.

$$v = \frac{s}{t} \qquad and \qquad s = r\theta = 2\pi r \quad (for\ one\ full\ rotation)$$

$$\therefore \omega = \frac{2\pi}{T} \qquad and\ so \qquad v = \frac{2\pi r}{T}$$

$$\boxed{v = r\omega}$$

Angular acceleration and linear tangential acceleration

The angular acceleration is given by:

$$\alpha = \frac{d\omega}{dt}$$

The linear tangential acceleration is given by:

$$a_t = \frac{dv}{dt}$$

when the rotational speed v is *changing*.

Since $v = r\omega$, then at any instant, this means that $\frac{dv}{dt} = r\frac{d\omega}{dt}$ so:

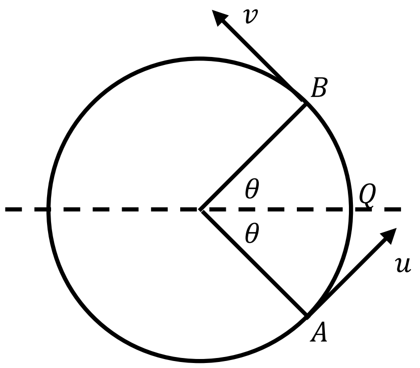
$$\boxed{a_t = r\alpha}$$

where the direction of a_t is at a tangent to the circular path of radius r.

Note: The unit for a_t is ms^{-2}.

Radial/Centripetal Acceleration

Consider a particle undergoing circular motion.



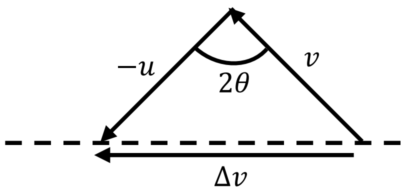
The particle travels from A to B in a time ***Δt*** and with a speed ***v***, thus $|u| = |v|$ and:

$\Delta v = v + (-u)$ which gives $\Delta v = v - u$.

Now,

$$\Delta t = \frac{arc\ AB}{v} = \frac{r2\theta}{v}$$

Average acceleration:

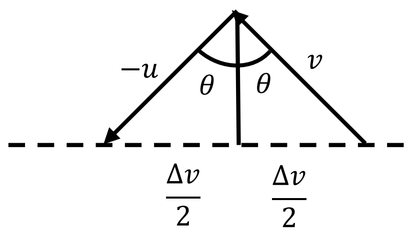


$$a = \frac{\Delta v}{\Delta t} = \frac{2vsin\theta}{\Delta t} = \frac{2vsin\theta}{\frac{r2\theta}{v}} = \frac{v^2 sin\theta}{r\theta}$$

As θ tends to zero, the average acceleration tends to the instantaneous acceleration at point Q.

$$a = \frac{v^2}{r}\lim_{\theta\to 0}\left[\frac{sin\theta}{\theta}\right] \quad but \quad \lim_{\theta\to 0}\left[\frac{sin\theta}{\theta}\right] = 1$$

since when θ is small and is measured in radians, sinθ = θ.



$$sin\theta = \frac{\Delta v}{2}$$

$$\boxed{a_r = \frac{v^2}{r} = r\omega^2}$$

since $v = r\omega$.

The **direction** of this acceleration is **always towards** the **centre** of rotation.

Radial/centripetal acceleration is either represented by a_r or $a_\perp$.

Note:

This is **not** a uniform acceleration. Radial acceleration continuously changes direction and its magnitude changes is the speed of rotation changes.
This motion is typical of many **central force** type motions e.g. planetary motion, electrons 'orbiting' nuclei and electrons injected at right angles to a uniform magnetic field which will be covered later in the course.
Thus any object performing circular motion at uniform speed must have a constant **centre-seeking** or **central** force responsible for the motion.

Central/Centripetal Force

Does a rotating body really have an inward acceleration (and hence an inward force)?

Argument

Most people have experienced the sensation of being in a car or a bus which is turning a corner at high speed. The feeling of being 'thrown to the outside of the curve' is very strong, especially if you slide along the seat. What happens here is that the friction between yourself and the seat is insufficient to provide the central/centripetal force needed to deviate you from the straight line path you were following before the turn. In fact, instead of being thrown outwards, you are, in reality, continuing in a straight line while the car moves inwards. Eventually you are moved from the straight line path by the inward (central) force provided by the door.

Magnitude of the Centripetal Force

$$F = ma \qquad but \qquad a = \frac{v^2}{r} = r\omega^2$$

Thus, the centripetal force can be calculated using:

$$F = \frac{mv^2}{r} = mr\omega^2$$

Examples

1. **A Car on a Flat Track**
   If the car goes too fast, the car 'breaks away' at a tangent. The force of friction is not enough to supply an adequate central force.

2. **A Car on a Banked Track**
   For tracks of similar surface properties, a car will be able to go faster on a banked track before going off at a tangent because there is a component of the normal reaction as well as a component of friction, F_r, supplying the central force.
   The central force is $R\ sin\theta\ +\ F_r\ cos\theta$ which reduces to $R\ sin\theta$ when the friction is zero. The analysis on the right hand side is for the friction F_r equal to **zero**.

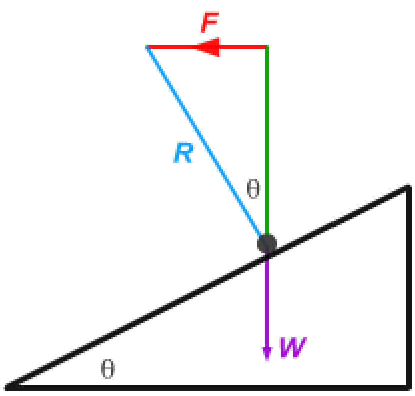
   R is the 'normal reaction' force of the track on the car.
   In the vertical direction there is no acceleration:

$$R\ cos\theta\ =\ mg$$

   In the radial direction there is a central acceleration:

$$R\ sin\theta\ = \frac{mv^2}{r}$$

$$\therefore\ tan\theta = \frac{v^2}{gr}$$

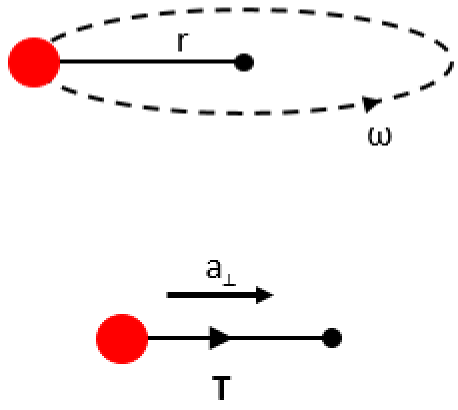


(This equation applies to all cases of 'banking' including aircraft turning in horizontal circles)

3. Objects Moving in a Horizontal Circle

The red ball is moving horizontally at a constant angular velocity, ω.
The only force acting horizontally on it is the tension in the string, T, so the tension provides the centripetal force which keeps the object in motion:

$$T = mr\omega^2$$



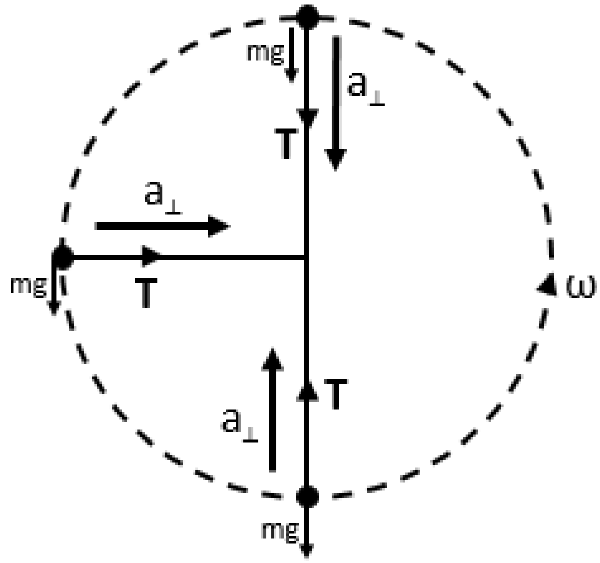
4. Objects Moving in a Vertical Circle

Weight, mg, always acts down.
When the object is at the top of the circle, both the weight and the tension are acting in the same direction.
The centripetal force is:

$$F = T_{top} + mg$$
$$mr\omega^2 = T_{top} + mg$$
$$T_{top} = mr\omega^2 - mg$$

When the object is at the bottom of the circle, the weight and the tension are acting in opposite directions.
The centripetal force is:

$$F = T_{bottom} - mg$$
$$mr\omega^2 = T_{bottom} - mg$$
$$T_{bottom} = mr\omega^2 + mg$$



5. Conical Pendulum

The string makes an angle ϕ to the vertical so:

$$sin\phi = \frac{r}{l}$$

$$Tcos\phi = mg$$

$$Tsin\phi = mr\omega^2$$

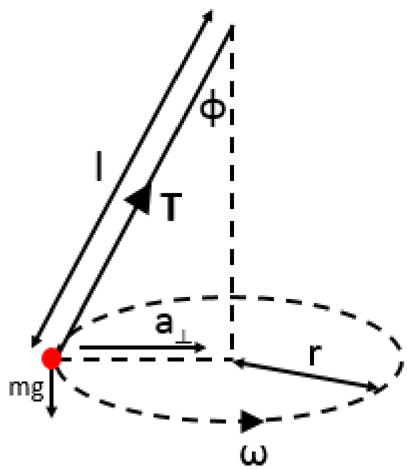
$$Tsin\phi = mlsin\phi\omega^2$$

$$\boxed{T = ml\omega^2}$$

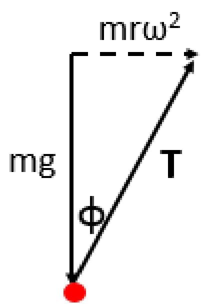
$$ml\omega^2 cos\phi = mg$$

$$l\omega^2 cos\phi = g$$

$$\boxed{cos\phi = \frac{g}{l\omega^2}}$$



The conical pendulum is moving in a horizontal circle



A free-body diagram shows the forces acting on the pendulum bob

Rotational Dynamics

Moment of a Force (Torque)

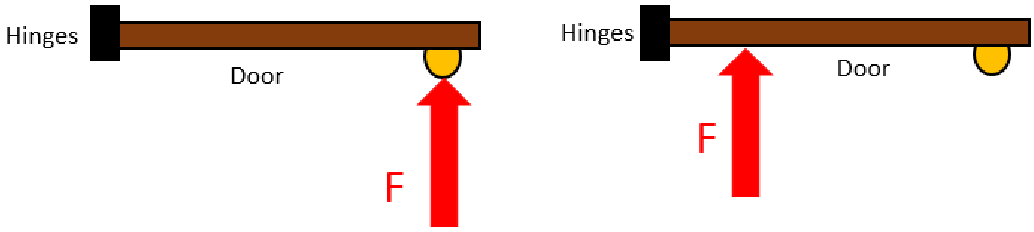
The **moment** of a force is the **turning effect** it can produce. It is also known as torque.

Examples of torques are:

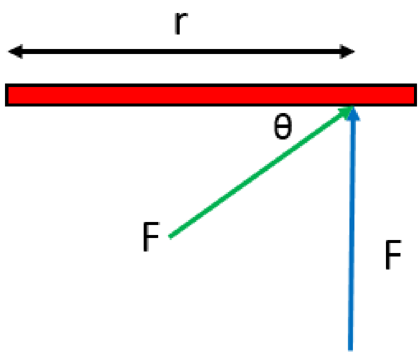
- Using a long handled screwdriver to 'lever off' the lid of a paint tin.

- Using a claw hammer to remove a nail from a block of wood or levering off a cap from a bottle.

- Opening a door



The same force can have different effects on the same rotating object depending on where it is applied. The same force has been applied to both doors.
The door on the left will open more quickly than the door on the right because the force has been applied at a larger radius to the axis of rotation.



If a force is applied perpendicularly to the radius of a rotation, the torque is calculated using:

$$T = Fr$$

where:

T = torque in newtons metres, Nm **(Not Nm^{-1})**
F = applied force in newtons, N
r = radius of rotation in metres, m

If a force is applied at an angle to the radius of a rotation, the torque is calculated using:

$$T = Fr\,sin\theta$$

Note the position of the angle θ in the diagram.

Torque is a **vector** quantity. The direction of the torque vector is at right angles to the plane containing both r and F and lies along the axis of rotation.

A force acting on the rim of an object will cause the object to rotate; e.g. applying a push or a pull force to a door to open and close.

An **unbalanced torque** will produce an **angular acceleration**.

Frictional torques oppose motion, in a similar way to frictional forces in linear motion.

Inertia

In linear dynamics an unbalanced force produces a linear acceleration. The magnitude of the linear acceleration produced by a given unbalanced force will depend on the mass of the object.
The word inertia can be loosely described as 'resistance to change in motion of an object'.
Objects with a large mass are difficult to start moving and once moving are difficult to stop.

Moment of Inertia

The moment of inertia I of an object can be described as its resistance to change in its angular motion.
The moment of inertia I for rotational motion is similar to the mass m for linear motion.

The moment of inertia I of an object depends on the mass **and** the distribution of the mass about the axis of rotation.

For a mass m at a distance r from the axis of rotation the moment of inertia of this mass is given by the mass m multiplied by r^2.

$$I = mr^2$$

where:

I = moment of inertia in kilogram metres squared, kgm^2
m = mass in kilograms, kg
r = distance between point mass and axis of rotation in metres, m

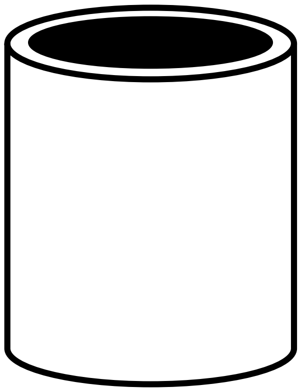
$I = mr^2$ is only used to calculate the moment of inertia of a point mass on a rotating body. This is because, for many other rotating objects, not all of the mass of the object is rotating at the same distance from the axis of rotation.

To calculate the total moment of inertia of a rotating body the following equations are used:

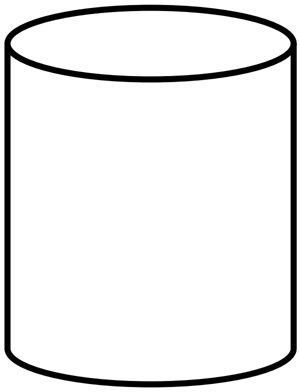
Scenario	**Equation**
Sphere about centre	$I = \frac{2}{5}mr^2$
Disc about centre	$I = \frac{1}{2}mr^2$
Rod about centre	$I = \frac{1}{12}ml^2$
Rod about end	$I = \frac{1}{3}ml^2$

where l is the length of the rod and all other quantities are as previously stated.

The following two cylinders are made of different metals but have the same mass and dimensions.
The first cylinder is hollow and the second cylinder is solid.



Hollow cylinder



Solid cylinder

The mass of the hollow cylinder is distributed around the radius, r.
This means that $I = mr^2$ can be used to calculate its moment of inertia.

The mass of the solid cylinder is evenly distributed between its axis of rotation and the radius, r.
This means it is a disc and $I = \frac{1}{2}mr^2$ is used to calculate its moment of inertia.

This means it will be *twice* as hard to change the motion of the hollow cylinder as the solid cylinder, even though they have the same mass and radius.

Torque and Moment of Inertia

An **unbalanced torque** will produce an **angular acceleration**. As discussed above, the moment of inertia of an object is the opposition to a change in its angular motion. Thus the angular acceleration α produced by a given torque T will depend on the moment of inertia I of that object.

$$T = I\alpha$$

Rotational Kinetic Energy, E_K

The total kinetic energy of a rotating body is the sum of the kinetic energies of each point mass on the rotating body.

$$E_K = \frac{1}{2}mv^2 \qquad and \qquad v = r\omega$$

$$so \qquad E_K = \frac{1}{2}mr^2\omega^2$$

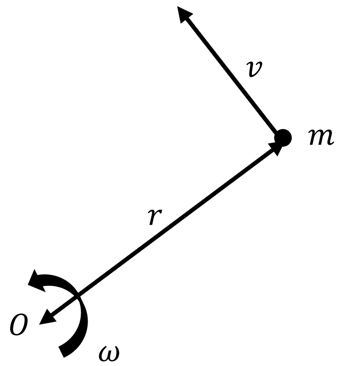
$$and \qquad I = mr^2$$

$$E_K = \frac{1}{2}I\omega^2$$

where all quantities and units are as previously stated.

Angular Momentum, L

The angular momentum L of a particle about an axis is defined as the **moment** of momentum.



A particle of mass m rotates at ω rads^{-1} about the point O, as shown.
The linear momentum of the particle is $p = mv$.

The moment of the particle $= mvr$ (r is perpendicular to v).

Thus the angular momentum of this particle:

$$L = mrv = mr^2\omega$$

For a rigid object about a fixed axis the angular momentum L will be the summation of all the individual angular momenta. Thus the angular momentum L of an object is given by $\sum(mr^2\omega)$. This can be written as $\omega\sum(mr^2)$ since all the individual parts of the object will have the same angular velocity ω. Also we have $I = \sum(mr^2)$.

The angular momentum, L, of an object is therefore the product of its moment of inertia and angular velocity:

$$L = I\omega$$

Angular momentum has the units kgm^{2}rads^{-1} (sometimes given as kgm^{2}s^{-1}).

Notice that the angular momentum of a rigid object about a fixed axis **depends** on the moment of inertia.

Angular momentum is a **vector** quantity. The **direction** of this vector is at right angles to the plane containing v (since $p = mv$ and mass is scalar) and r and lies along the axis of rotation.

Conservation of angular momentum

The **total** angular momentum before an impact will equal the **total** angular momentum after impact in the absence of **external torques**.

You will meet a variety of problems which involve use of the conservation of angular momentum during collisions for their solution.

Energy and work done

If a torque T is applied through an angular displacement, θ, then the:

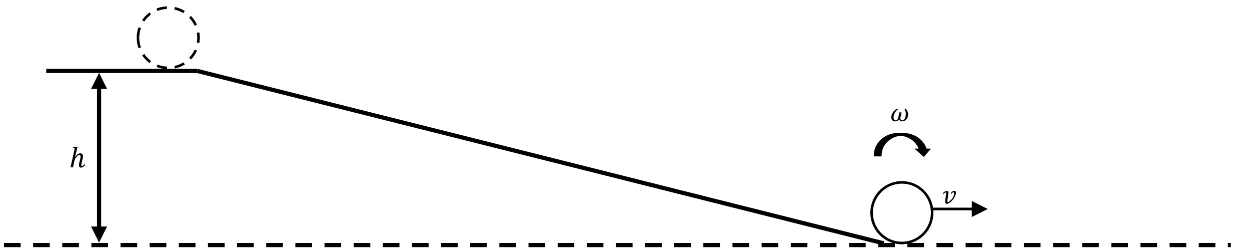
$$E_W = T\theta$$

Doing work produces a transfer of energy:

$$T\theta = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2$$

Objects Rolling down an Inclined Plane

When an object such as a sphere or cylinder is allowed to run down a slope, the potential energy at the top ($\boldsymbol{mgh}$), will be converted to both **linear** $\left(\frac{1}{2}\boldsymbol{mv^2}\right)$ and **rotational** $\left(\frac{1}{2}\boldsymbol{I\omega^2}\right)$ kinetic energy.



An equation for the energy of the motion (assume no slipping) is given below.

$$E_P \, at\ the\ top = Total\ E_k\ at\ the\ bottom$$

$$mgh = \frac{1}{2}mv^2 + \frac{1}{2}I\omega^2$$

The above formula can be used in an experimental determination of the moment of inertia of a circular object.

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, 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 brief historical note on these people follows.

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.

Nicolaus Copernicus (1510) insisted that the Sun and not the Earth was the centre of the solar system. He was the first to really challenge Ptolemy. He was the first to suggest that the Earth was just another planet. His great work published in 1543, "On the Revolutions of the Heavenly Spheres", had far reaching effects on others working in gravitation.

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, etc. In his mathematical analysis he required calculus - so he invented it as a mathematical tool!

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	$= 9.8\ ms^{-2}$
radius of the Earth, R_E	$= 6.4 \times 10^6$ m
radius of moon's orbit, r_M	$= 3.84 \times 10^8$ m
period, T, of moon's circular orbit	= 27.3 days = 2.36×10^6 s

Take: $\frac{R_E}{r_M} = \frac{1}{60}$

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.

$$a = g \propto \frac{1}{r^2}$$

Calculate the central acceleration for the Moon: use $a = \frac{v^2}{r}$ or $a = \frac{4\pi^2 R}{T^2}\ ms^{-2}$

Compare with the "diluted" gravity at the radius of the Moon's orbit according to the hypothesis:

$$\frac{1}{60^2} \times 9.8\ ms^{-2}$$

Conclusion

The inverse square law applies to gravitation.

Astronomical Data

Planet or satellite	Mass kg	Density kgm^{-3}	Radius m	Gravitational acceleration ms^{-2}	Escape velocity ms^{-1}	Mean distance from Sun m	Mean distance 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 only be explained if there existed a universal gravitational constant, given the symbol G.

We have already seen that Newton's of an inverse square law was correct. It also seems reasonable to assume that the force of gravitation will vary with the masses involved.

$$F \propto m \qquad F \propto M \qquad F \propto \frac{1}{r^2}$$

This gives:

$$F \propto \frac{Mm}{r^2}$$

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

where:

F = gravitational force in newtons, N
G = Universal Constant of Gravitation = **6.67 x 10^{-11} Nm^2kg^{-2}**
M = mass of star/planet in kilograms, kg
m = mass of planet/moon/satellite in kilograms, kg
r = separation between the centres of M and m in metres, m

Consider the Solar System

$$M = mass\ of\ the\ Sun$$
$$m = mass\ of\ orbiting\ planet$$

Force of attraction on a planet is:

$$F = \frac{GM_S m_P}{r^2}$$

Now consider the centripetal force if we take the orbit of the planet to be circular.

The centripetal force which maintains the planet's circular motion is provided by gravity. This means:

$$\frac{GM_S m_P}{r^2} = \frac{m_P v^2}{r}$$

Because $v = \frac{2\pi r}{T}$*:*

$$\frac{GM_S m_P}{r^2} = \frac{m_P}{r} \cdot \left(\frac{2\pi r}{T}\right)^2$$

$$\frac{GM_S m_P}{r^2} = \frac{4 m_P \pi^2 r}{T^2}$$

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

This is called Kepler's Third Law and it proves that the orbital period and radius of the orbit are related as follows:

$$T^2 \propto r^3$$

This means that $\frac{T^2}{r^3}$ *is a constant value and hence shows that G is a constant for all planets in the Solar System.*

Note:
- 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.
- Gravitational forces are 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 of mass m on its surface. The gravitational force of attraction can be given by two equations:

$$F = mg \qquad and \qquad \frac{GM_E m}{{R_E}^2}$$

where:

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

Thus:

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

$$M_E = \frac{gR_E^2}{G}$$

$$M_E = \frac{9.8 \times (6.4 \times 10^6)^2}{6.67 \times 10^{-11}}$$

$$M_E = 6.02 \times 10^{24}\ kg$$

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.

$$F_{gravity} = mg$$

We also take:

$$E_P = mgh$$

where for both calculations, g = constant = 9.8 Nkg^{-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.

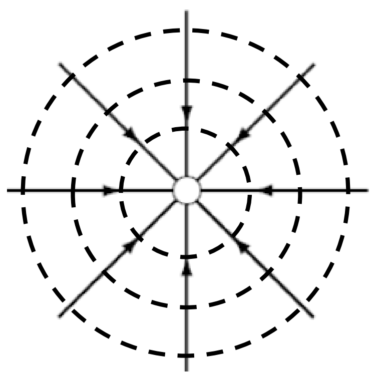
Definition of Gravitational Field Strength at a point:

This is defined to be the weight/force per unit mass at the point. i.e. $g = \frac{F}{m}$

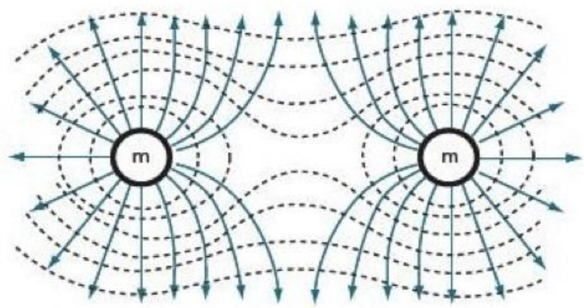
The concept of a field was not used in Newton's time. Fields were introduced by Faraday in his work on electromagnetism and only later applied to gravity.

Note that g and F above are both vectors and whenever forces or fields are added this must be done vectorially.

Field Lines



An isolated 'point mass'



Two equal 'point masses'

The equipotential lines join the points with equal gravitational field strength. They are the broken lines. The field lines show the direction that an object with mass placed in the gravitational field would travel.

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

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

An object's weight is the gravitational force exerted on it by the Earth:

$$\boldsymbol{W = mg \qquad where\ m = object\ mass}$$

$$F = \frac{GMm}{r^2} \qquad so \qquad mg = \frac{GMm}{r^2}$$

To find gravitational field strength, weight per unit mass, on any planet:

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

where:

M = mass of planet in kilograms, kg
r = radius of planet, m

Note:
$$g \propto \frac{1}{r^2}$$
above the surface of the Earth

The density of the Earth is not uniform and this causes an unusual variation of g with radii inside the Earth.

Note:
When calculating g at a height above the Earth, remember to add the height above the surface to the radius of the Earth (6.4×10^6 m) to find r.

Gravitational Potential, V

As the gravitational field strength varies with distance from the surface of a planet, we cannot always use $E_P = mgh$ to calculate gravitational potential energy – this only works over small distances where g is approximately constant.

The **gravitational potential, V,** at a point in space is defined as **the work done by external forces to move a unit mass from infinity to that point.**

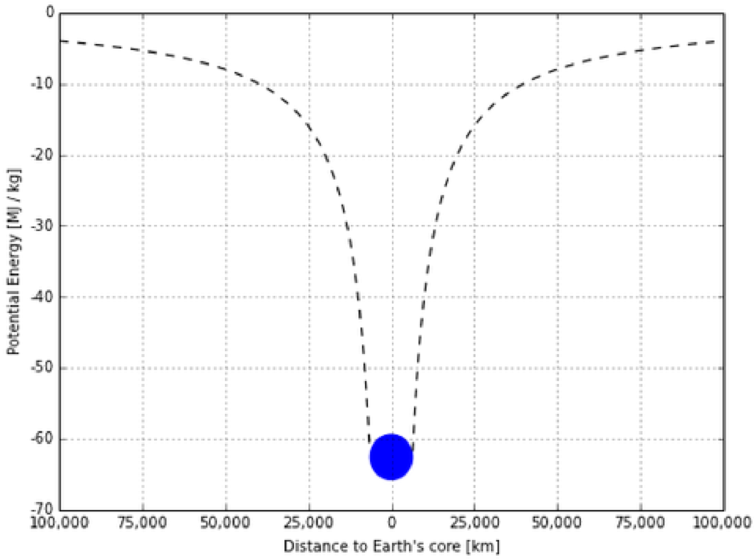
Note:
Infinity is where the force of gravity is zero.

The gravitational potential V at a distance r from a mass M is given by:

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

The unit for gravitational potential is Jkg^{-1}.
To calculate the gravitational potential energy of a mass m_1 which is a distance r from a mass m we need to use gravitational potential.
As the unit for V is Jkg^{-1} we know that it must be defined as $\frac{energy}{mass}$.

The Gravitational Potential 'Well' of the Earth



This graph gives an indication of how masses are 'trapped' in the Earth's field.

Gravitational Potential Energy

Rather than use $E_P = mgh$ for large distance from the Earth's surface, we use the idea of gravitational potential to find the potential energy of a mass m at a distance r from the centre of planet M.

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

$$\frac{E}{m} = -\frac{GM}{r}$$

$$\boxed{E_P = -\frac{GMm}{r}}$$

where:

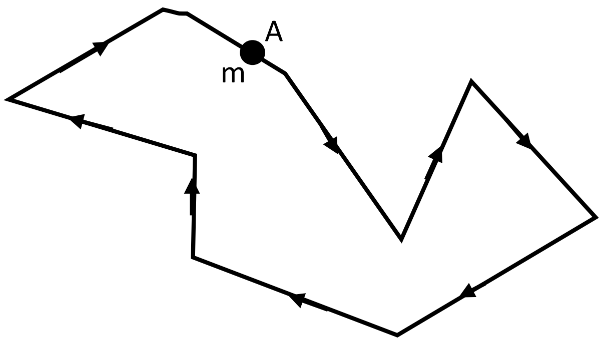
M = mass of planet, kg
m = mass of object, kg
r = distance between the centres of M and m, m

The negative sign occurs due to the fact that moving m away from M requires an increasing amount of work done overcoming the attractive force between the two masses. E_P increases to zero at infinity so must be negative at all points between m and infinity.

Conservative field

The force of gravity 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 m 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

Escape velocity is the minimum velocity required to just escape the planet's gravitational field and reach infinity with zero velocity.

$$E_K + E_P = 0$$

$$\frac{1}{2}mv^2 + \left(-\frac{GMm}{r}\right) = 0$$

$$\frac{1}{2}mv^2 = \frac{GMm}{r}$$

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

$$\boxed{v = \sqrt{\frac{2GM}{r}}}$$

Atmospheric Consequences:

$v_{r.m.s.}$ of H_2 molecules = 1.9 kms^{-1} (at 0°C)

$v_{r.m.s.}$ of O_2 molecules = 0.5 kms^{-1} (at 0°C)

When we consider the range of molecular speeds for hydrogen molecules it is not surprising to find that the rate of loss 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 kms^{-1}) is so small that any 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/small radius could have an escape velocity greater than the speed of light, $3 \times 10^8\ ms^{-1}$. 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 Einstein in his General Theory of Relativity (1915). Another physicist called Schwarzschild calculated the radius of a spherical mass from which light cannot escape. It is $r = \frac{2GM}{c^2}$.

Photons are affected by a gravitational field. There is gravitational force of attraction on the photon. Thus photons passing a massive star are deflected by that star and stellar objects 'behind' the star appear at a very slightly different position because of the bending of the photon's path.

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 of radius r and mass M. The energy of the photon, hf decreases as it travels to positions of greater gravitational potential energy 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.

Note:
This is known as the gravitational redshift, not the well-known Doppler redshift caused by the expanding universe.

If the mass and density of the body are greater than certain critical values, the frequency of the photon will 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.

Satellite Motion

Satellites, both natural and artificial, are kept in orbit by gravity. The gravitational force attracting the satellite is equal to the centripetal force which ensures that it maintains circular motion.

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

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

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

$$v = \sqrt{\frac{GM}{r}}$$

Note:
This equation is often confused with the equation for escape velocity – they are totally different!

Kepler's Third Law can also be applied to find the period of a satellite in orbit:

$$T^2 = \frac{4\pi^2 r^3}{GM}$$

Energy and Satellite Motion

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

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

Re-arranging, we get:

$$\frac{1}{2}mv^2 = \frac{GMm}{2r}$$

so:

$$E_K = \frac{GMm}{2r}$$

Note:
E_k is always positive.

But the gravitational potential energy of the system is:

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

Note:
E_P is always negative.

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.

Thus, the total energy of the satellite is:

$$E_T = E_P + E_K$$

$$E_T = -\frac{GMm}{r} + \frac{GMm}{2r}$$

$$\boxed{E_T = -\frac{GMm}{2r}}$$

Some 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 in equal times.
- The square of the orbital period of a planet is proportional to the cube of the semi-major axis of the orbit.

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 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.

Variation of "g" over the Earth's Surface

The greatest value for "g" at sea level is found at the poles and the least 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:

$$a = \frac{v^2}{r} \quad and \quad v = \frac{2\pi r}{T} \quad giving \quad a = \frac{4\pi^2 r}{T^2} = \frac{4\pi^2 \times 6.4 \times 10^6}{(24 \times 60 \times 60)^2} = 0.034\, ms^{-2}$$

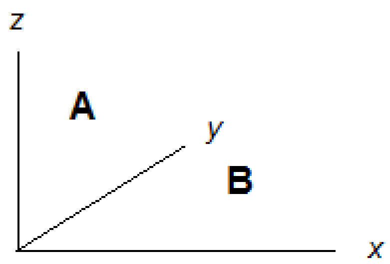
Observed values for "g": at poles = 9.832 $ms^{-2}$and at equator = 9.780 ms^{-2}.
This give a difference of $0.052\ ms^{-2}$.
Most of the difference has been accounted for. The remaining $0.018\ ms^{-2}$ is due to the non-spherical shape of the Earth. The equatorial radius exceeds the polar radius by $21\ km$. 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 $9.81\ or\ 9.82\ ms^{-2}$. 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.

Spacetime

The concept of spacetime relies on the fact that to give co-ordinates to an exact point, as well as the three dimensions x, y and z a fourth dimension t (for time) must be given so that we reach a particular point at the correct time.



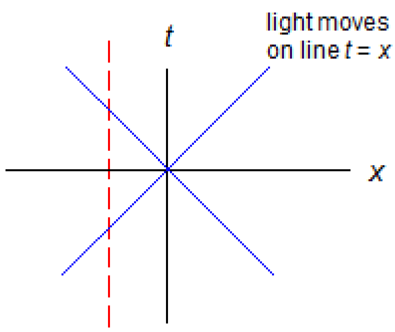
In Newtonian mechanics we use Euclidean space, consisting of three mutually perpendicular directions, often denoted by x, y, z.

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 impossible to draw four dimensions on paper, but one or more of the spatial dimensions can be suppressed with one dimension of space.

Spacetime Diagram

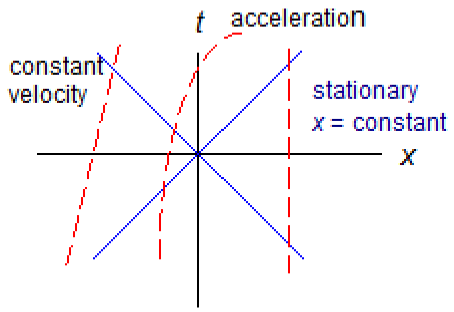


The above diagram is a representation of this. A light beam emitted from the origin will travel on the lines $t = x$ and $t = -x$. For an event E at the origin the region between the blue diagonal lines above the x-axis defines all events in the future that could be affected by E. Similarly, the region between the solid blue lines below the x-axis defines past events affecting E.

These two regions are referred to as timelike.

The regions to the left and right are called spacelike 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.

The dashed red vertical line represents the worldline of a stationary particle. Here the value of x remains the same as the time moves forward.



The above spacetime diagram has other worldlines added. Another name for a worldline 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 spacetime.

As can be seen from the diagram 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 gradient. Acceleration can be represented by a curve with changing gradient.

Principle of covariance

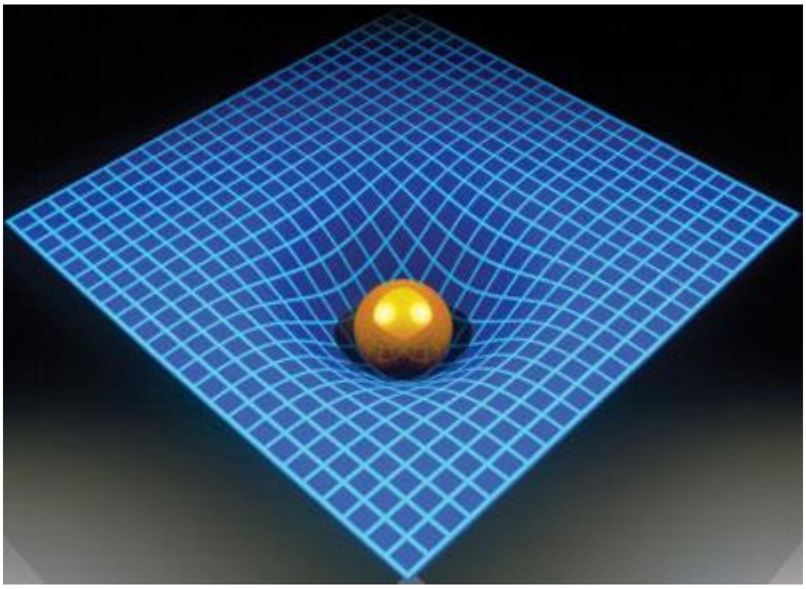
Physics is described by equations that put all spacetime co-ordinates on an equal footing.

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

General relativity

General relativity leads to the interpretation that mass curves space and what we perceive as gravity arises from the curvature of space.

As the mass of an object increases, the distortion of spacetime 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 spacetime. Collapsing stars can be very small and occupy a very small space, which in turn causes a point of infinite curvature that leads to an event horizon.

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 spacetime 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 distance at which light just escapes is termed the event horizon or Schwarzschild radius.

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.

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

where:

c = velocity of light, 3 x 10^8 ms^{-1}
M = mass, kg
G = Universal Gravitational Constant, 6.67 x 10^{-11} Nm^2kg^{-2}

The mass of the sun is approximately 2.0 × 10^{30} kg, G = 6.67 × 10^{-11} Nm^2kg^{-2} and c = 3.0 × 10^8 ms^{-1}, so:

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

$$r = \frac{2 \times 6.67 \times 10^{-11} \times 2.0 \times 10^{-30}}{(3 \times 10^8)^2}$$

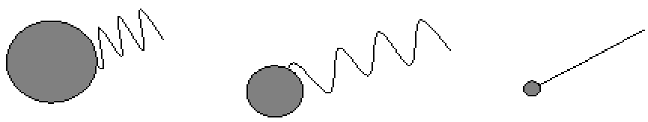
$$r = 2964\ m$$

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.’

Clocks run slow when in a stronger gravitational field.

In a strong gravitational field light appears to be Doppler shifted and ‘stretched’, but it still travels 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.



As mass is compressed into smaller volumes the gravitational strength at the surface increases. Seen by a distant observer a clock will appear to stop ticking at the event horizon.

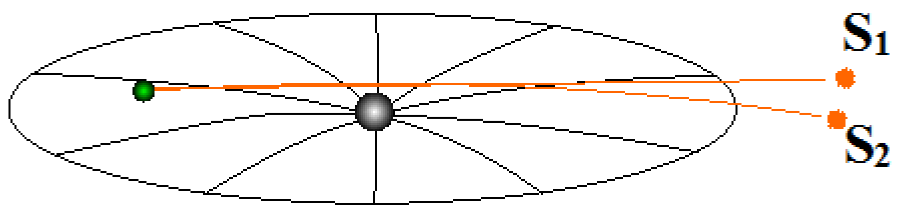
Global positioning systems

To achieve accuracy requirements global positioning systems (GPS) use the principles of general relativity to correct satellites’ atomic clocks.

Evidence for general relativity

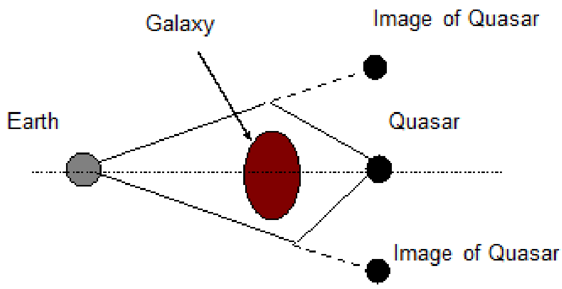
As has been previously stated general relativity indicates that a beam of light will experience curvature similar to that experienced by a massive object in a gravitational field. In order to see light being ‘bent’ by the Sun we need to view this effect during an eclipse.

This is done by measuring the position of a star during an eclipse (S2) and again when the Sun is in another part of the sky (S1). The angle between S1 and S2 can then be compared to the theoretical value.



Gravitational lensing

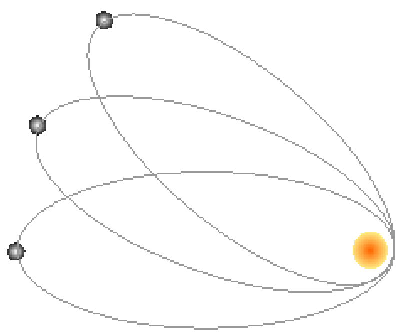
This effect can occur when light from a distant star is bent around a massive object on route to Earth. On Earth we might observe a circle or arc of light, although the original light comes from a single source, for example light from a distant quasar (quasi-stellar radio source).



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 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 spacetime near to the Sun is curved and this changes by a tiny amount the predicted orbits of the planets. 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.

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 ultimate fate.

Figure 1 An example of a spiral galaxy.

Throughout the Universe there are estimated to be something like 100 billion (1011) galaxies (Figure 1), each consisting of perhaps 1011 stars, meaning that there are in the order of 1022 stars in the observable Universe. Naked-eye observation from Earth reveals just a few thousand of these.

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 (1023 km) away. This means that travelling at the speed of light of 3.0 × 108 m s–1, light 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. For example Figure 2 shows a solar flare, an example of a process occurring in stars, such as our Sun.

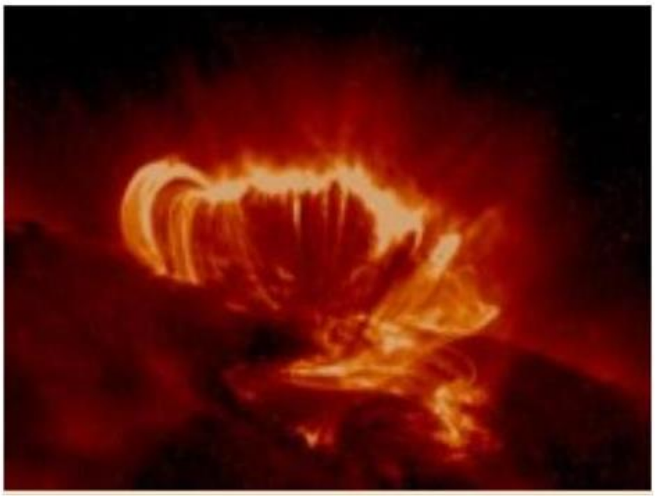


Figure 2 Solar flare.

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 = mc2 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 in Figures 3 to 5.

Fusion in the Sun

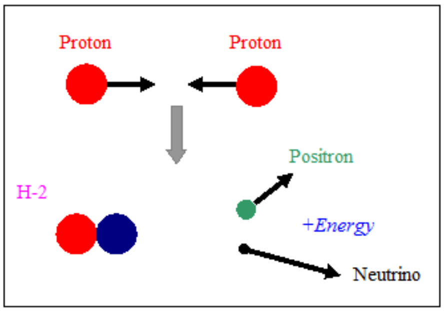


Figure 3 Two protons fuse and one is converted to a neutron to form deuterium (H-2). Energy is released.

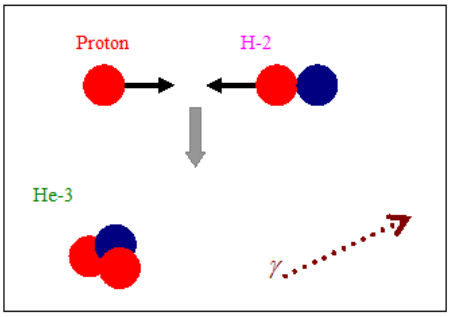


Figure 4 The deuterium captures another proton to form helium-3 (He-3). Energy is released as gamma rays

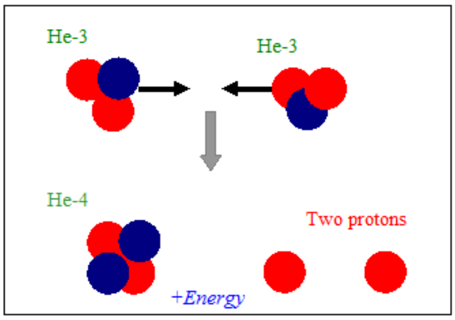


Figure 5 Two helium-3 nuclei fuse, giving helium-4 (He-4) and freeing two protons.

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

$$4\,({}^{1}_{1}H) \rightarrow {}^{4}_{2}He + 2\,e^{+} + 2\,neutrinos + energy$$

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, capable of overcoming the inter-nuclei repulsive electrostatic forces. 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 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 prodigious 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). Figure 6 shows one method of creating a fusion reaction on Earth, notice how different this is to that of our Sun.

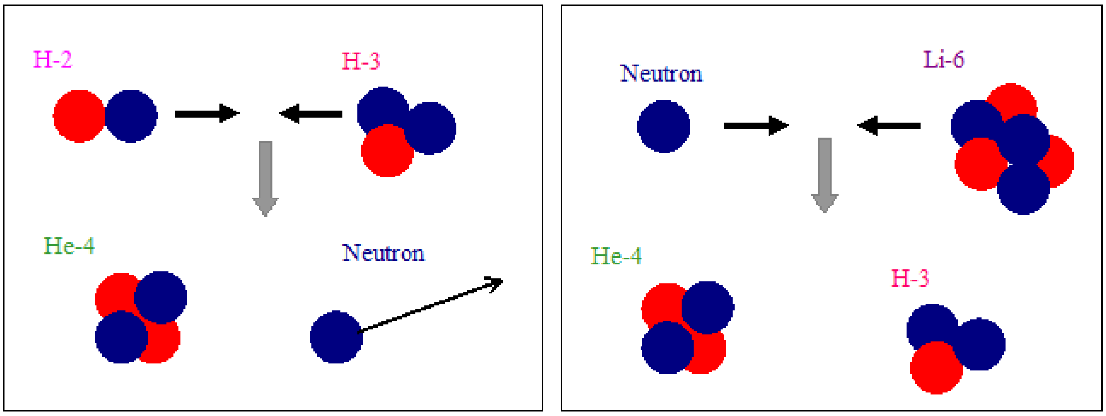


Figure 6 A possible method of fusion used on Earth. Deuterium and tritium (H-3) are heated to a very high temperature. The neutrons released fuse with lithium (Li-6) around the hot gases, breeding more tritium. Most of the useful energy comes from the deuterium–tritium reaction.

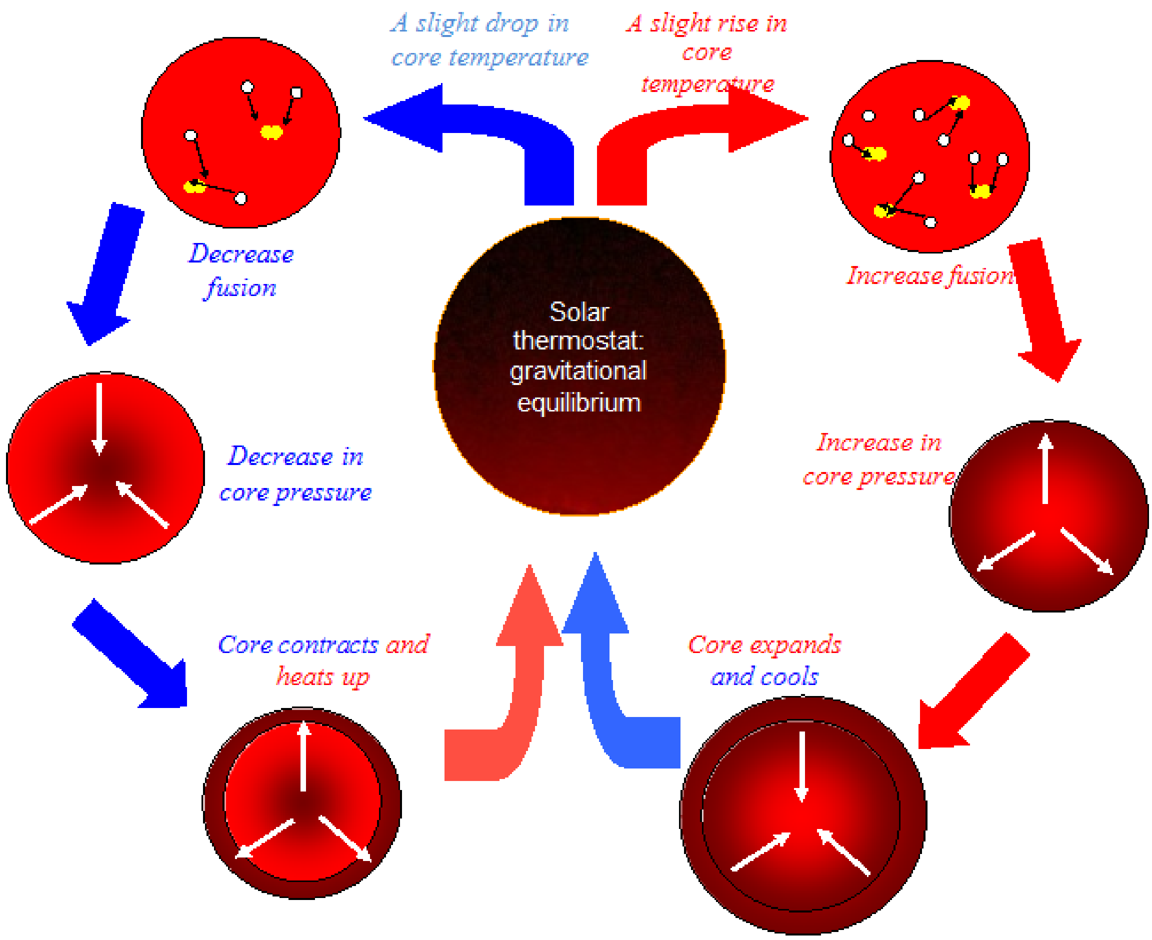


Figure 7 The solar thermostate.

The Structure of the Sun

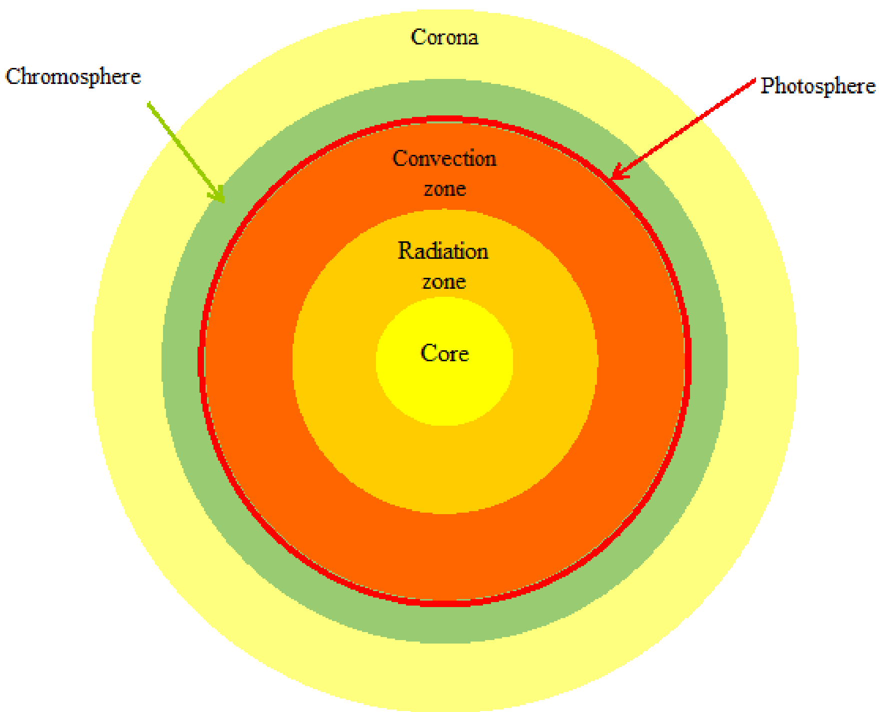


Figure 8 The solar zones.

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 Wm^{-2}, 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 (Figure 8) 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 of each layer are given in Table 1.

Table 1 Temperature of solar regions

Region	Typical Temperature (K)
Core	15×10^6
Radiation zone	10×10^6
Convection zone	10×10^6 to 6,000
Photosphere	5,800
Chromosphere	10,000
Corona	1×10^6

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) (Moore, 2002). White dwarfs (small, dense stars nearing the end of their lives) may have a radius of around 0.01 R_{Sun}, whilst at the other extreme supergiants may typically be 500 R_{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–Boltzman law.

$$P = \sigma T^4$$

where:

P is the power per unit area (Wm^{-2})
σ is the Stefan–Boltzman constant (5.67 × 10–8 $Wm^{-2}K^{-4}$)
T is the absolute temperature (K)

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

$$P = \sigma T^4$$
$$P = 5.67 \times 10^{-8} \times (5800)^4$$
$$P = 64 \times 10^6 \; Wm^{-2}$$

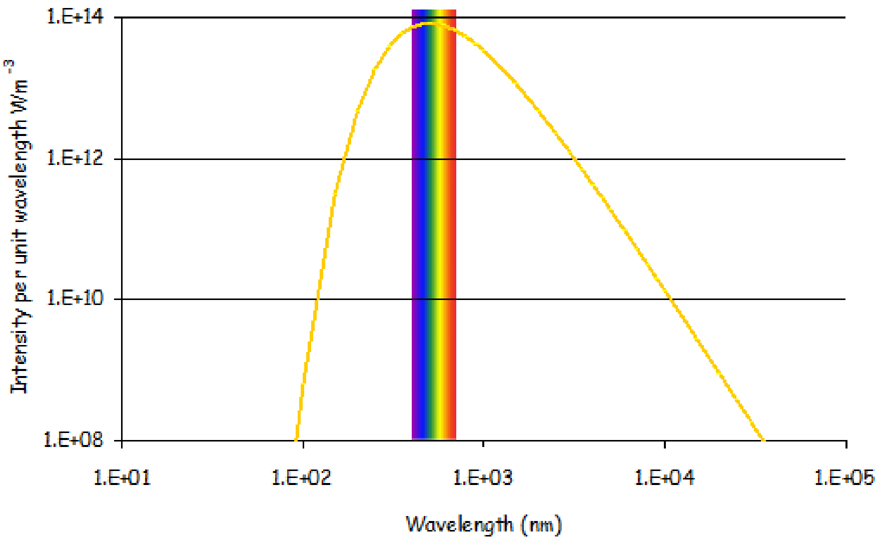


Figure 9 Black-body radiation from the Sun

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 in Figure 9. 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. Figure 9 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.

This curve is given by the Planck radiation law:

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5 \left(e^{\frac{hc}{\lambda kT}} - 1\right)}$$

where:

I(λ) is the radiation intensity per unit wavelength
h is Planck's constant
c is the speed of light
k is the Boltzman constant.

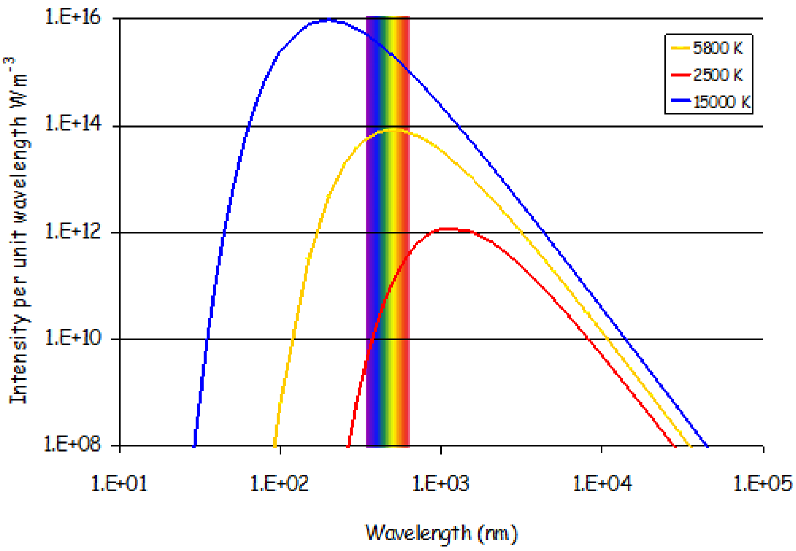
The maximum intensity in Figure 9 occurs at a wavelength of:

$$\lambda_{max} = \frac{2.9 \times 10^6}{T} \qquad (Wien's\ Law)$$

and 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 (Figure 10) 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 : 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 (Moore, 2002) or roughly 330,000 times that of the Earth. Other stars range in size from roughly 0.08 M_{Sun} to 150 M_{Sun}. 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 should this be?

Stars with a mass greater than 150 M_{Sun} 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 M_{Sun} 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 don't propose to cover here, but the outcome is that this outward pressure prevents the star from collapsing and raises the core temperature and pressure enough to sustain nuclear fusion.

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

Luminosity

The luminosity of a star is a measure of the total power the star emits. It is calculated by multiplying the power per unit area, P, by the surface area of the star, thus:

$$L = P \times 4\pi r^2$$

and as $P = \sigma T^4$:

$$\boxed{L = 4\pi r^2 \sigma T^4}$$

where:

L is the luminosity of the star (W)
P is the power per unit area (Wm^{-2})
r is the radius of the star (m)
σ is the Stefan–Boltzman constant (5.67 × 10–8 $Wm^{-2}K^{-4}$)
T is the absolute temperature (K)

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

$$L_{Sun} = P \times 4\pi r^2$$

$$L_{Sun} = 64 \times 10^6 \times 4\pi \times (6.96 \times 10^8)^2$$

$$L_{Sun} = 3.9 \times 10^{26}\ W$$

The Sun has a luminosity, L_{Sun} of 3.9 x 10^{26} W, and the range of stellar luminosities found throughout the universe is from around 10^{-3} L_{Sun} to 10^5 L_{Sun}.
Because the range of luminosities covers eight to nine orders of magnitude, the luminosity can also be described using the absolute 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 visual magnitude of a particular star would be when viewed from a distance of 10 parsecs (a parsec is 3.26 light years). The definition of luminosity has changed slightly over the years, but the modern scale can be obtained using:

$$L = 10\frac{4.83 - M}{2.5}L_0$$

where:

L_0 is the luminosity of our Sun
M is the absolute brightness of the star
L the luminosity of the star.

The form of this equation makes it easy to see that the scale is not linear. In fact, a difference of 5 in magnitude corresponds to a 100-fold difference in luminosity. It is also apparent from the equation that 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.

Apparent brightness

When viewed from Earth, the brightness of stars does not necessarily reflect their 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

$$Apparent\ brightness = \frac{L}{4\pi d^2}$$

where d is the distance from the star to our observation point (usually Earth).

To apply this to our Sun gives

$$Apprarent\ brightness = 3.839 \times \frac{10^{26}}{4\pi(150 \times 10^9)^2} = 1360\ Wm^{-2}$$

and 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 apparent 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 (ie 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 –1.44. 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 is –26.72.

Detecting astronomical objects

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 µm 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 number 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 wavelengths detectable 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).

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 Table 2 (Moore, 2002).

Table 2: 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.

Table 3: 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 – 6,000	Polaris	Yellow-white
G	6,000 – 5,000	Sun	Yellow
K	5,000 – 3,500	Arcturus	Orange
M	< 3,500	Proxima Centauri, Betelgeuse	Red

The order of the spectral classes has been remembered by generations of astronomy students using the mnemonic 'Oh, Be A Fine Girl/Guy, Kiss Me'.

This work was carried out at the Harvard Observatory by (mostly female) 'computers'. These individuals 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 listed in Table 3 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 (Figures 11 and 12), with the bright areas being regions where hotter plasma is welling up from below and dark regions those where cooler plasma is sinking.

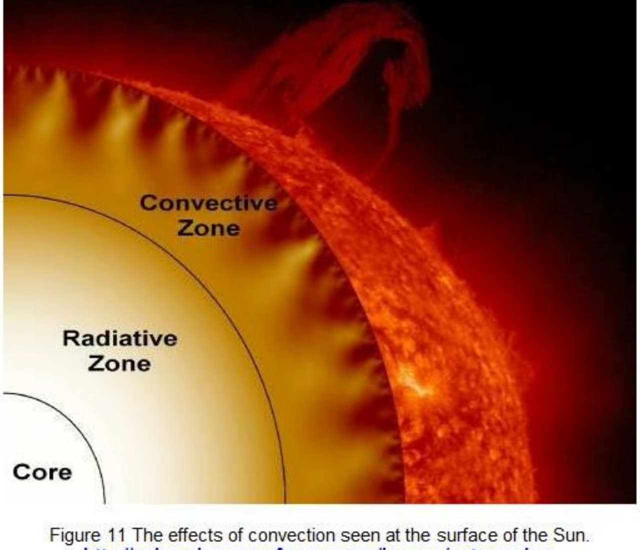


Figure 11 The effects of convection seen at the surface of the Sun.
http://solarscience.msfc.nasa.gov/images/cutaway.jpg.

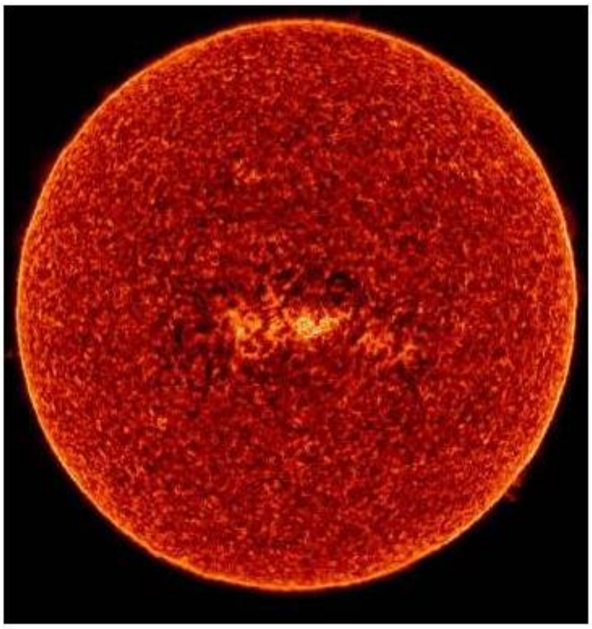
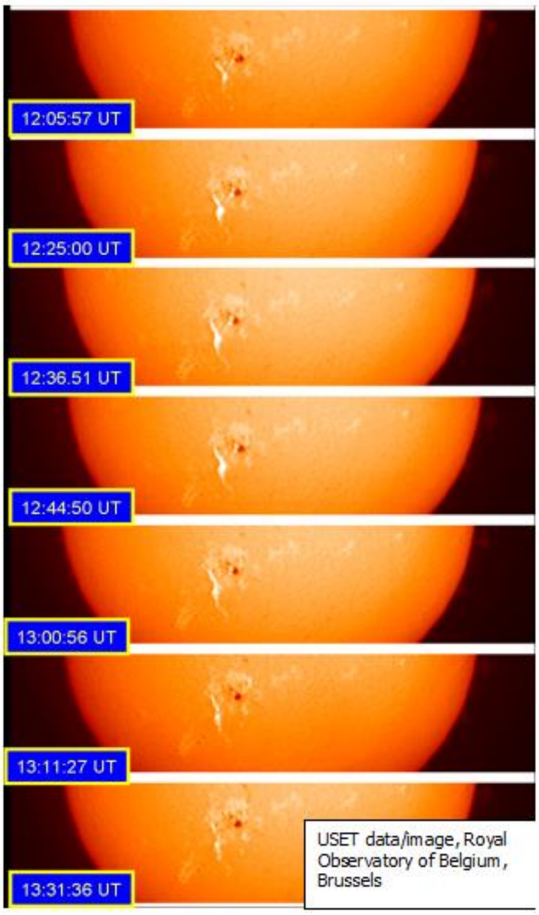


Figure 12 Convection currents observed at the solar surface.
http://solarscience.msfc.nasa.gov/images/sumer_S_VI.jpg.

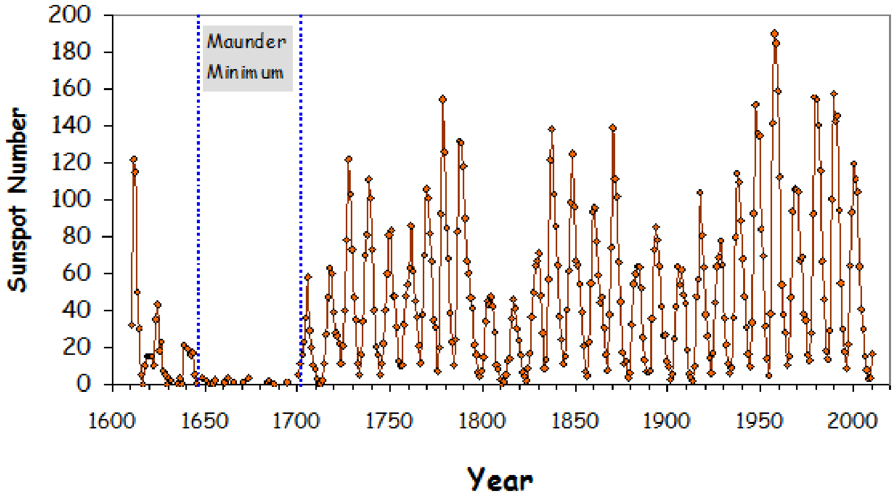
From time to time, larger dark areas appear, ranging in size from 1000 km to 100,000 km in diameter (with the Earth for comparison being approximately 13,000 km in diameter). These sunspots, which usually occur in pairs, are thought to arise when strong magnetic field lines emerge from the surface of 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. Figure 13 shows a time-sequence of a flare taken in August 2002 showing the movement of the flare away from the region of a sunspot.



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 (Figure 14).

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. 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 imperfectly 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 kms^{-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 dangerous to those aboard the International Space Station.

In recent years there have been several CME warnings, the most serious outcomes being: (i) in late 2003 a series of CMEs close together destroyed a satellite, disrupted others and caused problems with aircraft navigation systems; (ii) in 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 latter 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.

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 molecules, 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 × 1024 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 (Figure 15), 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 arose as a result of nuclear fusion in these first stars.

Figure 15 The Horsehead Nebula. Bright stars can be seen shining through the dust clouds.

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 then hydrogen and helium could have arisen in the Universe (Singh, 2004). The production of these heavier elements by nuclear fusion in a star of mass 25MSun is summarised in Table 4, which shows that progressively higher temperatures and densities are required to produce the heavier elements.

Table 4 Summary of nucleosynthesis occurring in a star of mass 25 M_{Sun}.

Stage	Temperature (K)	Density (gcm^{-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	$3 X 10^7$	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 in Table 4 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.

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 0.08MSun to 150MSun because out with 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 showed 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; Figure 16) 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.

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

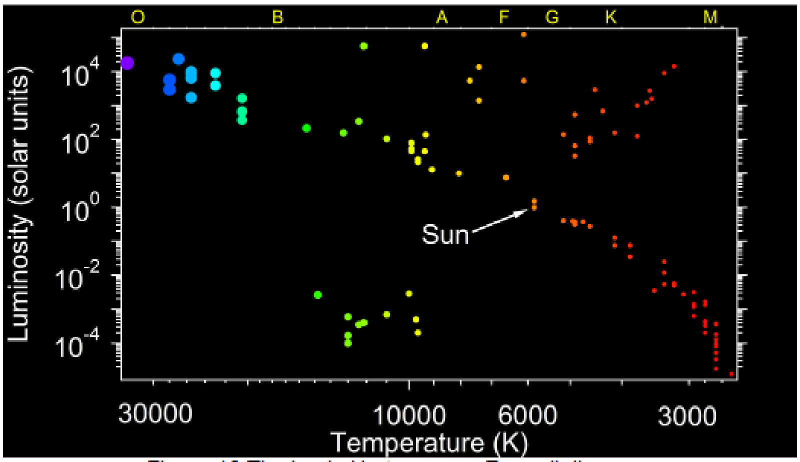


Figure 16 The basic Hertzsprung–Russell diagram.

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 (Figure 17).

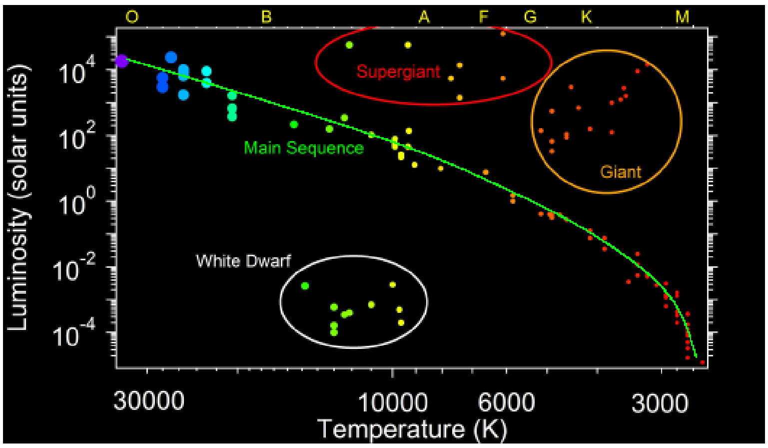


Figure 17 The principal star groups.

Main sequence

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 25000 K, whilst low mass stars may have surface temperatures as low as 2000 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 for. 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 Table 4, 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 (Figure 18), 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 stars

Supergiants

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 Table 4.

Figure 18 The glowing eye of planetary nebula NGC6751.

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 1011 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 ms^{-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 of such an entity is so strong that even light itself cannot escape. These objects are known as black holes.

The distance from within which nothing can escape a body is called the Schwarzschild radius and is given by :

$$r_{schwarzchild} = \frac{2GM}{c^2}$$

where:

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

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 (Figure 19). 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.

Figure 19 Crab Nebula.

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 million 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.