A-level Physics
Summer Independent Learning
Y11-12

Part 1: Compulsory work (pages 2 to 15)

Part 2: Strongly recommended work (pages 16 to 17)

Welcome to A Level Physics, please complete the following tasks ready for your first day at New College. You can either write on the document electronically, print the document out or write your notes and answers on paper to bring in for your first lesson in September.

You may have to **research** any knowledge or techniques you cannot immediately recall using common GCSE resources or other tutorials.

Please be aware that you will have an **assessment** on these topics shortly after beginning your A level Physics course and the knowledge covered is essential to understanding the subsequent content

Compulsory work

Watch the videos from alevelphysicsonline.com/practical-skills to complete the definitions below. You are encouraged to use diagrams to support your answers.	

Accuracy, Precision, Error and Uncertainty (link)

Define the following terms:

a. Accurate

b. Precise

Absolute uncertainty (link)

a. When measuring an object with a ruler (with a resolution of 1 mm), how many judgements are made?

b. What is the total **absolute uncertainty** for a ruler? Why?

c. What is the advantage of using callipers over a ruler?

d. Why might recording time have a different **absolute uncertainty** to that suggested by the apparatus?

Variables (link)

a. Define Independent Variable.

b. Define Dependent Variable

c. Define Control Variables

Control Variable, Fair Tests and Causation (link, Extra watching)

a. Describe the difference between the terms **Repeatable** and **Reproducible.**

Zero error (link)

a. What is a zero error? Provide an example

b. State one way that we can account for zero errors.

Parallax error (link)

a. Describe parallax errors

b. Describe **two** ways that you can reduce parallax errors when taking measurements.

Gradients & y-intercepts (link)

a. Describe how to find a y-intercept from a graph when your graph does not directly extrapolate to $x = 0$.

As a physicist you will develop your problem-solving skills, scientific writing, and analytical skills with an emphasis on improving your experimental techniques. The following is to prepare you for the course and may take some students longer to complete than others. If you find some of the maths tricky do not be intimidated. Practice before beginning the course.

> ‘But science is as much for intellectual enjoyment as for practical utility, so instead of just spending a few minutes, we shall surround the jewel by its proper setting in the grand design of that branch of mathematics called elementary algebra.’
>
> Richard P. Feynman, Feynman Lectures on Physics

ALGEBRA

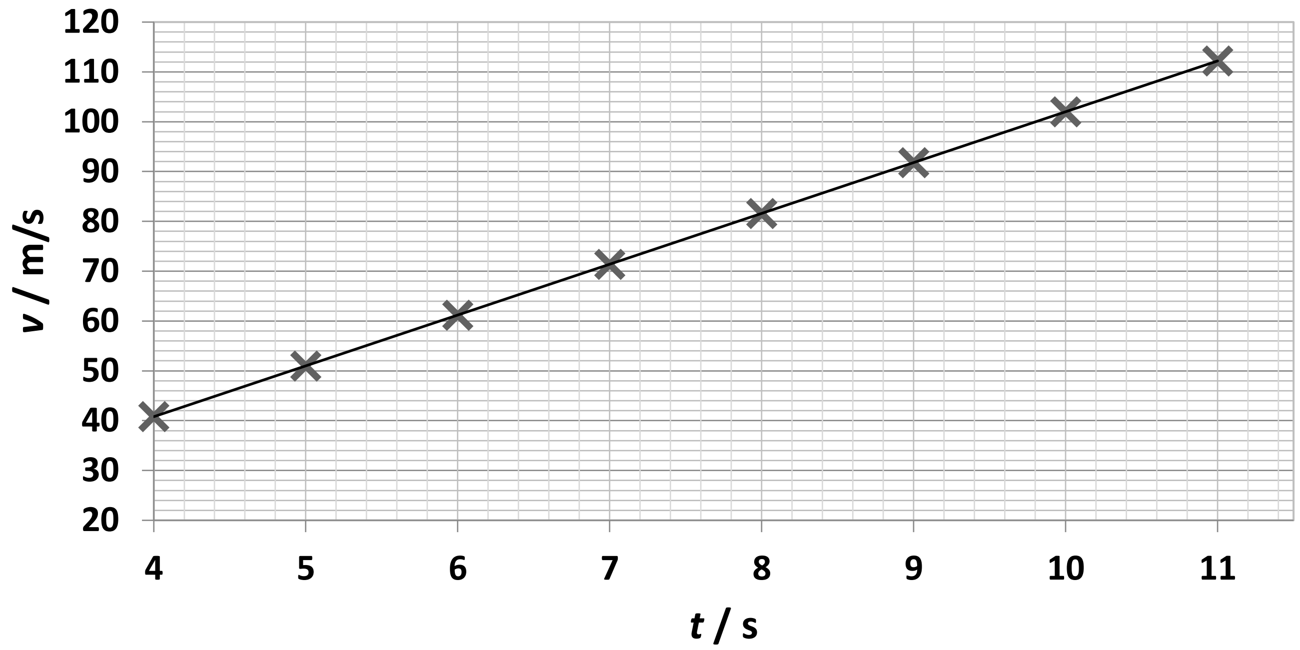
Q1. Identify the odd one out:

$$\frac{ab}{c},\quad a\left(\frac{b}{c}\right),\quad \left(\frac{1}{c}\right)ab,\quad a\left(\frac{1}{cb}\right),\quad a\left(\frac{1}{c}\right)b,\quad ab \div c,\quad (a \div c) \times b$$

Q2. Re-arrange the following equations to make the variable in bold the subject. Show your steps:

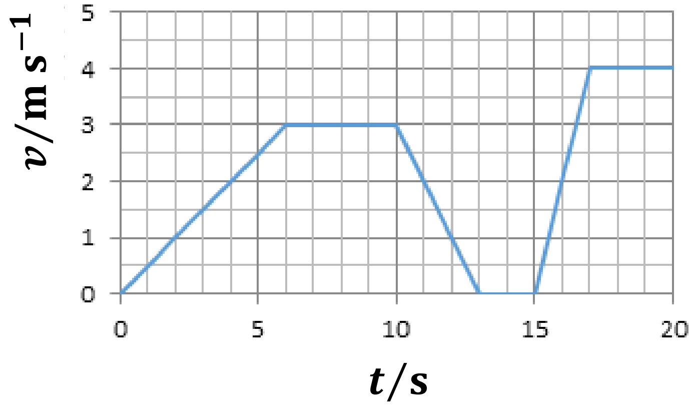
Equation	Working	Solution
$a = \frac{bc}{\boldsymbol{d}} + f$	$a - f = \frac{bc}{d}$ $d(a - f) = bc$ $d = \frac{bc}{a - f}$	$d = \frac{bc}{a - f}$
$V = I\boldsymbol{R}$		
$E_p = mg\boldsymbol{h}$		
$E_k = \frac{1}{2}\boldsymbol{m}v^2$		
$v^2 = u^2 + 2\boldsymbol{a}s$		
$E_k = \frac{1}{2}m\boldsymbol{v}^2$		
$V = \frac{4}{3}\pi\boldsymbol{r}^3$		

Q3. Calculate the following for the graph of velocity, v, in $\mathrm{m/s}$ against time, t, in s shown below:



a. gradient of the line to find the acceleration. (include an appropriate unit)	**b.** the area under the line to find the distance travelled in the interval $4 - 11$ s. (include an appropriate unit)	**c.** the v-intercept (i.e. at $t = 0$ s) to find the initial velocity. (include a sensible unit)

Q4. The graph below shows the velocity-time graph for a cat moving down a corridor.



Describe the motion of the cat referencing its velocity and acceleration.

DEFINITIONS

1. Link each term to the correct definition on the right:

Hypothesis	The maximum and minimum values of the independent or dependent variable
Dependent variable	A variable that is kept constant during an experiment
Independent variable	The quantity between readings, eg a set of 11 readings equally spaced over a distance of 1 metre would give an interval of 10 centimetres
Control variable	A proposal intended to explain certain facts or observations
Range	A variable that is measured as the outcome of an experiment
Interval	A variable selected by the investigator and whose values are changed during the investigation

2. Link each term to the correct definition on the right:

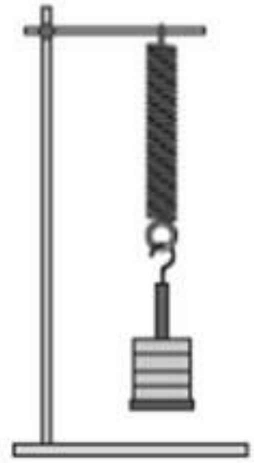
True value	The range within which you would expect the true value to lie
Accurate	A measurement that is close to the true value
Resolution	Repeated measurements that are very similar to the calculated mean value
Precise	The value that would be obtained in an ideal measurement where there were no errors of any kind
Uncertainty	The smallest change that can be measured using the measuring instrument that gives a readable change in the reading

3. Link each term to the correct definition on the right:

Random error	Causes readings to differ from the true value by a consistent amount each time a measurement is made
Systematic error	When there is an indication that a measuring system gives a false reading when the true value of a measured quantity is zero
Zero error	Causes readings to be spread about the true value, due to results varying in an unpredictable way from one measurement to the next

INVESTIGATING SPRINGS

A group of students investigated how the extension of a spring varied with the force applied. They did this by hanging different weights from the end of the spring and measuring the extension of the spring for each weight.



The results are below.

Weight added to the spring / N	Extension of spring / cm			
	Trial 1	Trial 2	Trial 3	Mean
2	3.0	3.1	3.2	
4	6.0	5.9	5.8	
6	9.1	7.9	9.2	
8	12.0	11.9	12.1	
10	15.0	15.1	15.12	

a. What do you predict the result of this investigation will be?

b. What are the independent, dependent and control variables in this investigation?

c. What is the difference between repeatable and reproducible?

d. What would be the most likely resolution of the ruler you would use in this investigation?

e. Suggest how the student could reduce parallax errors when taking her readings.

f. Random errors cause readings to be spread about the true value. What else has the student done to reduce the effect of random errors and make the results more precise?

g. Another student tries the experiment but uses a ruler which has worn away at the end by 0.5 cm. What type of error would this lead to in his results?

h. Calculate the mean extension for each weight.

i. A graph is plotted with force on the *y* axis and extension on the *x* axis. What quantity does the gradient of the graph represent?

STANDARD FORM

1. Write the following numbers in standard form.

a. 379 4

b. 0.0712

2. Use the A-level data booklet to write the following as ordinary numbers.

a. The speed of light

b. The charge on an electron

3. Write one quarter of a million in standard form.

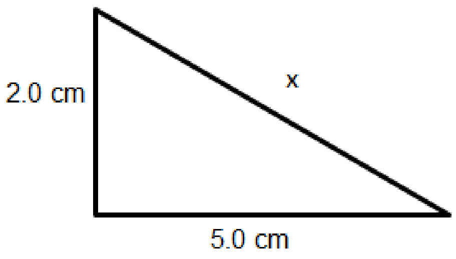
4. Write these constants in ascending order (ignoring units).

- Permeability of free space
- The Avogadro constant
- Proton rest mass
- Acceleration due to gravity
- Mass of the Sun

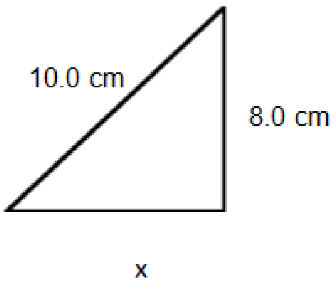
A-level data booklet

TRIGONOMETRY & PYTHAGORAS

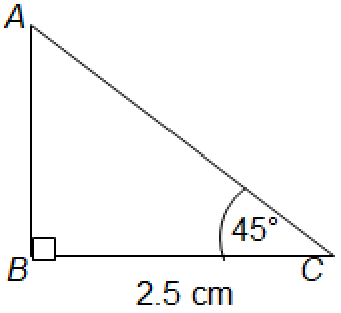
1. Calculate the length of side x.



2. Calculate the length of side x.

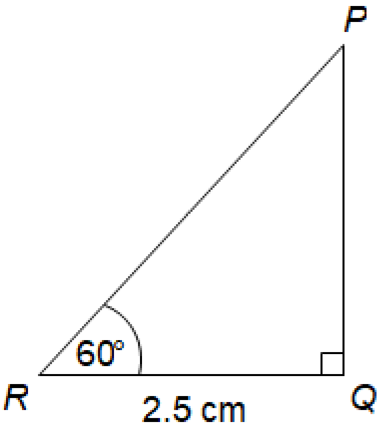


3. Calculate length AB



(not drawn to scale)

4. Calculate length PR

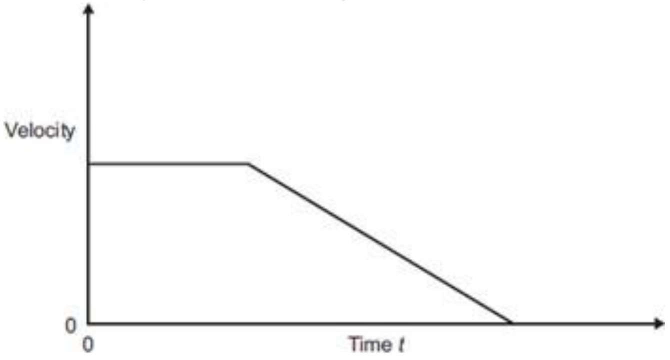


(not drawn accurately)

GRADIENTS & AREAS

1. A car is moving along a road. The driver sees an obstacle in the road at time $t = 0$ and applies the brakes until the car stops.

   The graph shows how the velocity of the car changes with time.

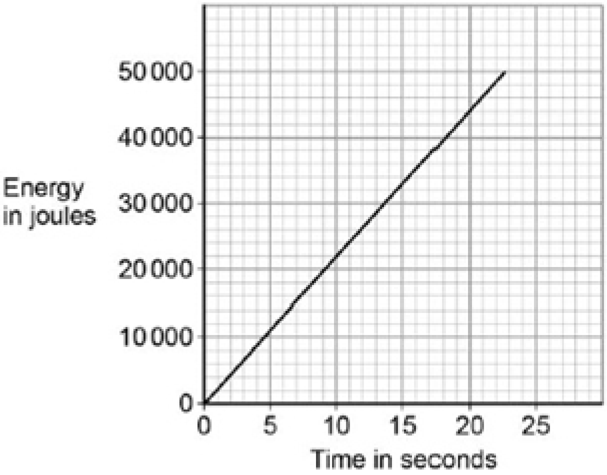


From the list below, which letter represents:

- the negative acceleration of the car
- the distance travelled by the car?

a. The area under the graph
b. The gradient of the sloping line
c. The intercept on the y axis

2. The graph shows how the amount of energy transferred by a kettle varies with time.



The power output of the kettle is given by the gradient of the graph.

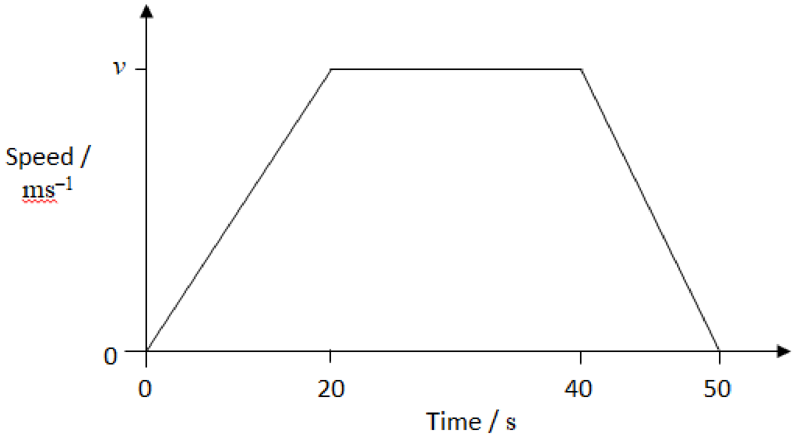
Calculate the power output of the kettle.

3. The graph shows the speed of a car between two sets of traffic lights.

It achieves a maximum speed of *v* metres. per second. It travels

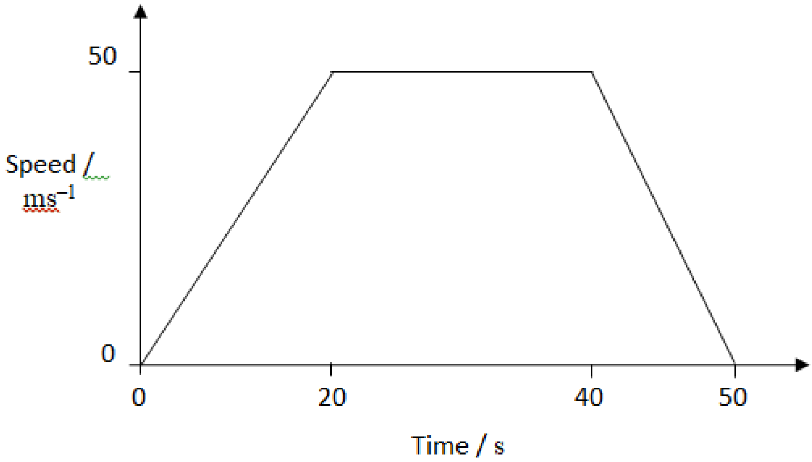
for 50 seconds.

The distance between the traffic lights is 625 metres.



Calculate the value of v.

4. The graph shows the speed of a train between two stations.



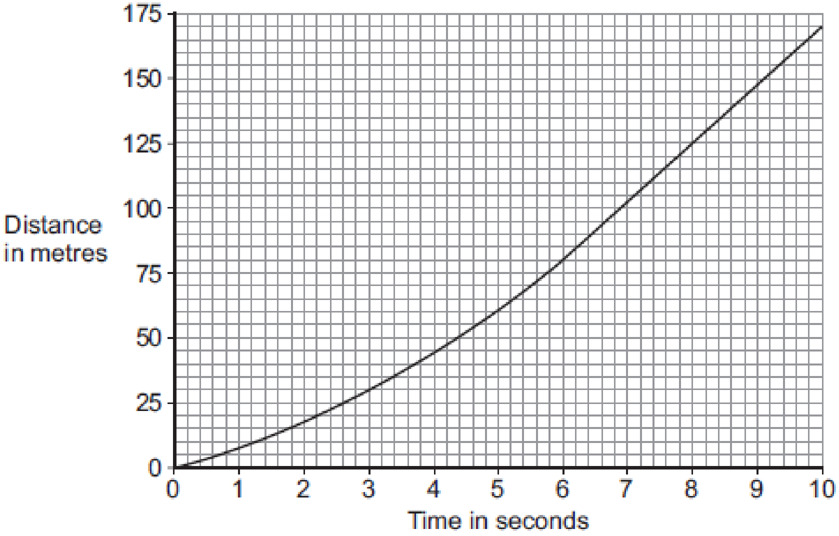
(not drawn accurately)

Calculate the distance between the stations.

USING & INTERPRETING DATA

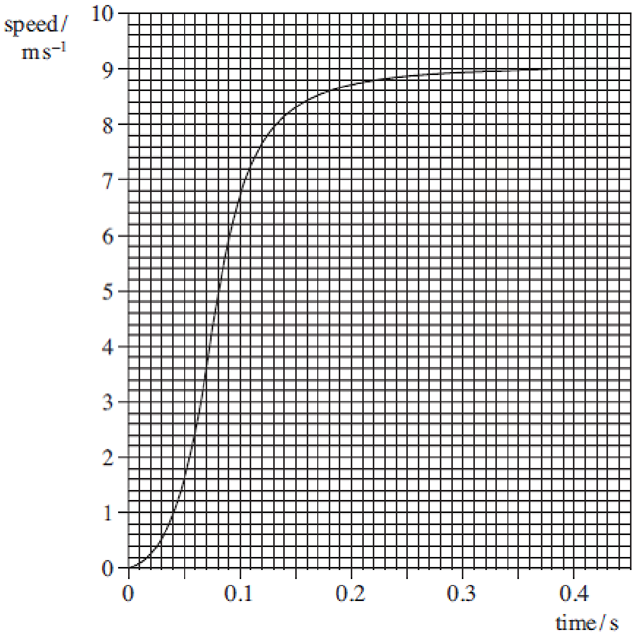
1. The graph shows the motion of a car in the first 10 seconds of its journey.

Figure 1



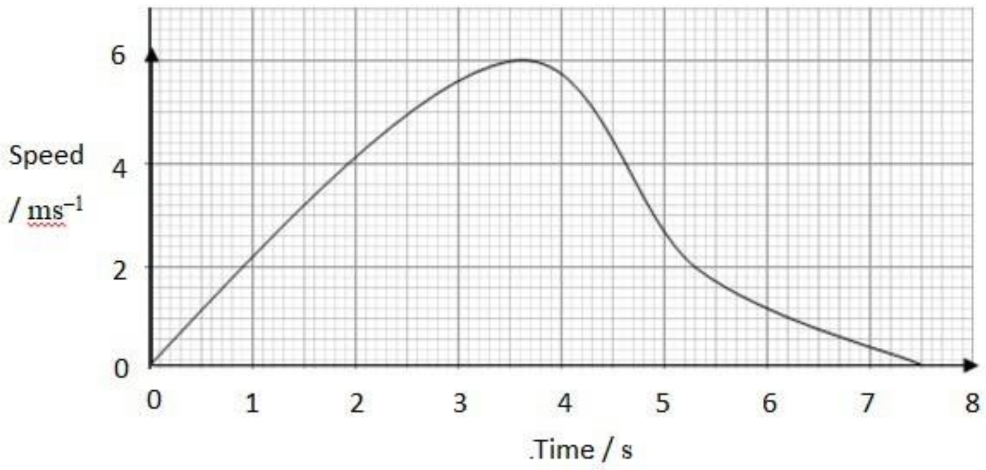
Use the graph to calculate the maximum speed the car was travelling at.

2. The figure below is a speed–time graph for a sprinter at the start of a race.



Determine the distance covered by the sprinter in the first 0.3 s of the race.

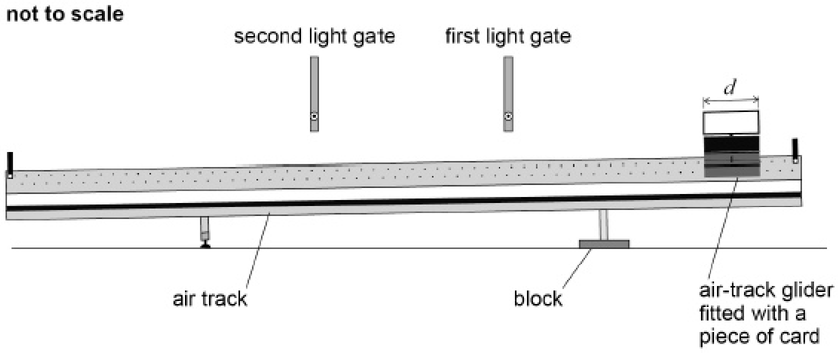
3. The graph shows the speed–time graph of a car.



Use the graph to determine:

a. the maximum speed of the car
b. the total distance travelled
c. the average speed for the journey.

4. The diagram shows the apparatus used by a student to measure the acceleration due to gravity (*g*).



In the experiment:

- a block is used to raise one end of the air track
- an air-track glider is released from rest near the raised end of the air track and passes through the first light gate and then through the second light gate
- a piece of card of length d fitted to the air-track glider interrupts a light beam as the air-track glider passes through each light gate
- a data logger records the time taken by the piece of card to pass through each light gate and also the time for the piece of card to travel from one light gate to the other.

a. The table gives measurements recorded by the data logger.

Time to pass through first light gate / s	Time to pass through second light gate / s	Time to travel from first to second light gate / s
0.50	0.40	1.19

The length d of the piece of card is 10.0 cm.

Assume there is negligible change in velocity while the air-track glider passes through a light gate.

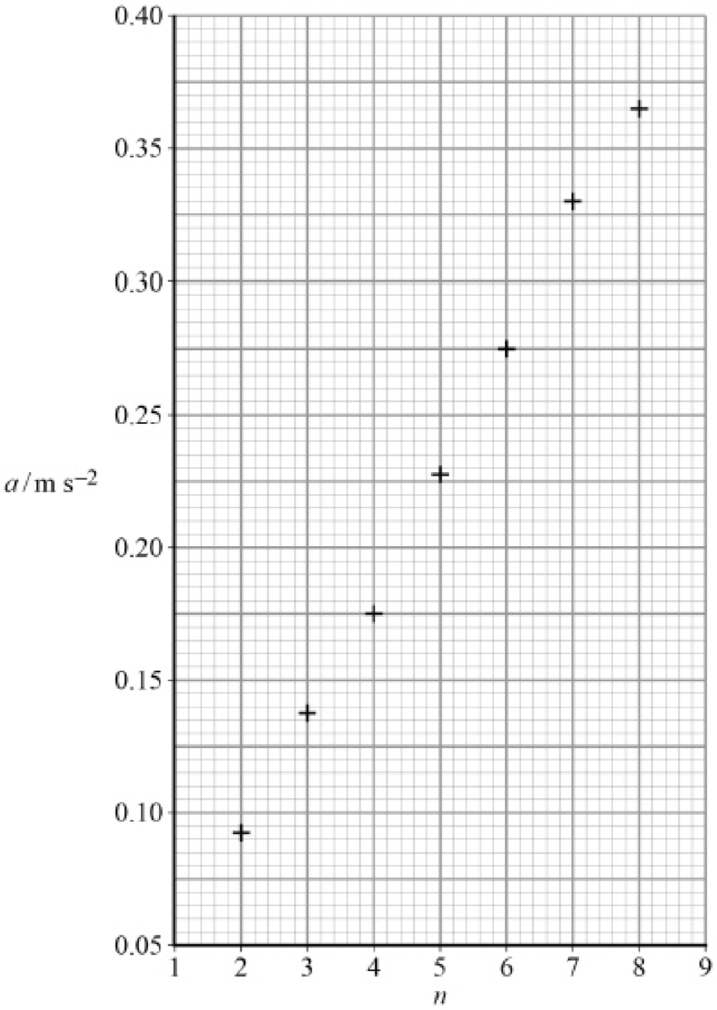
Determine the acceleration a of the air-track glider.

b. Additional values for the acceleration of the air-track glider are obtained by further raising the end of the air track by using a stack consisting of identical blocks.

Adding each block to the stack raises the end of the air track by the same distance.

Below is a graph of these results showing how a varies with n, the number of blocks in the stack

Below is a graph of these results showing how *a* varies with *n*, the number of blocks in the stack.



Draw a line of best fit and then determine the gradient of your line (A).

c. It can be shown that, for the apparatus used by the student, g is equal to $\frac{2A}{h}$ where h is the thickness of each block used in the experiment.

The student obtains a value for g of 9.8 $\mathrm{m/s^2}$

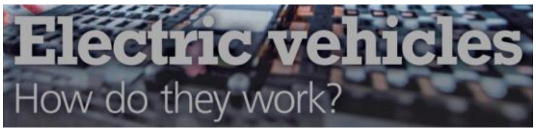
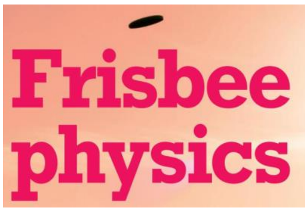
Calculate h

Strongly recommended summer learning

As a student at New College there is an expectation that you read around your topics and preview upcoming topics before lessons.

When you join New College you will have access to the excellent Physics Review articles via our library as well as digital issues. These articles are tailored towards A-level students and cover a broad range of topics. These provide more depth and context to your learning.

A selection of articles have been chosen for you as you transition from year 11 to year 12. They can be found at the end of this document.

Task: Read 3 of the articles and answer the following questions:

1. Which article did you find most interesting?

2. What did you find interesting about the article?

3. Did it link with your GCSE topics? If so, which topics?

4. Which topics that you have read about have inspired you to read further?

There are some excellent resources for more general interest physics; here are a few to get you started:

Infinite Monkey Cage (podcast)	**Kurzgesagt** (YouTube)	**The Bomb** (podcast)	**Veritasium** (YouTube)

Pointing a space telescope

Peter Main

A space telescope's pointing is accurate to a few millionths of a degree, and it must remain steady for long periods. Peter Main explains how this is achieved

Hubble image of the Messier 100 spiral galaxy at a distance of 55 million light years

EXAM LINKS

The terms in bold link to topics in the **AQA**, **Edexcel**, **OCR**, **WJEC** and **CCEA** A-level specifications, as well as the **IB**, **Pre-U** and **SQA** exam specifications.

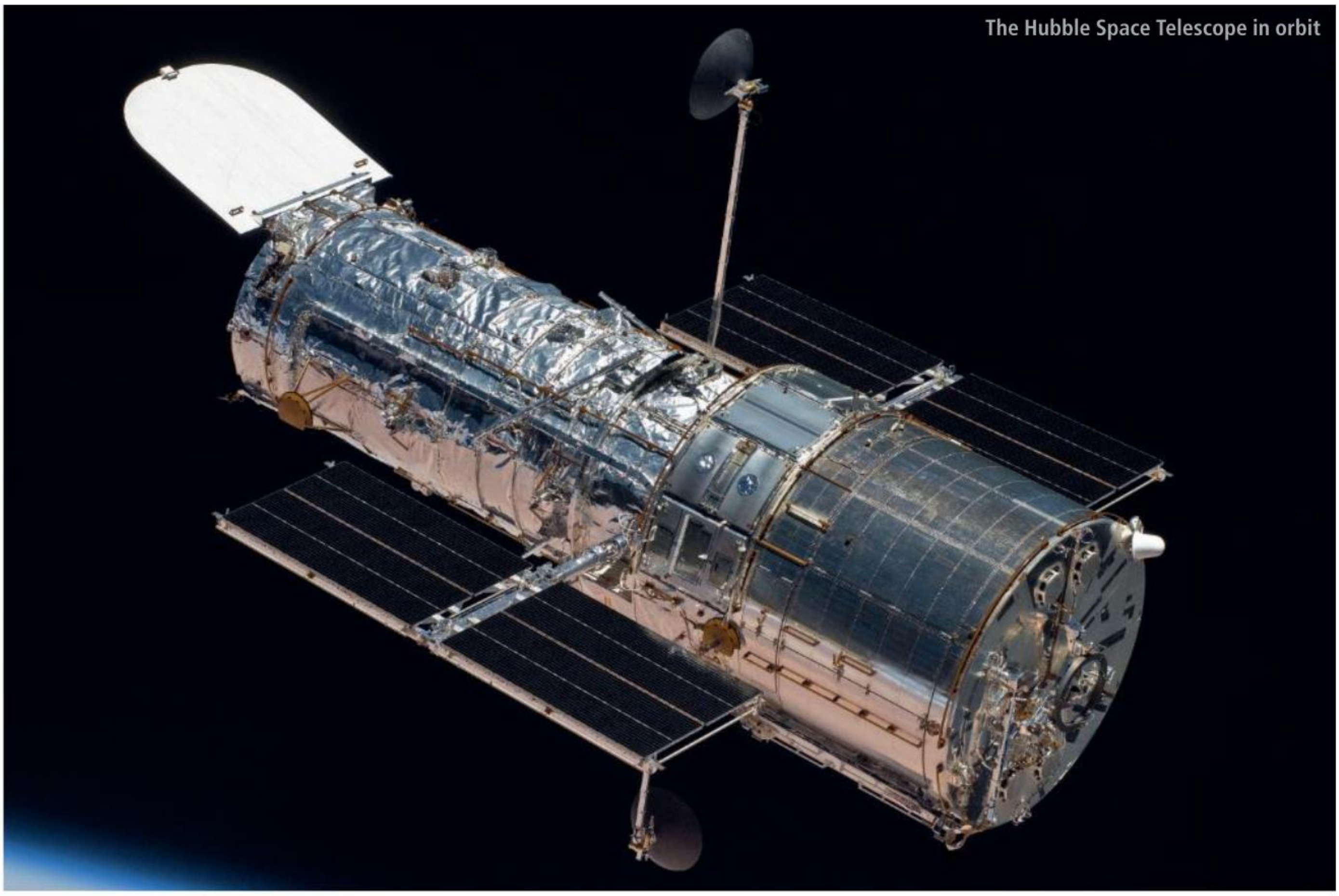
The Hubble Space Telescope **senses** its orientation with a **magnetic** compass and a star map, and uses the **conservation of angular momentum** to control its pointing with high **precision**.

You may be familiar with the stunning pictures received from the Hubble Space Telescope. Hubble is a technological wonder of the modern world. Over the last three decades its pictures and the spectroscopy that comes with them have revolutionised almost every aspect of astronomy. We now look forward to the soon-to-be-launched James Webb Space Telescope, which is designed to fulfil a related and complementary role. It promises to give us even more spectacular results.

The James Webb and Hubble telescopes require long exposures, during which they must remain precisely oriented. Not only that, but each telescope, with a mass of several tonnes, must be pointed in the right direction and be able to return, if necessary, to precisely the same tiny part of the sky. Have you ever wondered how that is done? This can be explained using the example of the Hubble telescope.

Precision

There are two parts to the orientation of the Hubble telescope — the sensors and the actuators. The sensors detect the present orientation, and the actuators move the telescope to point it in the right direction. When the actuators have done their job,



The Hubble Space Telescope in orbit

PhysicsReview*Extras*

Get practice-for-exam questions based on this article at

astronomers need to know the accuracy of the orientation. They have nothing to worry about because the pointing of the telescope is incredibly accurate. If the telescope were in York, it could aim at a point in London and reliably get within 1 cm of the target. Putting it another way, that is within 0.000 002° of the correct direction. In addition, the orientation is so stable that it can easily be held steady for at least 24 hours.

Sensors

There are several types of sensor that make up the pointing control system. These are the Sun sensors, the gyroscopes, the magnetic sensors, the star trackers and the fine guidance system. They each work over a different range of angular adjustment to bring the telescope from a random orientation to one that is precisely calculated (Figure 1).

The Sun sensors give a crude indication of orientation, but they are most important for the protection of the telescope. Hubble's optical system is extremely sensitive, so the telescope must always be pointed at least 50° away from the Sun. In addition, the Sun sensors indicate the optimum orientation of the solar panels that use the Sun's radiation to generate electricity.

Hubble's gyroscopes are among the most accurate and stable ever built (Box 1). They keep a very sensitive watch over any changes in orientation, measuring the direction and rate of rotation. The spacecraft has a complement of six gyros, of which three are active and three are kept as spares. Being mechanical devices, they are subject to wear and eventual failure. They have all been replaced during service visits by the Space Shuttle. However, during Hubble's lifetime, improvements in manufacture have increased their lifespan.

Just as a compass needle on Earth's surface points towards magnetic north, the magnetic sensing system acts as Hubble's compass. Instead of using a magnetised needle, Hubble's system

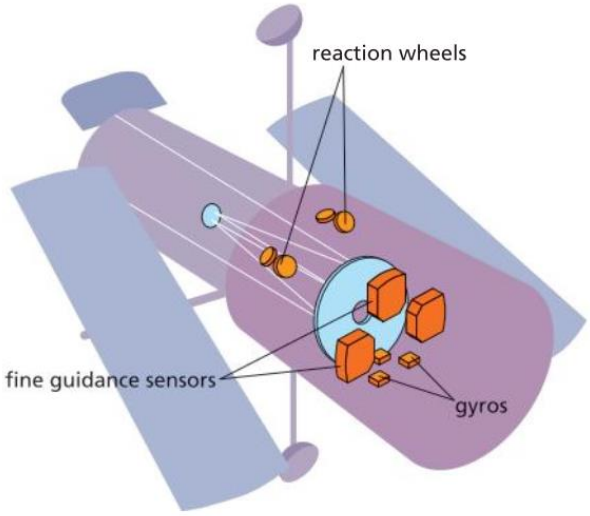


Figure 1 The positions of Hubble's sensors

Box 1 Angular momentum and gyroscopes

In linear motion, momentum is defined as:

momentum = mass × velocity

There is an equivalent definition of angular momentum in circular motion. The equivalent of mass in rotational motion is moment of inertia, so the definition of angular momentum is:

angular momentum = moment of inertia × angular velocity

The simplest example of moment of inertia, I, is of a point mass m a distance r from the axis of rotation:

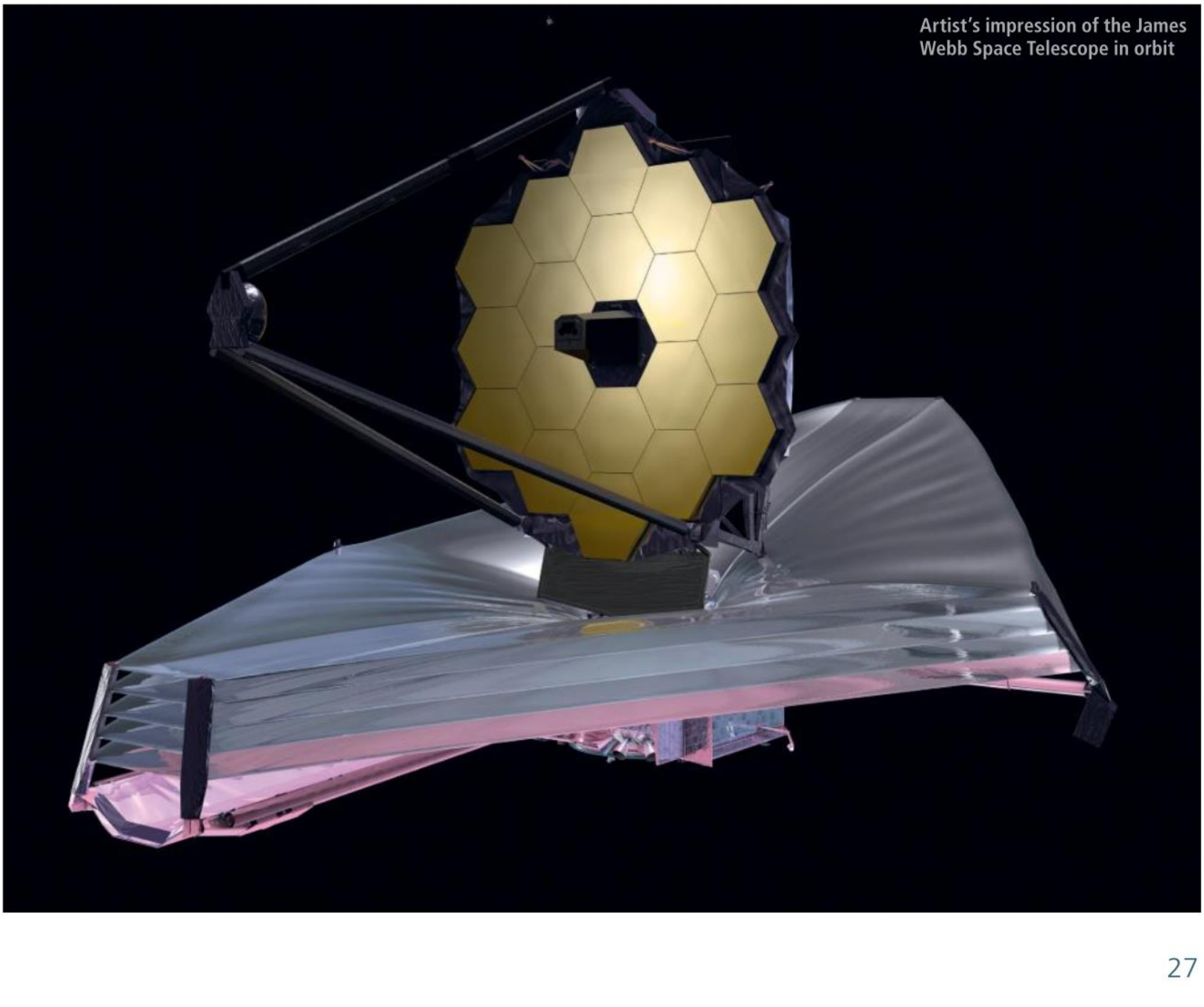
$$I = mr^2$$

Angular momentum is an important quantity in physics because, like linear momentum, it is conserved. This property is exploited in a gyroscope. A simple gyroscope consists of a spinning wheel or disc mounted on an axle. Because it is spinning, it possesses angular momentum, which is a vector quantity with the vector pointing along the axis of rotation. The conservation of angular momentum therefore requires that both the magnitude and direction of the vector remain constant. In other words, the direction of the axis of rotation is fixed in space unless operated upon by an external force.

uses two magnetometers that measure magnetic field strength and direction. Together with some clever electronics, they serve to point the telescope in the correct direction with an error of no more than 0.6°.

Star trackers are available to amateur astronomers, enabling them to see ever fainter objects. However, the fixed-head star trackers on the Hubble telescope take the technique to a whole new level. They determine Hubble's orientation by measuring the locations and magnitudes (brightnesses) of stars in their field of view. Hubble's onboard computer has a vast catalogue of 15 million star positions, together with their magnitudes, which can be matched to the stars seen by the star trackers. This information increases the accuracy of Hubble's orientation to within 0.017°.

One of the most remarkable pieces of Hubble's equipment is the fine guidance system. It is Hubble's most accurate pointing sensor, made up of three fine guidance sensors (FGSs), each the size of a small piano. Only two are required for the operation of the telescope — the third acts as a spare. However, it is normally assigned the secondary task of determining star positions with high accuracy — an important tool for locating extra-solar planets and determining distances to stars. The FGSs find and maintain a lock on two guide stars, one for each FGS. The pointing parameters used are updated 40 times per second to


Artist's impression of the James Webb Space Telescope in orbit

Hubble images of: (top) a giant cluster of thousands of stars called Westerlund 2, measuring between 6 and 13 light-years across; (bottom left) galaxy NGC 2442, nicknamed the Meathook Galaxy owing to its extremely asymmetrical and irregular shape; and (bottom right) the Great Barred Spiral Galaxy, NGC 1365, at around 60 million light-years from Earth

During Servicing Mission 3B in March 2002, astronauts Michael Massimino (left) and James Newman (right) replaced one of Hubble's four reaction wheel assemblies, which help turn the spacecraft

ensure that the spacecraft's orientation does not change. Before a more powerful computer was installed during one of the maintenance visits, this took a large fraction of the available computing power.

If 40 updates per second sounds like overkill, the designers of Hubble would tell you otherwise. The orbit is at a height of 547 km. Even at this altitude there is enough air to disturb Hubble's orientation by a measurable amount. Constant adjustment for air resistance is therefore a necessity. The pressure of the Sun's radiation is also a disturbing factor, which requires constant adjustment of orientation. The level of stability and precision that the FGSs provide gives Hubble the ability to remain pointed at a target with no more than a 0.000 002° deviation over extended periods of time.

Actuators

There are two systems that rotate and physically point the telescope — these are the reaction wheels and magnetic torquers. Neither uses thrusters because the resulting exhaust causes a fog that could interfere with the field of view, and may coat the optics. Also, they require propellant, which is a finite resource.

Let us look at the four reaction wheels first. Each one has a mass of 40 kg with a diameter of 60 cm, and spins in its own unique direction, driven by an electric motor. The wheels rotate the telescope about its centre of mass using the conservation of angular momentum (Box 1). If the spin of a reaction wheel decreases, its angular momentum decreases. Therefore, by conservation of angular momentum, the telescope will rotate about the same axis but in the opposite sense to compensate. Similarly, if the wheel speeds up, the telescope's rotation will be in the reverse direction.

Since the rotation axes of the four reaction wheels point in different directions, Hubble can use combinations of them to point itself at any location in the sky. Only three wheels are necessary for this, but the extra wheel provides a degree of fault tolerance. The spin speed of the reaction wheels is so closely controlled that they can keep the orientation of the telescope to within the required 0.000 002°. Since there is rarely any need to change direction quickly, the maximum rate at which Hubble rotates is approximately 90 degrees in 15 minutes — about the speed of the minute hand of a clock.

The second way to adjust the orientation of the Hubble telescope is by means of the four magnetic torquer bars, all oriented in different directions. Each consists of an eight-foot iron rod wrapped in wire. An electric current flowing through the wire creates a magnetic field, as from a bar magnet. This makes it behave like a giant compass needle, so it produces a torque (a rotational force) in a direction that would align it with Earth's field. The strength of the torque varies with the current through the wire, which is under computer control. The atmospheric drag on Hubble tends to increase the reaction wheel speeds to maintain orientation. The torquer bars are mainly used to reduce these speeds. However, in the event of reaction wheel failure, the torquers can be used, together with two wheels, to orient the telescope without loss of accuracy.

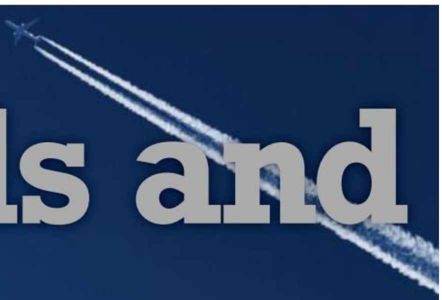
RESOURCE

A video on the orientation of the Hubble telescope can be found at:

Vibrating solar panels

A problem with the accuracy of the telescope pointing system was discovered shortly after launch in 1990. Because of its low orbit, Hubble alternates between being in full sunlight and passing through Earth's shadow. The temperature difference on the solar panels is 45°C, with the transition occurring in less than a minute each time. The consequent expansion and contraction set the panels vibrating, which affected the main telescope. The pointing control system was not set up to deal with this kind of movement and about 10 minutes of observing time was lost on each orbit until the vibration died down. The vibration was tiny but was enough to degrade the pointing precision by a factor of 10. Rewriting the control software was complicated and, when it was implemented, it took up nearly all the available computing power. A permanent solution to the problem was found by redesigning the solar panels and replacing them on the first service visit by the Space Shuttle in 1993. There would always be unforeseen problems to deal with, but it was a triumph of design that this could initially be overcome remotely from the ground.

Peter Main is on the academic staff of the University of York and is also a member of the Physics Review editorial board.



Vapour trails and

Ron Holt

Recent research indicates that aircraft vapour trails can have a significant impact on climate change. Ron Holt explains how vapour trails are formed and how they contribute to global warming

EXAM LINKS

The terms in bold link to topics in the **AQA**, **Edexcel**, **OCR**, **WJEC** and **CCEA** A-level specifications, as well as the **IB**, **Pre-U** and **SQA** exam specifications.

Vapour trails form when water vapour **condenses** to form droplets and then **freezes** to create ice crystals, which may **sublime** to return directly to vapour. A combination of atmospheric **pressure**, **temperature** and **relative humidity** determines whether these processes occur.

On a clear day you can often see numerous white trails behind high-flying aircraft that criss-cross the sky. These white trails are called contrails, from the combination of the words 'condensation' and 'trails'.

Contrails are produced by engine exhaust from aircraft that are cruising at altitudes several miles above the Earth's surface. The combustion of hydrocarbon fuels within the aircraft engine produces a significant amount of water vapour and other gases and particles. The air temperature is very low at high altitudes, which provides the conditions for contrails to form, first as liquid water and subsequently as ice. Depending on atmospheric conditions, contrails can develop into more substantial cirrus-type clouds that can affect the Earth's temperature and climate.

What are contrails?

Contrails are white, linear plumes that are clearly visible behind aircraft that are flying at cruising altitude, where fuel efficiency is best. Contrails form when the gases and particles from the hot jet engine exhaust mix with the low-temperature atmospheric air. The hot exhaust gases include water vapour (which is invisible), which rapidly cools and condenses into water droplets and subsequently freezes to form ice crystals. These are visible as a long, white cloud or contrail (Figure 1).

The time taken for exhaust water vapour (invisible) to cool enough to condense and form an ice contrail (visible) accounts for the short gap that is clearly seen between the aircraft engine and the formation of the contrail. The way a contrail is formed and evolves depends on a number of factors, including the content and composition of the aircraft exhaust gases and, perhaps more importantly, on the aircraft altitude (i.e. atmospheric pressure), atmospheric temperature and the amount of moisture within the atmosphere (i.e. the relative humidity) surrounding the aircraft.

Little was known about contrails during the very early days of aircraft production in the 1920s. They were, however, studied extensively during World War Two and in subsequent years by scientists in the military and meteorologists, because the telling white trails showed opponents where your planes were. The correlations between pressure, temperature and relative humidity and the formation of contrails were depicted on a chart developed by H. Appleman in 1953 (known as the Appleman chart). It has been used extensively to predict the formation of contrails (Box 1) given specific values for pressure, temperature and humidity.

Contrail formation occurs at altitudes typically between 8000 m and 12 000 m (26 000 ft to 40 000 ft), i.e. the cruising altitude of commercial aircraft on long-haul flights, where water vapour within the exhaust plume interacts with the cold ambient air at temperatures typically around −15°C. As the exhaust air cools rapidly in the cold local air, the newly formed water droplets suddenly freeze to form the ice particles that form the white contrail lines seen.

Contrails can also form at lower altitude due to rapid changes in air pressure emanating from curved surfaces, as encountered in take-off and landing and in airplane displays. Such vortices add no water to the air but result from a decrease in pressure and subsequent decrease in temperature until the dew point is reached. However, such visible water droplets quickly evaporate, and the visible vortex disappears.

Types of contrail

The development and lifecycle of contrails depend on the surrounding atmospheric conditions. Short-lived contrails only

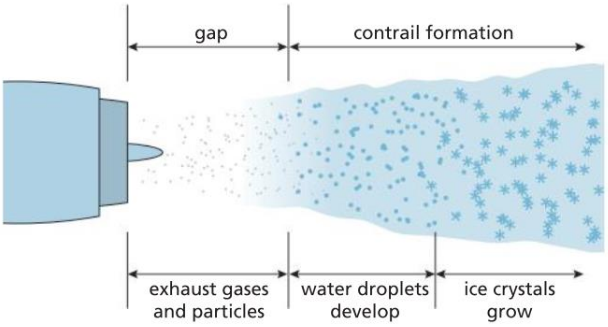


Figure 1 Formation of a contrail

last a few minutes and therefore have relatively short plumes. These are formed under conditions where the air is dry (i.e. low humidity), which signifies that only a small amount of additional water vapour is available from the atmosphere. A short-lived contrail that is formed under these conditions will sublime relatively quickly from ice crystals directly into water vapour, and will consequently fade away quite rapidly.

A contrail whose plume remains linear, visible for longer (at least 10 minutes) and perhaps shows signs of developing into a broader trail is known as a persistent contrail (or persistent non-spreading contrail). Persistent contrails remain visible for long after the aircraft has passed because the atmosphere around the aircraft is more humid and therefore already contains plenty of water vapour, so the ice crystals do not turn to water vapour so quickly. They tend to be more developed, a little broader and perhaps fluffier in character than short-lived contrails.


A typical linear contrail from an aircraft cruising at high altitude, showing vapour trails from all four engines

Box 1 The Appleman chart

The Appleman chart (Figure 1.1) shows the relative humidity needed for contrail formation at a given atmospheric pressure and temperature.

Relative humidity (RH) is a measure of the amount of water vapour in the air. When RH is zero the air contains no water vapour. When RH is 100%, the air is 'saturated' with water vapour and can contain no more.

The blue area represents the typical cruising altitude for commercial aircraft. Persistent contrails tend to form when the relative humidity is between 60% and 70%, as indicated by the darker region.

Most commercial aircraft on long-haul flights tend to fly at between 8 km and 12 km (26 000–40 000 ft) i.e. in the upper troposphere and lower stratosphere, where the air pressure lies between 35 kPa and 20 kPa. This range of cruising altitude increases fuel efficiency via reduced drag, but with still enough oxygen available for efficient combustion.

$$1\,\text{kPa} = 10^3\,\text{Pa} = 10^3\,\text{Nm}^{-2}$$

A commercial aircraft flying at a flight level of 340 (FL340) translates to an atmospheric pressure of 25 kPa. At sea level atmospheric pressure is 1.014×10^5 Pa (≈ 100 kPa). If the external temperature was −56°C, then the Appleman chart would indicate a contrail. If the relative humidity was between 1% and 3% then the contrail would be a short-lived contrail that sublimes and fades away relatively quickly.

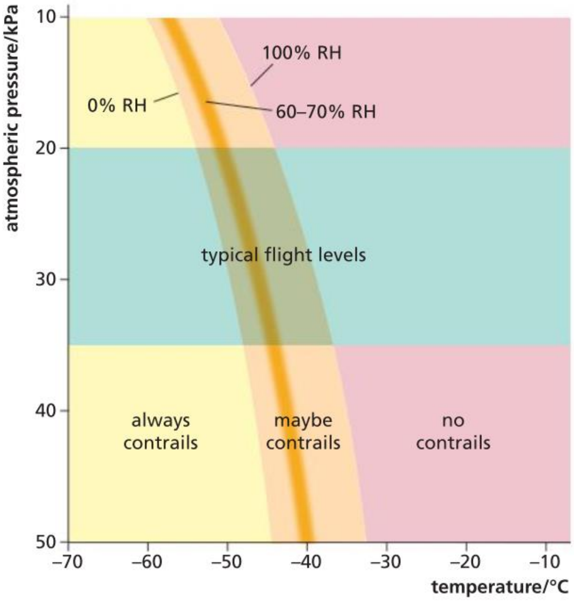


Figure 1.1 The Appleman chart


Aircraft-induced clouds can make a significant contribution to cloud cover

Persistent contrails that remain after many hours show considerably more development and are similar in character and structure to natural cirrus clouds (high, thin and wispy clouds — see 'Clouds', Physics Review Vol. 18, No. 2, pp. 2–6). For this reason, they are generally referred to as contrail cirrus (or persistent spreading contrails). Atmospheric conditions for these to form and evolve include moist air, where the humidity is high (i.e. greater than that needed for ice condensation to occur), so that more water from the air freezes on the ice crystals, causing the contrails to grow. Due to further mixing with the surrounding air enhanced by air turbulence, as well as solar heating, contrail cirrus may form large clouds several tens of square kilometres in area and several hundreds of metres thick. Persistent contrails and contrail cirrus are collectively known as aircraft-induced clouds (AICs). It is these that can cause a change in global cloud cover, affecting the temperature balance and structure in the lower atmosphere.

Aircraft exhaust and contrail formation

Aircraft engines burn hydrocarbon fuels (e.g. kerosene) at high temperatures. The resultant combustion produces exhaust gases that contain a significant amount of water vapour and carbon dioxide (CO_2), as well as small amounts of unburnt hydrocarbon fuel (HC), carbon monoxide (CO), nitrogen oxides (NO_x), sulfur dioxide (SO_2), soot and particulate matter. The essential ingredients for contrail development are water vapour, sulfur gases, particulate matter and soot, as these create the tiny particles (cloud condensation nuclei, CCN) that are needed for water droplet formation. This is in addition to those particles naturally present in the atmosphere at these altitudes.

RESOURCE

An academic research paper on contrails:

Table 1 shows the amounts of gases and particulate matter within a typical jet exhaust that are emitted per tonne of kerosene consumed. A typical commercial aircraft on a long-haul flight of 10 hours uses approximately 75 tonnes of fuel. The figures in Table 1 depend on operating conditions of the engine, aircraft altitude, and atmospheric temperature and humidity.

Contrails and global warming

Recent studies using both observation and modelling are beginning to highlight the role that persistent contrails and contrail cirrus have in terms of global warming. Natural cirrus clouds and aircraft-induced clouds (AICs) are all high-level clouds composed of ice crystals, and they all form and evolve in ice-supersaturated regions of the atmosphere, which are sufficiently cold and moist.

Table 1 Exhaust products per tonne of kerosene fuel burned

Process	Product intake/exhausted	Mass/kg
Intake	Kerosene fuel	1000
	Cold air intake	315000
Exhaust	Cold air exhausted	264000
	Hot air exhausted	47600
	Carbon dioxide (CO_2)	3200
	Water vapour (H_2O)	1200
	Nitrogen oxides (NO_x)	11.0
	Sulfur dioxide (SO_2)	1.0
	Carbon monoxide (CO)	0.8
	Hydrocarbons (HC)	0.2
	Particulate matter and soot	0.04

A change in global cloud cover due to an increase in AICs creates an imbalance between the short-wave solar radiation from the Sun reaching the Earth's surface and the long-wave radiation emitted from the Earth's surface reaching space. This imbalance, known as radiative forcing (RF), is critical to the overall effects of global warming. In drier or warmer air (ice-subsaturated regions) contrails may still form but these tend to be of the short-lived variety, and because of their small area coverage and lifespan have little or no impact on radiative forcing and hence on global warming.

The effects and impact of AICs have been highlighted in a recent research article (see Resource box), which predicts radiative forcing to increase by a factor of three from 2006 to 2050. This is the result of an anticipated increase in air traffic volume over all main air traffic areas in Europe and North America, and an even larger increase over many areas in eastern Asia.

Whilst contrails pose a threat in terms of global warming, it should not be forgotten that aircraft emissions also include CO_2, hydrocarbons, CO, nitrogen oxides, sulfur oxides and black carbon, all of which interact with themselves and with the atmosphere around them. These emissions from the combustion of jet fuel lead to an inevitable increase in greenhouse gases, which warm the lower atmosphere and the Earth's surface. Such emissions have increased substantially over the past 20 years.

Future changes

Globally there are in excess of 100 000 aircraft flights every day, with the majority of these leading to the formation of contrails. Currently this is set to rise by an average of 5% per year. Economic predictions suggest that aviation fuel usage, and therefore CO_2 and water vapour emissions, may double by 2050. Other projections suggest increases by a factor of three or four. Whatever the outcome, the impact of AICs will be substantial.

There may be solutions to these problems, but these will undoubtedly incur a cost. For example, aircraft on long-haul flights could cruise at lower altitudes to eliminate contrail formation, but this would lead to an increase in exhaust contaminates as well as increased CO_2 emissions. Aircraft could perhaps be rerouted to avoid regions of high humidity and ice supersaturation, but again the additional rerouting distances will lead to increases in CO_2 emissions. Perhaps aircraft engine technology may be developed and advanced to improve engine efficiency and reduce the water vapour and contaminants (e.g. sulfur) emitted in the exhaust gas. This may be aided by using biofuels, which produce fewer soot particles than standard jet fuel.

All these scenarios are currently being looked at because reducing aviation's contribution to global warming is clearly of paramount importance. The potential magnitude of the problem and the growing impact that aviation is having on our global climate, as briefly described here, have clearly highlighted the need to further understand the complex problems associated with the formation and evolution of contrails.

Ron Holt is a physicist, teacher and author.

AT A GLANCE

Nanotechnology

Nanotechnology is the study and manipulation of objects in the approximate size range 1–100 nm (1 nm = 1×10^{-9} m).

Electron microscopes are important instruments for nanotechnologists. Both particle and wave models are needed to explain how they work. In a scanning transmission electron microscope (1), an electron gun (2) accelerates electrons to high energy. The electron beam is steered using magnetic lenses (3) and focused to a fine spot on the object being studied, which absorbs and scatters electrons. Sensors detect the intensity of the beam transmitted through the object, and the spot is scanned in a grid pattern to build up an image (4).

A microscope's resolution (the smallest detail it can distinguish) is limited by instrumental effects and by the radiation that it uses (5). The newest STEM instruments can achieve a resolution of about 0.05 nm.

1 A scanning transmission electron microscope (STEM)

2 Electron gun

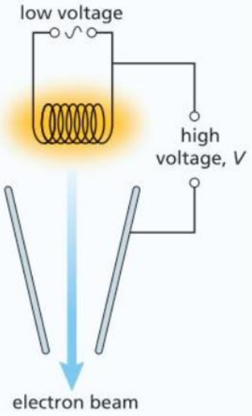
Electrons accelerated through a potential difference, V, gain kinetic energy E_k:

$$E_k = eV$$

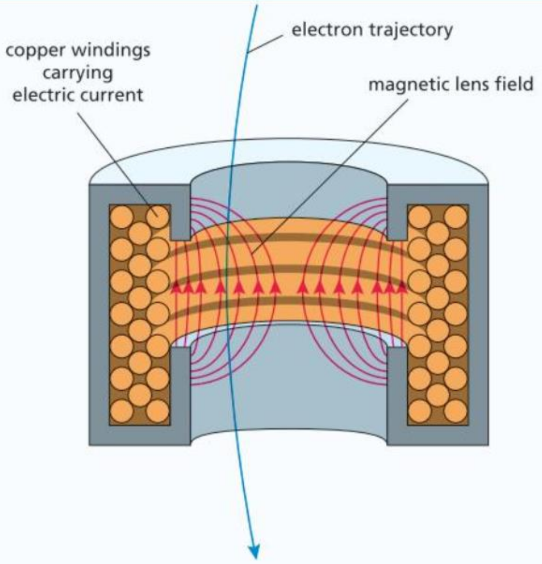
where the electron charge $e = 1.60 \times 10^{-19}$ C.

1 eV (electronvolt) is the energy transferred to an electron that is accelerated through 1 V.
1 eV = 1.60×10^{-19} J.

STEM instruments typically accelerate electrons through about 100 kV, giving them kinetic energies of about 100 keV.



3 Magnetic lens



An electron moving with speed v in a magnetic field B experiences a force F at right angles to its direction of motion and to the field:

$$F = Bev\sin\theta$$

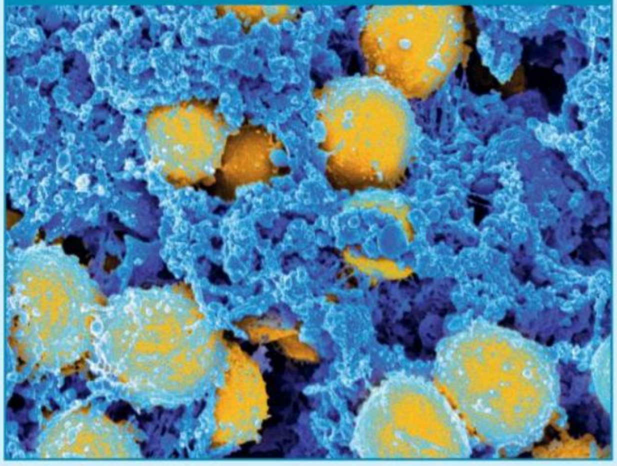
where θ is the angle between the field and the electron's motion. This force steers the electrons in a curved path.

Current nanotechnology research is wide-ranging, with STEM images now giving us an unprecedented view of the world on an atomic scale (6).

The latest electron microscopy technique, cryo-EM, involves flash-freezing a solution of biological molecules to reduce their random thermal motion. The resulting STEM images allow scientists to see the 3D shapes of protein molecules (7), leading to a better understanding of how malfunction can be targeted with drugs.

Atomic and close-to-atomic scale manufacturing (ACSM) promises to have applications in such diverse fields as quantum computing and nanomedicine. Perhaps one day nanobots will even be able to hunt and destroy viruses (8).

4 SEM image of *Staphylococcus aureus* bacteria (yellow) on skin. Each bacterium has a diameter of around 500 nm



5 Electron wavelength

The electrons form a beam of radiation with wavelength λ, which is their de Broglie wavelength:

$$\lambda = \frac{h}{p}$$

where p is momentum, and h is the Planck constant:
$h = 6.63 \times 10^{-34}\,\text{J s}$

For an electron with mass m ($9.11 \times 10^{-31}\,\text{kg}$) moving at a speed v that is much less than the speed of light ($c = 3.00 \times 10^{8}\,\text{m s}^{-1}$):

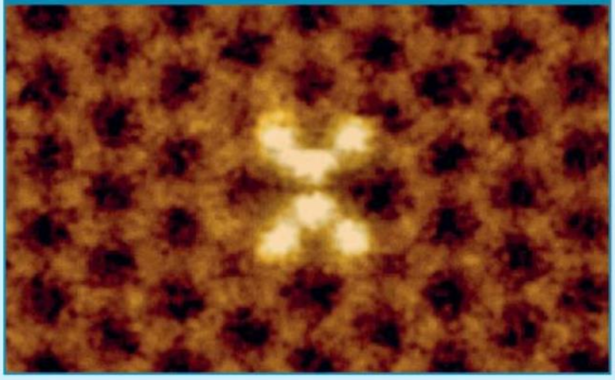
$$p = mv$$
$$= \sqrt{2mE_k}$$
$$\lambda = \frac{h}{2mE_k}$$

A 100 keV electron has $\lambda \approx 4 \times 10^{-12}\,\text{m}$.

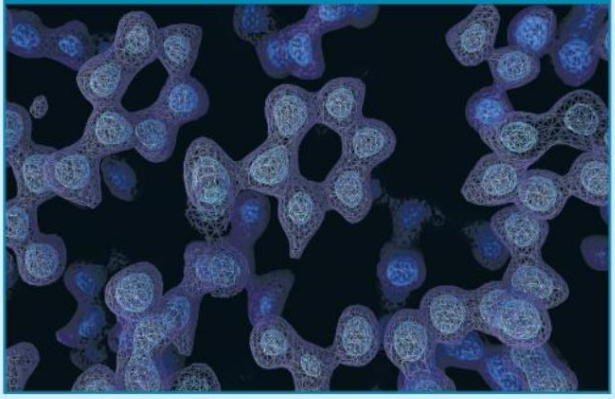
PhysicsReview*Extras*

Download this poster and get practice-for-exam questions at [redacted]

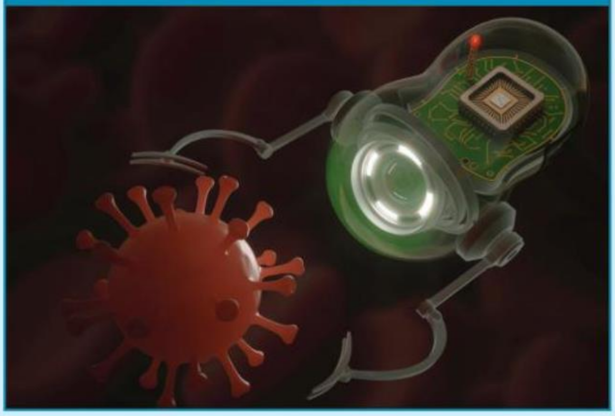
6 Silicon atoms (yellow) in a graphene sheet. The distance between atomic centres is about 0.142 nm



7 A cryo-EM image of the protein apoferritin, provided by Paul Emsley, MRC Laboratory of Molecular Biology



8 Artist's impression of a nanobot capturing a virus with a diameter of about 100 nm



Structural colour

Mike Follows

We perceive colour all around us, and animals use it to attract mates or warn off predators. Mike Follows explains how colour can be produced without chemical pigments

EXAM LINKS

The terms in bold link to topics in the **AQA**, **Edexcel**, **OCR**, **WJEC** and **CCEA** A-level specifications, as well as the **IB**, **Pre-U** and **SQA** exam specifications.

Structural colours are produced by **superposition** and **interference** when **light waves** undergo multiple **reflection.**

Colours are associated with different wavelengths of visible light. Some colours are due to *pigments* that absorb light at particular wavelengths. For example, a substance impregnated with a pigment that absorbs blue light will look yellow when illuminated with white light, because white minus blue gives yellow. Bioluminescent creatures such as fireflies, on the other hand, emit light due to a *chemical reaction*.

But most of the colour we see is *structural colour,* produced when light interacts with microscopic particles or microstructures. This interaction can include scattering, dispersion, diffraction and interference. Diffraction gratings, thin films or multilayers (of thin films) and other structures that repeat themselves can give rise to colours that change depending on the viewing angle. This is called *iridescence*. Because there is none of the absorption of light associated with pigments, structural colours can be startlingly vivid.

Structural colour was discovered by Robert Hooke. In his 1665 book *Micrographia* he described the iridescence of a peacock's tail feathers as 'fantastical'. He was astonished that the colours disappeared when he immersed the feathers in water, demonstrating that pigments played no role. Some creatures can change their structural colour. For example, chameleons

PhysicsReview*Extras*

Get practice-for-exam questions based on this article at

Figure 1 A dress made from Morphotex

can change the spacing between guanine crystals in their skin, which changes the wavelength (and colour) of the light reflected.

New developments in materials technology are bringing us structurally coloured fabrics. The dress shown in Figure 1 is made from Morphotex — a fabric inspired by the morpho butterfly. The colour is produced without chemical dye, and changes subtly with viewing angle.

Structural colours all around us

Scattering of light by particles that are much smaller than the wavelength of light is called Rayleigh scattering (p. 34). The sky is blue because of Rayleigh scattering by air molecules (e.g. nitrogen and oxygen), which scatter blue light more strongly than red. When the Sun is overhead, sunlight passes through the 10 km thickness of the atmosphere. When the Sun is close to the horizon, sunlight has a longer journey through the atmosphere, so more blue light is scattered out of our line of sight to the Sun, which leaves it looking red.

Scattering by particles with sizes comparable to the wavelength of light is known as Mie scattering. Mie scattering by water droplets accounts for the white or grey appearance of clouds. The Lycurgus Cup, a Roman artefact dating from around 400 AD (Figure 2), is another example of structural colour due to Mie scattering from randomly distributed particles. The glass of the cup is dichroic — it is an opaque jade green when it reflects light and a translucent ruby colour when it transmits light. X-ray analysis shows that the light scattering is due to 50–100 μm diameter silver-gold alloy nanoparticles, with the precious metals present in concentrations of around 100 parts per million (ppm). Some Venetian glass gives a brilliant red colour due to scattering from dispersed gold nanoparticles.

Rainbows form when raindrops disperse and reflect light. Dispersion — the separation of white light into its component colours — occurs because higher-frequency light (the blue end of the visible spectrum) is refracted more than low-frequency (red) light. As shown in Figure 3, the light entering the raindrop is dispersed and totally internally reflected at least once off the back surface of the raindrop. It is then refracted again as it leaves the raindrop.

Soap bubbles

The colours of a soap bubble arise because light is reflected from both sides of a thin film. The colour depends on the angle from which a surface is seen.

Imagine a light ray striking the film of soap at normal incidence (i.e. at right angles to the surface) so that both the incident ray and reflected rays are normal to the soap film. Some light reflects off the outer (or top) surface of the film, retracing the path of the incident ray. Of the light transmitted, some reflects off the inner (bottom) surface of the film.

Figure 2 The Lycurgus Cup **(a)** when it reflects light and **(b)** when it transmits light

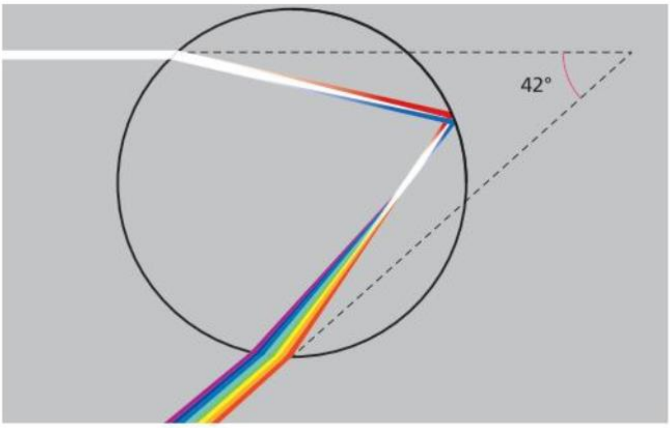


Figure 3 Light that is dispersed and totally internally reflected in a raindrop contributes to a rainbow

Figure 4 shows this with not-quite normal incidence, making it easier to see what is going on.

The two reflected rays interfere, as discussed in Box 1. The transmitted ray has travelled an extra distance corresponding to twice the thickness of the film. Ignoring the role of phase shift for a moment, the two reflected rays will constructively interfere when their path difference is an integer (whole) number of wavelengths, so that a crest from one reflected ray coincides with a crest from the other reflected ray:

$$m\lambda = 2nt \tag{1}$$

where m is an integer, λ is the wavelength, n is the refractive index of the soap film and t is its thickness. Equation 1 shows that the wavelength (and therefore colour) of the reflected light depends on the thickness of the film.

Box 1 Interference

Interference results from the superposition of waves, when two or more waves meet at a given point in space. The displacement at that point is the *algebraic* sum of the individual displacements due to each wave.

Imagine two waves with identical amplitude and frequency. In Figure 1.1a the crest of one wave (solid line) meets the trough from the second wave (dotted line) and we have destructive interference: the sum of a crest and a trough is zero and the two waves are cancelled at that point. In Figure 1.1b the crest from one wave (solid line) meets a crest from a second wave (dotted line). The resultant wave has double the amplitude and we have constructive interference.

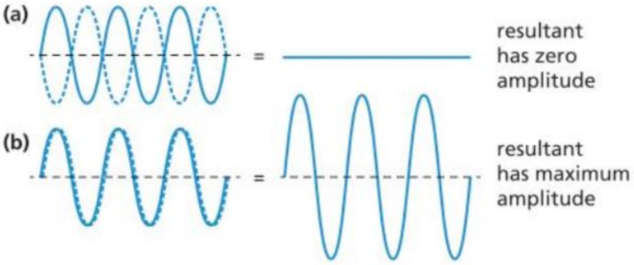


Figure 1.1 **(a)** Destructive interference. **(b)** Constructive interference

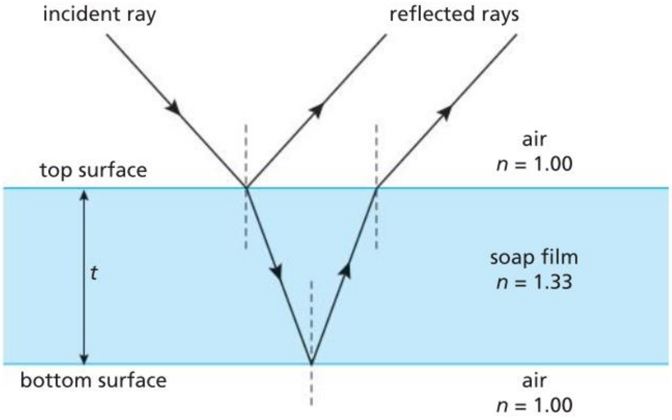


Figure 4 A ray of light reflected off the top and bottom surfaces of a soap film

A slight complication is that the light reflecting off the top surface of the soap film undergoes a 180° phase shift (so that a crest becomes a trough and vice versa). This happens when any wave encounters a medium of higher refractive index, so the condition for constructive interference requires a slight modification to Equation 1, though the essential physics remains the same:

$$(m - \tfrac{1}{2})\lambda = 2nt \tag{2}$$

Increasing the thickness increases the wavelength that produces constructive interference, so the colour will change from, say, green to yellow.

Multilayers

Multiple thin films are the most common method for producing vivid, metallic colours in biological systems. The wing casings of many beetles are a good example. Increasing the number of layers leads to more intense colours, but costs more energy because the beetle has to grow more body tissue.

Multilayers can be either narrowband (Figure 5a) or broadband (Figure 5b). For simplicity, refraction at the boundaries between layers is not shown. The layers in narrowband multilayers have


A large soap bubble

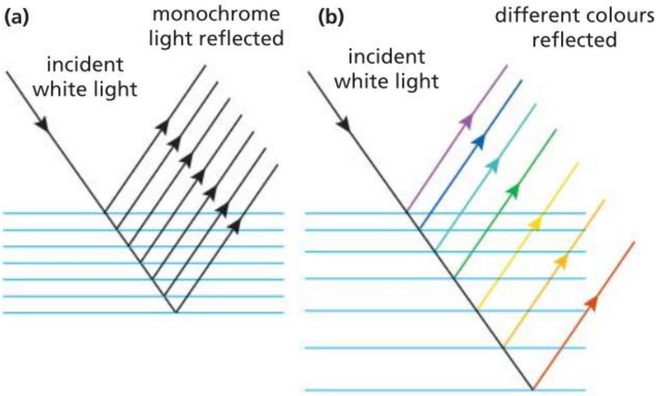


Figure 5 Multilayers: **(a)** narrowband and **(b)** broadband

a uniform thickness and reflect light over a narrow range of wavelengths (and hence colours). If the lowest layer is dark, it absorbs any remaining light, making the reflected colour very vivid. In contrast, the scales of silver-coloured fish have broadband multilayers whose layers vary in thickness. Thick layers reflect red light while thin layers reflect violet, as shown. This means that almost all wavelengths are reflected and, by reflecting virtually all the light in their environment, these fish are much less visible, providing the best possible camouflage in the open ocean, where there is nowhere to hide. Goldfish reflect all colours except the blue end of the spectrum.

Photonic crystals

Sometimes multilayers are referred to as photonic crystals. Natural selection over millions of years has led to photonic crystals with multilayers in one, two and three dimensions, as shown left to right in Figure 6. The diagram shows just two different values of refractive index (represented as two colours) with layers of uniform thickness in all dimensions. Reality is much more complicated and subtle, providing a genuine challenge to researchers attempting to mimic their physical properties, as the following examples illustrate.

Nature's colour tricks: reflecting bowls

Each 100 µm wing scale of the emerald swallowtail butterfly has a honeycomb array of dimples a few micrometres across. These dimples or bowls are lined with a stack of thin films, with eleven layers of protein separated by air, each 75 nm thick. A yellow dot is reflected when light incident along the normal strikes the bottom of each bowl. Two 45° reflections off opposite sides of the bowl result in the retro-reflection of a blue ring, as shown in Figure 7. The result is iridescence, where increasing the viewing angle from 0° to 45° shifts the reflected colour from

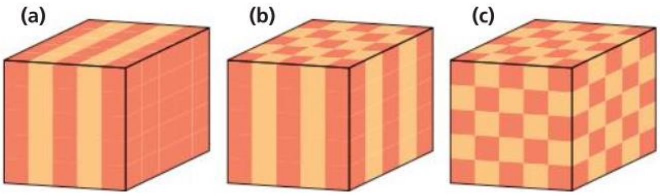
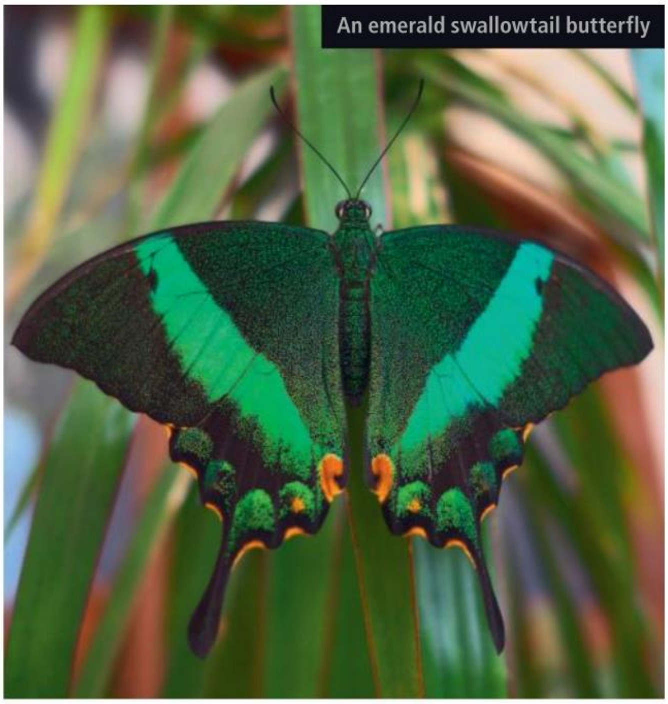


Figure 6 Photonic crystals: **(a)** in one dimension, **(b)** in two dimensions and **(c)** in three dimensions


An emerald swallowtail butterfly

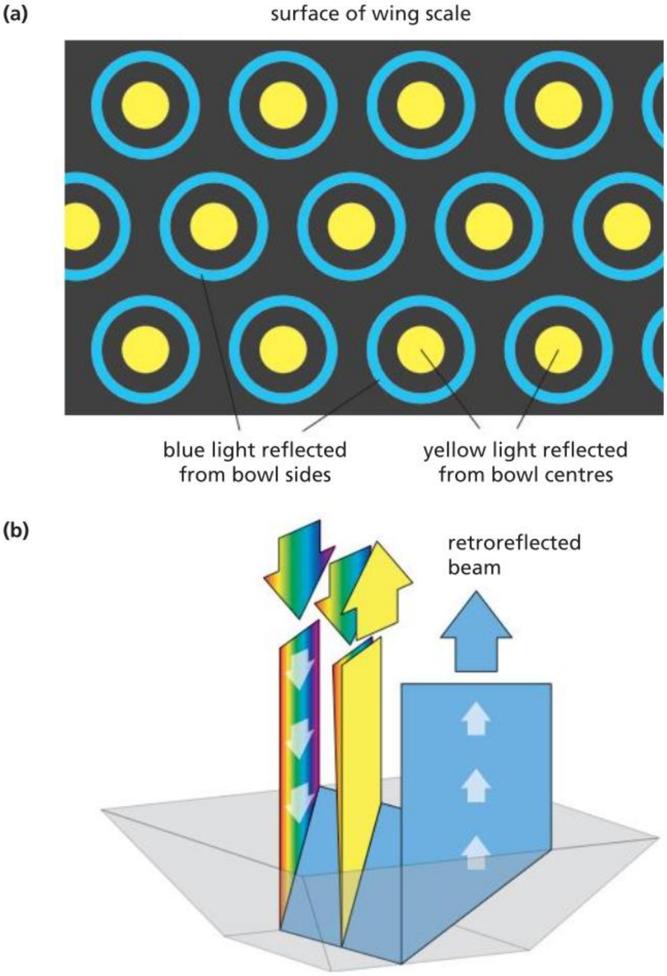


Figure 7 **(a)** The structural colours produced by the honeycomb array on the wing scale of an emerald swallowtail butterfly. **(b)** A ray diagram to show what is happening in an individual bowl


The iridescent dogbane leaf beetle

yellow to blue. Our naked eyes are unable to resolve the spots from the rings, so the colours are mixed and we perceive the wing as green.

Each reflection changes the polarisation of light. Seen through crossed polarising filters, the yellow dot disappears, leaving the blue ring. This is how it works. Light is polarised (using a Polaroid filter) and shone onto the wing. The reflection is viewed through another Polaroid (an analyser) held with its plane of polarisation perpendicular to the first Polaroid. The yellow light cannot pass through the analyser but the blue light can because its plane of polarisation is rotated through 45° both times it is reflected off the side of the bowl, so that its plane of polarisation is parallel to that of the analyser. (See At a glance, Physics Review Vol. 28, No. 4, pp. 16–17.) This would be difficult to counterfeit, so a team at Georgia Institute of Technology in the USA is trying to mimic this structure for use on bank notes.

Morpho butterflies: living jewels

The dorsal (top) side of a male morpho butterfly's wings are an iridescent cobalt blue (Figure 8a) and can be seen from several hundred metres. Closing the wings hides the dazzling display from potential predators. This is the butterfly that gives its name to the Morphotex™ fabric shown in Figure 1.

Each morpho wing scale is composed of photonic crystals, typically 5 μm across. Each crystal is composed of a microscopic 'Christmas tree' protein structure (Figure 8b). It is difficult to model mathematically how light interacts with each crystal but, if each branch of the tree acts as a thin film about 110 nm thick, there is a reflectance maximum at 49°. Each crystal has a slightly different orientation, the layers within each crystal are not parallel and the thickness within each layer varies — characteristics referred to as 'fabrication noise' by researchers struggling to replicate the structure and its properties. Scientists are trying to fabricate such tree structures for use as chemical sensors. Wetting the surface would change the ratio of the refractive indices and shift the location of the reflectance maxima.

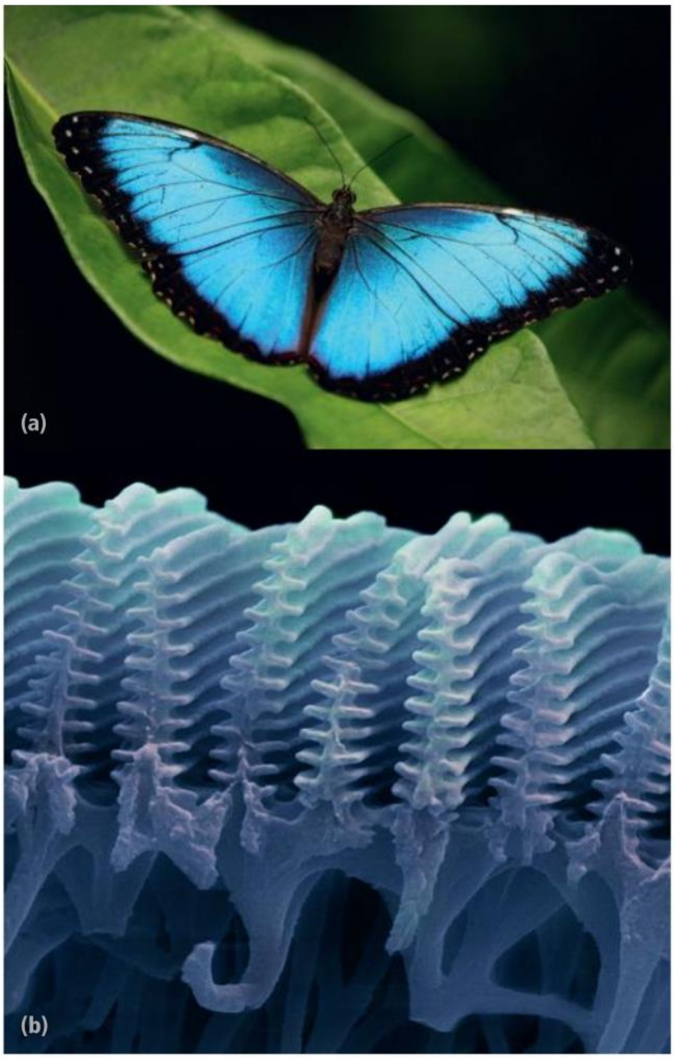


Figure 8 **(a)** The cobalt-blue wings of a morpho butterfly. **(b)** A scanning electron micrograph showing the 'Christmas tree' structure

Final comments

One advantage of new materials designed to have structural colour is that dyes are not used, saving water and energy, and reducing pollution. Structurally coloured clothes are colourfast, so they do not fade with repeated washing or exposure to UV light. Plastics would be much easier to recycle if surface structures replaced chemical pigments — these are under development. So structural colours are not only pleasing to look at, but can also help us address some important environmental issues.

RESOURCES

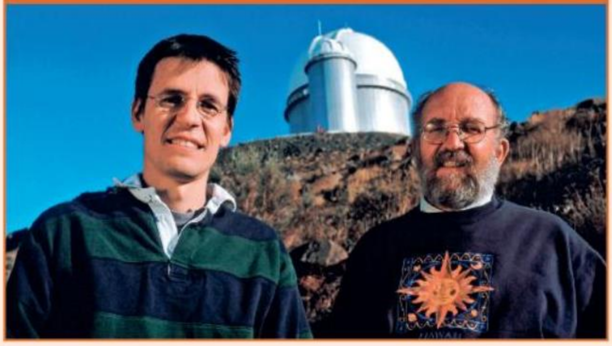
The photograph in Figure 1 was provided by Donna Sgro, textile designer and lecturer at the University of Technology, Sydney, Australia. See for more information and images.

Mike Follows teaches physics at King Edward's School, Birmingham, and is a keen science author.

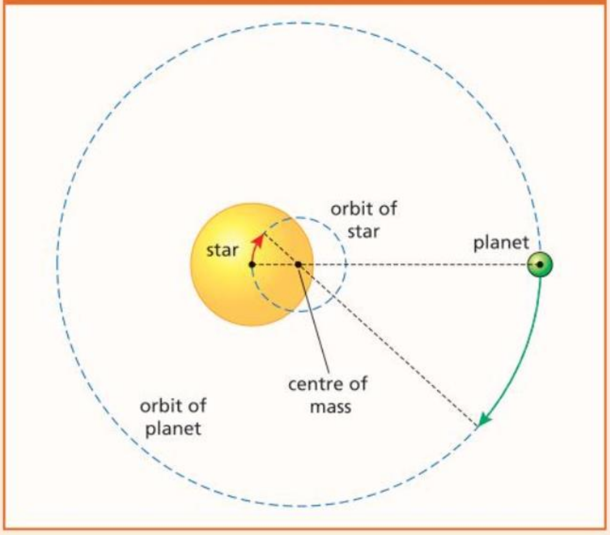
AT A GLANCE

Exoplanets

1 Didier Queloz (left) and Michel Mayor at La Silla Observatory in Chile, which hosts the TRAPPIST exoplanet programme

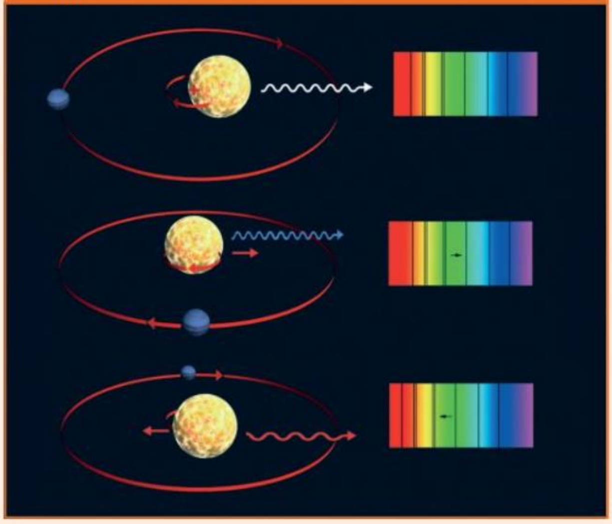


2 A planet and star orbit around their common centre of mass

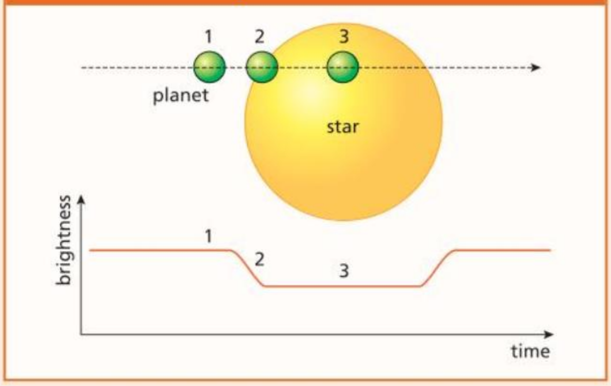


In 2019 Michel Mayor and Didier Queloz (1) received a share of the Nobel prize in physics 'for the discovery of an exoplanet orbiting a solar-type star'. Working at the University of Geneva, they pioneered techniques for detecting planets orbiting stars other than the Sun. Following their discovery of the first exoplanet in 1995, other astronomers joined the search and by June 2020 there were 4141 known exoplanets. Of the 2.5×10^{12} stars in the Milky Way, about 3000 are now known to have planetary systems.

3 Absorption lines in a star's spectrum are Doppler shifted as the star wobbles



4 A transiting exoplanet causes a slight dip in a star's observed brightness



The first exoplanets were found using *Doppler spectroscopy*. When a planet orbits a star both move around their common centre of mass (2). As the star wobbles towards and away from an observer, the absorption lines in its spectrum are Doppler shifted (3):

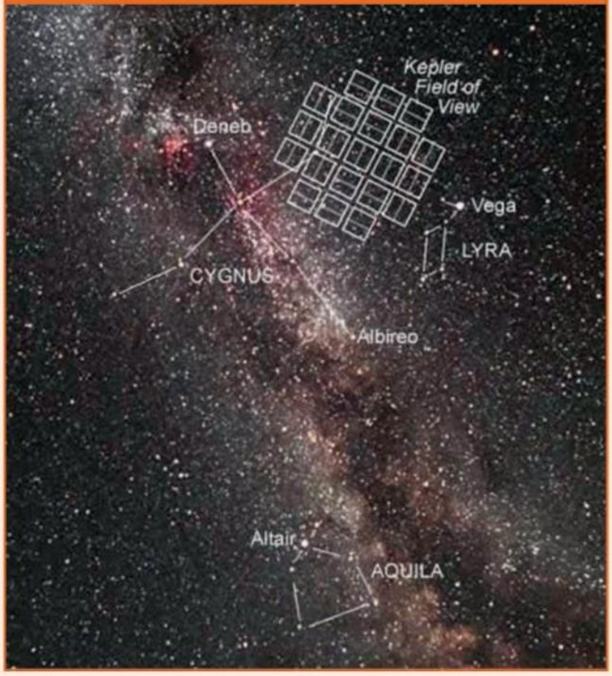
$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c}$$

where $\Delta\lambda$ is the change in wavelength, λ the unshifted wavelength, c the speed of light and v the star's velocity along the observer's line of sight, known as its *radial velocity*.

An observer far outside the solar system might detect Jupiter by seeing our Sun's radial velocity change by about $12\,\mathrm{m\,s^{-1}}$ with a period of about 12 years. Planets with smaller mass or in larger orbits (lower orbital speed) are harder to detect, and this method only finds planets whose orbits are seen from the side.

Most exoplanets are now discovered using *transit* measurements. When a planet passes in front of a star there is

5 The Kepler telescope surveyed about 500 000 stars in a small part of the Milky Way



6 Combined infrared and optical observations of exoplanet 2M1207b (lower left) in orbit

a slight dip in the star's observed brightness (4). NASA's Kepler space telescope monitored stars in a small part of the Milky Way (5) and detected nearly 3000 exoplanets.

Astronomers use observational data to calculate the masses and orbital radii of exoplanets. Many known exoplanets have masses similar to Jupiter's or larger. A few large, nearby exoplanets have been imaged directly, for example 2M1207b, which has a mass about five times that of Jupiter and an orbital radius of about 55 AU. It orbits a brown dwarf star about 230 light years from Earth (6, 7).

Earth-sized exoplanets have been found, and some orbit their stars in the *habitable zone*, where temperatures allow liquid water to exist (8). But many exoplanets are quite unlike those in our solar system.

In March 2020 the media carried news of an exoplanet whose temperature exceeds 2400 °C, which is hot enough for iron to liquefy (9). Such exotic planets are causing scientists to rethink their theories of how planets might have formed.

7 Unit conversions

Name and symbol	Definition	SI equivalent
Astronomical unit, AU	Mean Earth–Sun distance	1.496×10^{11} m
Light year, ly	Distance light travels in a vacuum in 1 year (3.154×10^{7} s)	9.461×10^{15} m

8 Artist's impression of an Earth-like planet in the TRAPPIST-1 system, discovered by the TRAnsiting Planets and Planetesimals Small Telescope at La Silla Observatory

9 Artist's impression of the exoplanet Wasp-76b, where molten iron may fall as rain

PhysicsReview*Extras*

Download this poster at

[redacted]

Measuring gravity's effect on time

Peter Main

The faulty launch of a pair of satellites gave an opportunity to measure how gravity affects the passage of time. The results agreed with the predictions of Einstein's theory of general relativity

Exam links

The terms in bold link to topics in the **AQA**, **Edexcel**, **OCR**, **WJEC** and **CCEA** A-level specifications, as well as the **IB**, **Pre-U** and **SQA** exam specifications.

Two satellites in elliptical **orbits** were used to measure **time dilation** in a **gravitational field**.

When things go wrong at the launch of a satellite, they usually go very wrong, leading to the loss of the satellite or rendering it useless for the job it was designed to do. However, the faulty launch of a pair of satellites in 2014 from the European Space Agency (ESA) launch site in French Guiana turned out to be the start of an ambitious and unplanned experiment.

The satellites were destined for the Global Navigational Satellite System, known as the Galileo Project, which is the European equivalent of the American GPS. A frozen fuel line in the fourth stage of the *Soyuz* rocket meant the satellites were launched in the wrong direction, into highly elliptical orbits that were totally unsuitable for their intended purpose. What happened next was a most remarkable recovery of a failed launch which, incidentally, presented an opportunity to test an important feature of Einstein's theory of general relativity.

PhysicsReview*Extras*

Get practice-for-exam questions based on this article at

Recovery of orbital stability

The satellites were meant to be in the same circular orbit 23 000 km above the Earth's surface but displaced 180° from each other. Instead, they were in a highly elliptical orbit with a closest approach to Earth (perigee) of 13 700 km, rising to a furthest distance (apogee) of 25 900 km. In addition, the orbit was incorrectly inclined to the Earth's equator. There was a substantial list of things to do to determine the actual orbit, reorient antennae to set up robust radio links, bring the satellites under control and make them safe. All this was achieved in about a month, then further action to change the orbit and recover the situation could begin.

The satellites had enough fuel on board for small course corrections during their planned 12-year life span, but that was insufficient to move them into their correct orbit. A recovery plan was devised by a collaboration of space-flight specialists from Germany, France, Italy and Britain. One of the concerns was the lowest orbital height of 13 700 km, exposing the satellites to the harmful radiation of the Van Allen radiation belts (see below). Such intense radiation would degrade the satellites' electronics, leading to premature failure.

The recovery scheme therefore involved a series of manoeuvres, which raised the lowest point of the orbit by more than 3500 km.

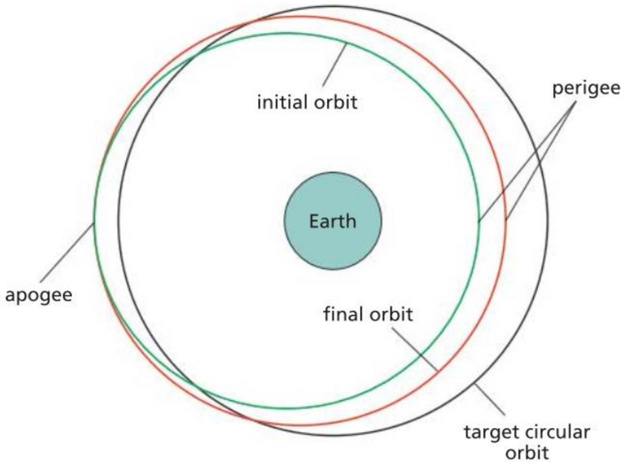


Figure 1 **The initial, target and final satellite orbits. The Earth's centre of gravity is at one focus of the elliptical orbits**

This greatly reduced the satellites' radiation exposure, ensuring a long-term reliable performance. In addition, the orbit was more circular, which helped the satellites to fit into the Galileo navigational system and gradually positioned them to be 180° from each other round the orbit. It took about 7 months to achieve this.

Figure 1 shows the initial and final orbits of the satellites as well as the desired orbit. Now the satellites were in a position where, with a certain amount of ingenuity, they could be used for their original purpose.

The Van Allen radiation belts

The discovery of the Van Allen radiation belts makes an interesting story. In January 1958, at the height of the Cold War, the USA launched its first Earth satellite, *Explorer 1*. It was 3 months after the Russians had launched their first satellite, *Sputnik 1*, and 2 months after *Sputnik 2* was successfully launched. The Americans were desperate to catch up with the Soviet Union, but they had been held back by launch failures.

James Van Allen of Iowa University was given the urgent job of designing the scientific payload of *Explorer 1*, which he made as basic and as light as possible. In fact, the satellite just contained some temperature sensors, an acoustic sensor and a Geiger counter with a tape recorder, as well as the essential communications equipment. The tape recorder was to record the output of the Geiger counter while the satellite was out of contact with the ground station.

Van Allen was expecting to detect only cosmic rays with the Geiger counter, but instead it recorded a rapid build-up of radiation, and then suddenly nothing. It was thought (and confirmed later) that the radiation was so intense that it had overwhelmed the counter, making it unable to register anything. It is now well known that the Earth is surrounded by radiation belts consisting of energetic charged particles, mostly electrons, captured from the solar wind by the Earth's magnetic field. Although the Americans were second into space, they were the first to make a discovery about the Earth's space environment.

A cross-section of the radiation belts is shown in Figure 2. There are three altogether and they clearly follow the shape of the Earth's magnetic field. Their closest approach to the Earth's surface is near the magnetic poles, where they can interact with the upper atmosphere to give rise to the aurora borealis (northern lights) and aurora australis (southern lights).

The experiment

Now back to the satellites. Because they were designed to be used in a navigational system, they contained atomic clocks, which were highly accurate — to 1 second in about 30 million years, or roughly 0.1 nanoseconds per day (1 nanosecond = 10^{-9} s). Einstein's theory of gravity — his theory of general relativity (Physics Review Vol. 29, No. 3, pp 2–6) — predicts that time progresses at a rate that depends upon the strength of the gravitational field.

The satellites were in an orbit that changed height from about 17 000 km to 26 000 km. Over that distance the acceleration due to Earth's gravity changes by a factor of 1.9. So, according to the theory, from the point of view of an observer on Earth, the clocks in the satellites would run faster at a height of 26 000 km than they would at 17 000 km. The difference in rate, although tiny, should be quite measurable with the instruments onboard. This is known as gravitational time dilation and is the basis of the redshift used by astronomers.

It should be pointed out that there is no doubt about Einstein's theory because it has been used by astronomers and cosmologists for a long time with spectacular success. However, this aspect of the theory — the change in the rate at which time passes — has rarely been measured directly. The last time was over 40 years ago by *Gravity Probe A* (see below). All other applications, such as its successful use in the GPS system, have simply assumed the theory to be correct and accepted its predictions. These satellites in an elliptical orbit therefore presented an unexpected but golden opportunity to test Einstein's theory to a much higher precision than was previously possible. This could only be done because the experiment did not interfere with the normal running of the satellites.

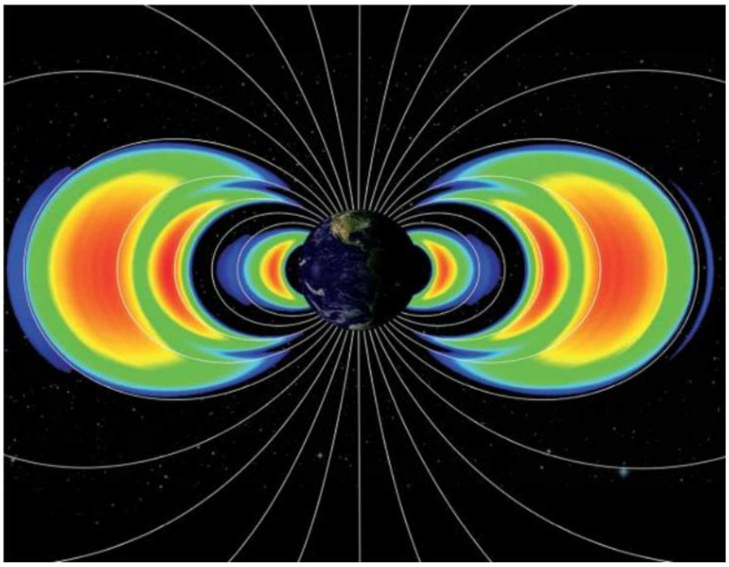


Figure 2 **Cross-section of the Van Allen radiation belts superimposed on the Earth's magnetic field lines**

Gravity Probe A

The last occasion when time dilation was measured accurately was in 1976. It was a space-based experiment, carried out by NASA and the Smithsonian Astrophysical Observatory. The space probe, carrying a hydrogen maser, was launched into a suborbital flight lasting just under 2 hours and reaching a height of 10 000 km. A maser is like a laser except that it produces microwaves instead of visible light. The microwaves in this case had an extremely stable frequency, equivalent to a clock that is accurate to 1 second in 50 million years.

The frequency of the microwaves received from the probe was not only altered by time dilation, but also by the Doppler shift caused by the probe moving relative to the ground. After correction for the Doppler shift and an adjustment given by special relativity, the frequency was compared with that produced by an identical maser on the ground. The frequency difference was then compared with that predicted by general relativity. It was found that the maser frequency increased in the weaker gravitational field along the probe's path by an amount predicted by Einstein's theory to a precision of about 100 parts per million — ten times better than the previous determination of the effect.

Procedure and results

To reduce experimental error, it was important to know the position and movement of each of the navigation satellites as accurately as possible. This was done by taking laser-ranging measurements. Each satellite was already fitted with laser retroreflectors, sometimes called corner cube reflectors (Box 1). Their properties mean that a beam of light will always be reflected back to its source, no matter what the angle of incidence might be. Using laser ranging, it became clear that the greatest source of error in predicting the position of the

Box 1 Retroreflectors

A retroreflector is a device that reflects light back to its source, for any direction of incidence. It consists of three mutually perpendicular reflective surfaces, arranged to form the corner of a cube. Figure 1.1 illustrates this and shows two different rays being reflected back along their direction of incidence. Astronauts on the *Apollo 11*, *14* and *15* missions left retroreflectors on the Moon as part of the Lunar Laser Ranging Experiment, which measures the distance between Earth and Moon to millimetre precision (Physics Review Vol 18, No. 1, pp. 12–15).

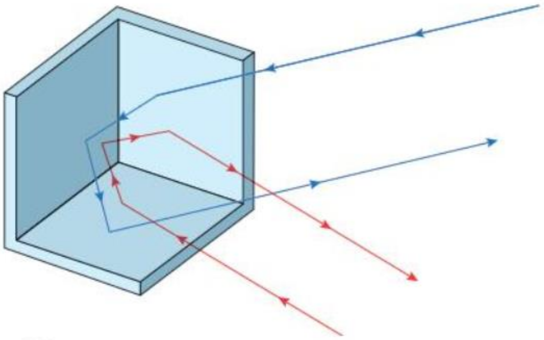
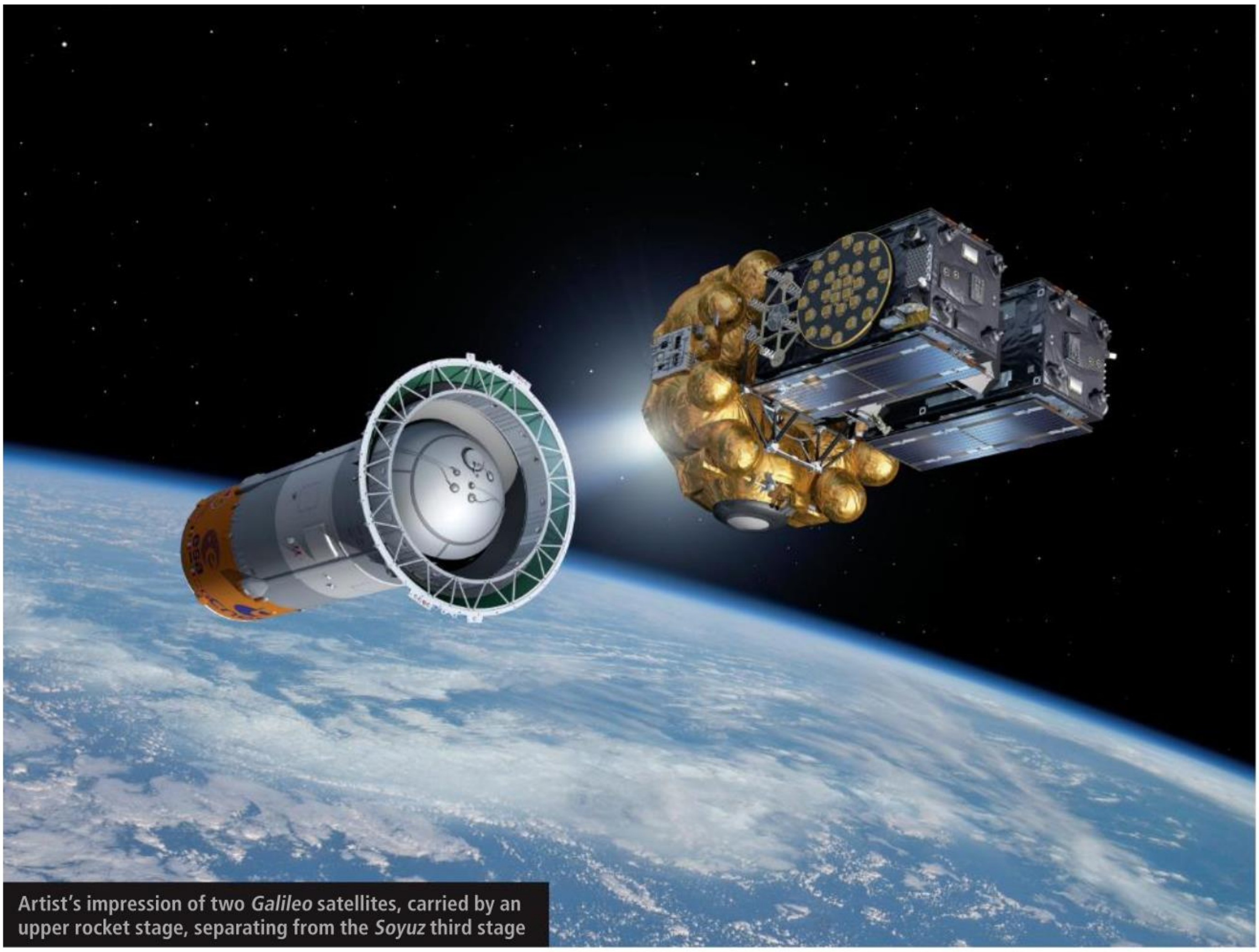


Figure 1.1 A corner cube reflector

Artist's impression of a fully operational *Galileo* satellite.



Artist's impression of two *Galileo* satellites, carried by an upper rocket stage, separating from the *Soyuz* third stage

satellite was due to the effect of the Sun's radiation. Photons possess momentum and they impart momentum to any surface on which they are incident, including the satellite. Ignoring this effect could put the predicted position of the satellite in error by about 1 km per day.

The atomic clock frequency of each satellite was recorded as a function of height for 1000 days and compared with the frequency of an identical atomic clock on the ground. The difference in frequency was then compared with the prediction of general relativity. It was found that general relativity predicted the gravitational time dilation correct to about 20 parts per million — the most precise determination to date and five times better than that obtained previously.

Comment

Why bother testing a theory that has performed so well for about 100 years? The testing continues — there is now an experiment on the International Space Station that aims to improve the measurement of time dilation to a precision of about two parts per million. This is a factor of ten better than using the navigation satellites. Physics progresses when a theory gives inaccurate or wrong predictions. Experiments can then be devised that will show what went wrong, enabling physicists to amend the inaccurate theory or develop a new one.

The theory of gravity is certainly worth testing because there are aspects of gravity that remain a mystery. For example, neither dark matter nor dark energy can be explained. Finding an inaccurate prediction of general relativity will hopefully give clues as to where to look for a better theory.

A related problem is that no way has yet been found to combine general relativity with the highly successful quantum theory to produce a theory of quantum gravity. It appears that our formulation of either general relativity or quantum theory (or both) must change before that can happen. The person who achieves this will certainly deserve a Nobel prize.

References and further reading

The inside story of the faulty launch, recovery and exploitation of the satellites:

An excellent video of the time dilation experiment:

Peter Main is on the academic staff of the University of York and is also a member of the Physics Review editorial board.

Electric vehicles

How do they work?

Susan Street

Susan Street looks at the development of the electric vehicle (EV), considers the physics involved in the rechargeable battery and the motor, and discusses the importance of this engineering revolution

Exam links

The terms in bold link to topics in the **AQA**, **Edexcel**, **OCR**, **WJEC** and **CCEA** A-level specifications, as well as the **IB**, **Pre-U** and **SQA** exam specifications.

The performance of EVs and rechargeable batteries is described in terms of **charge**, **energy** and **power** in **electric circuits**, **efficiency**, and **electromagnetic induction**.

You may be surprised to read that EVs have been around in Britain for over a century. The first all-electric car with a rechargeable battery was invented in 1884 by Shropshire ironworker Thomas Parker (better known for electrifying the London Underground). His lead-acid battery is still used in every petrol or diesel car, but the internal combustion engine soon delivered a much longer-range journey and by 1935 EVs had all but disappeared.

PhysicsReview*Extras*

Get practice-for-exam questions based on this article at

This century has seen a revival in EV manufacture, driven by the need to improve air quality and cut carbon emissions, and by the development of better batteries.

Types of EV

There are three types of electric vehicle:

- The hybrid electric vehicle (HEV) has a petrol or diesel engine and an electric motor to reduce fuel consumption. The battery is recharged as the brakes are applied.
- The battery in the plug-in hybrid electric vehicle (PHEV) is recharged from the National Grid, and its electric motor drives the wheels. Its petrol engine is a back-up for driving the wheels and charging the battery.
- The all-electric vehicle (EV or AEV) uses a rechargeable battery to operate a high-power electric motor to drive the wheels.

Portable energy

Cells and batteries are portable energy stores. A cell's emf is the amount of energy that it transfers to each coulomb of charge (Box 1). The way the chemicals in the cells release and move electric charge determines both the cell's emf and the potential difference (pd) it delivers to a working circuit. (The pd is slightly lower than the emf because energy is dissipated in the cell as a result of its internal resistance — see *Skillset* on pages 6–8.)

When cells are joined in series to make a battery, the battery emf is the sum of the individual cell emfs, and the pd it supplies is the sum of the cell pds.

Box 1 summarises some key electrical equations and SI units. To compare car batteries, we use larger units of charge, energy and power. Three quantities are typically used to measure battery performance: *charge capacity*, *energy density* and *efficiency*.

Charge capacity is the total charge that flows through a battery while it is operating. One ampere-hour (Ah) is the charge that flows when a battery provides a current of 1 amp for 1 hour:

$$1\,\text{Ah} = 3600\,\text{C}$$

Energy output and storage are measured in kilowatt-hours, kWh, the unit found on domestic energy bills. 1 kWh is the energy transferred by a device with a power of 1 kW operating for 1 hour:

$$1\,\text{kWh} = 3.6 \times 10^6\,\text{J}$$

If we know current in amperes and and pd in volts, then:

$$\text{power in kW} = \frac{\text{current in A} \times \text{pd in V}}{1000} \times \text{number of cells} \quad (1)$$

$$\begin{array}{l}\text{energy stored in}\\ \text{a battery (kWh)}\end{array} = \frac{\text{battery charge capacity} \times \text{pd}}{1000} \times \begin{array}{l}\text{number}\\ \text{of cells}\end{array} \quad (2)$$

A battery's energy density is its total stored energy divided by its mass, usually expressed in kWh per kg:

$$\text{energy density} = \frac{\text{energy stored in kWh}}{\text{mass of battery in kg}} \quad (3)$$

Box 1 Electrical equations and SI units

In SI units, charge, q, is expressed in coulombs (C) and current, I, in amperes (A). Current is the rate of flow of charge:

$$I = \frac{q}{t} \quad (1.1)$$

$$1\,\text{A} = 1\,\text{C}\,\text{s}^{-1}$$

$$q = It \quad (1.1a)$$

$$1\,\text{C} = 1\,\text{A}\,\text{s}$$

Power, P, is the rate of transfer of energy, E; it has SI units of watts (W).

$$P = \frac{E}{t} \quad (1.2)$$

$$1\,\text{W} = 1\,\text{J}\,\text{s}^{-1}$$

$$E = Pt \quad (1.2a)$$

$$1\,\text{J} = 1\,\text{W}\,\text{s}$$

Potential difference, V, and emf are measures of the energy, E, transferred per unit charge, and have SI units of volts (V).

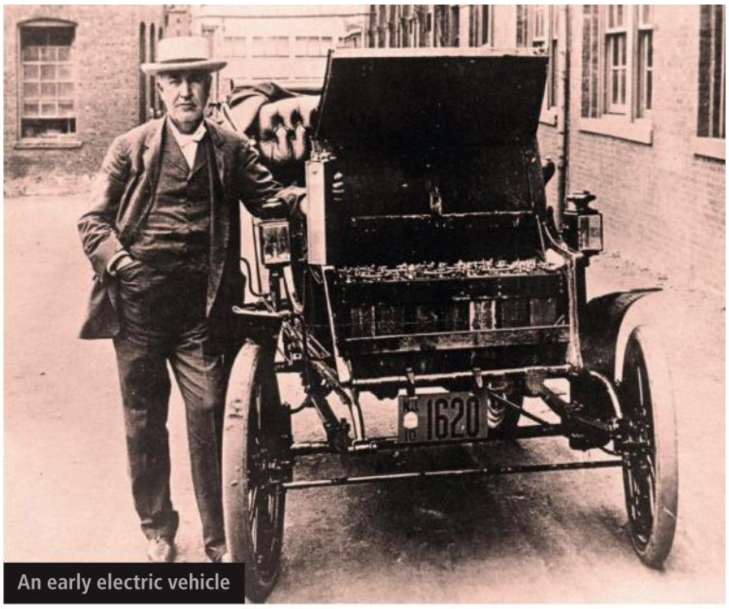
$$V = \frac{E}{q} \quad (1.3)$$

$$1\,\text{V} = 1\,\text{J}\,\text{C}^{-1}$$

power = joules per coulomb × coulombs per second

$$P = VI \quad (1.4)$$

$$E = qV = VIt \quad (1.5)$$



An early electric vehicle

A battery's efficiency is defined as:

$$\begin{array}{l}\text{battery}\\ \text{efficiency}\end{array} = \frac{\text{useful energy output during discharge}}{\text{energy transferred during charging}} \quad (4)$$

Efficiency is often expressed as a percentage.

The rechargeable battery

The main components of each cell are:

- a chemical solution (an electrolyte) that conducts electricity due to the presence of charged atoms (ions)
- plates (electrodes) that allow a chemical reaction to take place in the electrolyte and provide contacts for electrons to enter and leave the cell

Parker's electrolyte was dilute sulfuric acid; one electrode was lead and the other lead coated with lead dioxide. His solution contained positive ions (H^+) and sulfate ions (SO_4^{2-}). He connected six 2.1 V cells in series to make a 12.6 V battery. Lead-acid batteries are typically 70% efficient and have an energy density of 30–40 Wh kg^{-1}, which is too low for modern EVs.

The main successor to Parker's battery is the lithium-ion (Li-ion) battery. The electrolyte is an organic solution of lithium. One electrode is a lithium metal oxide with manganese, nickel and cobalt added to increase charge density and give stability. The other electrode is graphite.

The positive ions are lithium Li^{2+}. During discharge they move from anode to cathode through the electrolyte and a separator, which is a polymer barrier full of tiny (30–100 nm) pores holding electrolyte but allowing the ions to pass. Lithium ions are absorbed by the cathode, electrons are released from the anode and the resulting flow of electrons in the external circuit allows the motor to do work. Charging is the reverse process. Electrons are drawn away from the cathode and positive ions are attracted to the anode until it is packed with ions and the cell is fully charged again.

The Li-ion battery is 80–90% efficient and, since lithium has a low density (533 $\text{kg}\,\text{m}^{-3}$), it has an energy density of 200 Wh kg^{-1}. It is the go-to battery for EVs.

The Nissan Leaf is a popular EV

The Nissan Leaf

One popular EV is the Nissan Leaf, first produced in 2010. It is a good example of recent achievements in this field.

The basic Leaf has a Li-ion battery pack containing 48 modules (connected in series) of four cells (two in series and two in parallel) under the seats. Each cell of 3.75 V is rated at 32.5 Ah. Using Equation 2 to calculate the energy stored:

$$\text{energy} = 32.5\,\text{Ah} \times 3.75\,\text{V} \times 48 \times 4 = 23\,400\,\text{Wh} = 23.4\,\text{kWh}$$

Only 80% of this is available because Li-ion batteries are destroyed if completely discharged. The storage capacity is temperature dependent, increasing in hot weather.

The Li-ion battery discharges as it delivers an electric current to the motor. To recharge the cells, an electric current in the opposite direction is used to reverse the chemical reaction and increase the stored energy.

Recharging the battery

The main method of recharge is from an external electricity supply.

There are two on-board charging inlets under a flap in the bonnet. The cabling and connector depend upon the power rating of the supply you are using. Off-board charging cables are supplied at rapid-charging stations, with electronics that match up to the car's inlet with a 'handshake'. Fast on-board charging at home or at the workplace involves connecting a cable to a highly rated socket. For overnight slow charging in the home, a three-pin socket is sufficient. Table 1 compares how long it *can* take (official and real-world data vary) to charge a 24 kWh Li-ion battery to 80% capacity with different supply power ratings.

The motor and transmission

When a wire carries an electric current in a magnetic field, it experiences a force at right angles to both the current and field. When the wire is coiled and mounted on a shaft it becomes the rotor (rotating drive) of a motor, spinning within a surrounding magnetic field provided by a stationary electromagnet called the stator.

EV motors use alternating current (AC) to supply the stator so its magnetic field alternates too. To produce a constant rotation rate in one direction, the rotor must be synchronised with the stator's changing magnetic field (see 'Dynamos in cars — why AC?', PHYSICS REVIEW Vol. 26, No. 3, pp. 18–21).

The synchronous AC motor in the basic Leaf has a power of 80 kW (compare this with a maximum of 1.4 kW for a rechargeable power drill). It produces a twisting force (torque) of 280 Nm to a drive shaft that turns the wheels. The maximum rate of rotation is 10.4×10^3 revolutions per minute (rpm), varying as the car accelerates and decelerates. A single reducing gear with a ratio of 8:1 reduces this maximum to 1.3×10^3 rpm. An electric motor can provide maximum torque even at low speeds, so a conventional gearbox is not needed. With a wheel diameter of 63.6 cm, the vehicle's top speed is about 156 km h^{-1}.

Table 1 Battery charging times

Charging type	Power rating/kW	Time/hours
Rapid	50	0.5
Fast	22	1.5
Fast (home)	7	4
Slow	3	12

An inverter is included between the battery and the motor unit. This is an electronic controller that uses signals from the single 'epedal' to speed up the motor for acceleration when the pedal is pushed down and causes the motor to act as a generator when the pressure on the pedal is released for deceleration, at the same time discharging or recharging the battery, respectively. It also converts the direct current (DC) from the battery into AC for driving the motor and AC back to DC again for battery charging.

Regenerative braking uses the motor as a dynamo, reducing the vehicle's kinetic energy and recharging the battery. As regenerative braking gets less effective as speed decreases, standard friction brakes are applied for the final halt or for an emergency stop.

Performance

To increase efficiency, drag caused by air resistance is reduced as much as possible. The Leaf has a tapered nose, a flat underbody and two rear spoilers, giving the vehicle a drag coefficient of 0.28 — similar to that of a bullet.

A common performance measure for EVs is kWh used per 100 km. A typical figure for combined city and open road driving is about 18 kWh per 100 km: a fully charged 24 kWh battery should give a range of about 130 km.

There are losses within the charging system in the vehicle but in principle a more highly rated battery drives a more powerful motor and delivers a longer journey. All the measurements are fraught with variables, such as terrain, speed and driver competence.

Next steps

As demand for EVs rises, the network of street charging stations in towns and cities and multi-charging-point stations countrywide is expanding, with maps and apps available for drivers to plan for longer journeys. Check out your local availability at Zap-Map ('). In the home, wall chargers for EVs will become the norm.

A rapid charge port

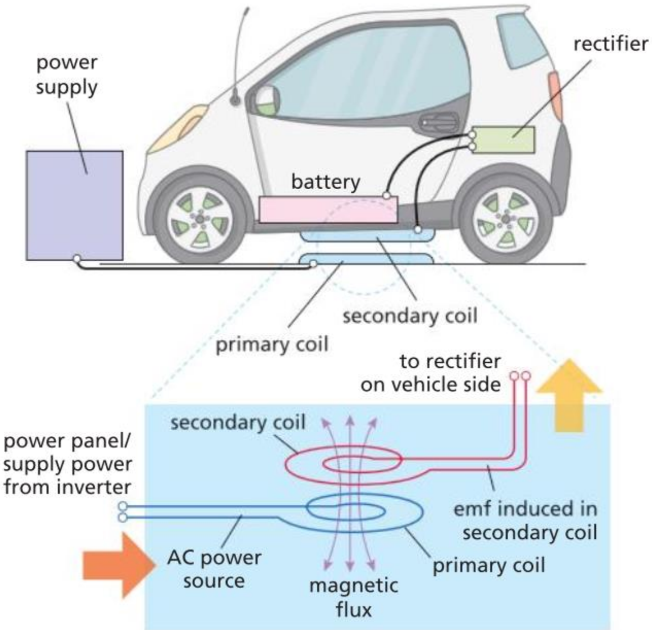


Figure 1 Wireless recharge

Manufacturers are vying to design more efficient vehicles, both domestic and commercial, with longer ranges. The next generation Leafs, for example, have 40 kWh and 62 kWh batteries.

Future EVs might use a wireless charging system with an air transformer. At a chosen parking space, for example outside the home or workplace, an alternating magnetic field is generated with mains AC in a primary coil fixed in the ground (Figure 1). The changing magnetic flux is directed vertically and links with a secondary coil attached to the base of the car above it, where it induces a secondary current, which is fed to an AC/DC rectifier in the car's charging port. The challenge will be to install the necessary charging infrastructure, to make it efficient and to ensure that the National Grid can supply sufficient energy when required (see 'Will electric cars break the National Grid?', Physics Review Vol. 28, No. 3, pp. 8–11).

The environment

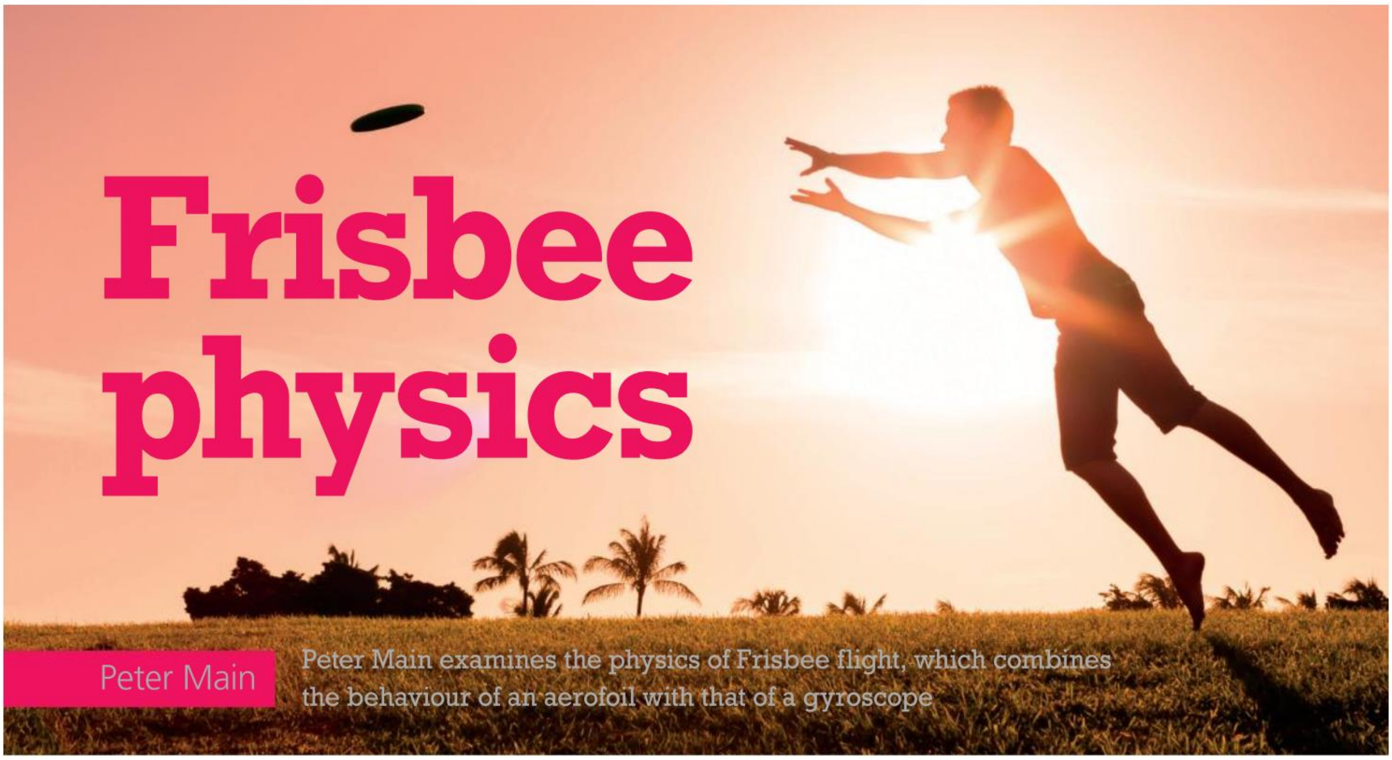
The AEV is emission free at the point of use — it has no exhaust. The clamour to reduce greenhouse gases is rising as climate change bites. The UK government has adopted a 'Road to Zero' strategy that aims for:

- 50% EV (PHEV and EV) new car sales by 2030
- no new petrol/diesel cars in production by 2040
- 100% EV by 2050

Whether these targets can be achieved will depend to a certain extent on the necessary charging infrastructure.

Across Europe the installation of thousands of charging stations is well underway. However, electric cars are only as green as their power supply. Worldwide, a concerted approach to halt the burning of fossil fuels, both on the roads and at the power stations, is the ultimate aim for tomorrow's drivers.

Susan Street is a former physics teacher.



Frisbee physics

Peter Main

Peter Main examines the physics of Frisbee flight, which combines the behaviour of an aerofoil with that of a gyroscope

Exam links

The terms in bold link to topics in the **AQA**, **Edexcel**, **OCR**, **WJEC** and **CCEA** A-level specifications, as well as the **IB**, **Pre-U** and **SQA** exam specifications.

The **conservation** of **mass**, **energy** and **momentum**, together with **Newton's second and third laws of motion**, help explain the lift provided by an aerofoil. Spin stablises a Frisbee flight because of of **angular momentum conservation**.

For decades Frisbees have been a source of fun for people of all ages. A skilful thrower can make them follow interesting curved trajectories, go in a straight line or even appear to hover in mid-air. These simple plastic discs can be thrown long distances and, best of all, are inexpensive — over 100 million are sold each year.

The entertaining aspects of a Frisbee can be explained in terms of just two physical concepts — it is both an aerofoil and a gyroscope.

The Frisbee story

Before we get into the physics, there is an interesting story about how Frisbees started. It all began in 1871 in William Russell Frisbie's small bakery in Connecticut, USA. Frisbie's delicious fruit pies were popular at the nearby Yale University. The students enjoyed throwing the empty metal pie plates around when they discovered they could make them fly long distances. The plates quickly became known as 'Frisbies'.

Production of the plastic flying discs started in the 1950s when they were marketed by the Wham-O Manufacturing Company in California. They called them 'Frisbees' — the slight change in name was to avoid trademark problems. Note that Frisbee is a registered trademark of Wham-O Manufacturing, hence the capital letter. Other firms manufacture them but are not allowed to call them Frisbees — they are known as 'flying discs'.

PhysicsReview*Extras*

Get practice-for-exam questions based on this article at

Aerodynamic lift

A Frisbee acts as an aerofoil as it flies through the air. A common example of an aerofoil is an aeroplane wing, which provides lift to an aircraft. How it provides lift is quite complicated, involving conservation of mass, energy and momentum. There is no straightforward explanation, and many textbooks make mistakes when they attempt a simplification.

The original Frisbie's pie tin

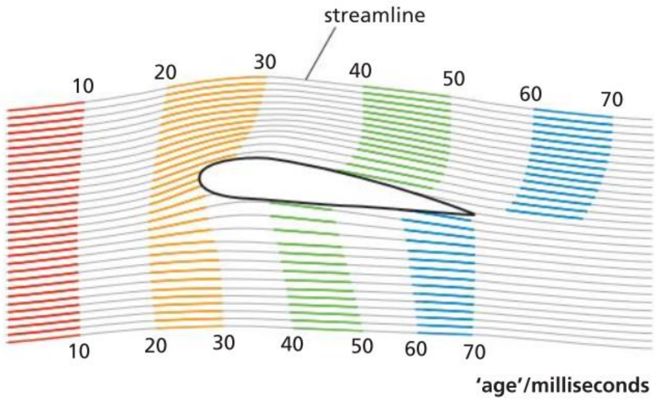


Figure 1 **The calculated flow of air over an aerofoil. The air follows streamlines with varying speed**

Let us just accept the accurately calculated flow of air over a wing, as shown in Figure 1. The air flows from left to right, following the continuous streamlines. Vertical slices of air, initially at equal time intervals, are represented by streamers of different colours. The streamers are stretched out over the top of the wing, showing that the air has increased its speed. In fact, the maximum velocity produced by the aerofoil is approximately twice that of the undisturbed air.

Conversely, the slices of air travelling along the underside of the wing are going more slowly. This means the two parts of each slice do not meet after travelling around the wing. Figure 1 shows that the upper part of the blue slice has moved well ahead of the lower part when it reaches the back of the wing.

Along a streamline, the static air pressure, p, is related to the dynamic pressure (kinetic energy per unit volume), $\frac{1}{2}v^2\rho$, by conservation of energy, giving the equation:

$$p + \tfrac{1}{2}v^2\rho = \text{constant} \qquad (1)$$

where v is the speed of the air and ρ is its density. The equation shows that if v is increased, then p must decrease to compensate. The faster-flowing air on top of the wing must therefore be at a lower pressure than the slower air underneath. This difference in pressure creates the upward force on the wing that we call lift.

A flying Frisbee and an aircraft wing are both aerofoils

Another way of looking at this is by considering the rising air immediately in front of the wing compared with the air immediately behind the wing, which is falling. Clearly, there has been a change of momentum of the air. The rate of change of momentum, according to Newton's second law of motion, is proportional to, and in the same direction as, the applied force. The force must therefore be in a downward direction. The force is exerted by the wing and Newton's third law tells us there must be an equal upward force on the wing, which is the lift.

Notice that neither of these interpretations of Figure 1 explains how the flow pattern occurs in the first place. A nice animated diagram of the flow of air over an aerofoil can be found at [illegible] **_(force)**.

Forces on the Frisbee

The total aerodynamic force on a flying disc can most conveniently be described in terms of the two components of *lift* and *drag*. The drag force is in a direction exactly opposite from the direction of flight, and the lift force is upwards at right angles to this. The vector sum of the two is the total aerodynamic force on the disc, which acts through the *centre of pressure*. Because the air pressure changes across the disc, the centre of pressure does not correspond to the geometric centre of the disc, but is closer to the leading edge, as seen in Figure 2.

The remaining force on the disc is its *weight*, the gravitational force that acts vertically downwards through the *centre of mass*. By symmetry, the centre of mass coincides with the geometric centre of the disc. Since the centre of pressure and the centre of mass do not coincide, the forces acting through them tend to flip the Frisbee over. With a successful throw this does not happen, so we need to discover what keeps it stable.

Gyroscope

You may well have played with a spinning top or a gyroscope. Their behaviour can be quite fascinating. One of the first things to consider is the rate of spin, which is a scalar quantity and can be expressed in radians per second. If the direction of the axis of rotation is included in the description it becomes a vector quantity, with the vector pointing along the rotation axis. It is then called the *angular velocity*.

In linear motion, multiplying velocity by mass gives momentum. The equivalent in rotational motion is angular momentum, which is also a vector quantity. When a Frisbee is thrown, it is given a lot of spin, so it has a substantial amount of angular momentum.

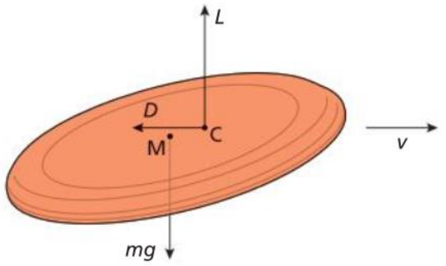


Figure 2 **The forces acting on a Frisbee travelling with a velocity *v*. M is the centre of mass through which the weight, *mg*, acts. C is the centre of pressure through which the forces of lift, *L*, and drag, *D*, act**

Gyroscopic action helps to keep the Frisbee stable in flight

Moment of inertia in rotational motion is equivalent to mass in linear motion. For a point mass m at a distance r from the axis of rotation, its *moment of inertia* I is:

$$I = mr^2 \qquad (2)$$

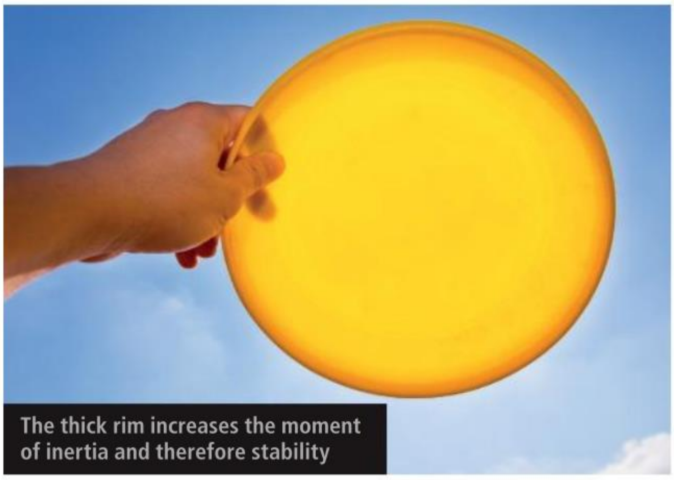
Multiplying angular velocity, ω, by moment of inertia, I, gives *angular momentum*:

$$L = I\omega \qquad (3)$$

Just as linear momentum is conserved when no external resultant force acts, angular momentum is also conserved when there is no turning force (torque) acting on an object. This requires that both the rate of spin and the direction of the axis of rotation do not change (ignoring friction with the air). Requiring the axis of rotation to remain constant is exactly what is needed to stabilise the Frisbee and stop it from flipping over. Without spin, the Frisbee cannot fly.

Construction

The shape and construction of a Frisbee have important effects on its performance. We have already seen that its shape allows it to act as an aerofoil, but let us look at the purpose of the thick rim. First, and most important, it allows us to grip and throw the disc. Without it, the throw would be a lot less effective.



The thick rim increases the moment of inertia and therefore stability

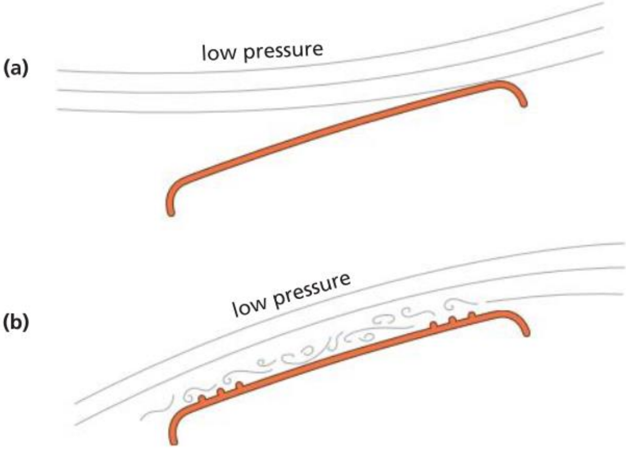


Figure 3 **Streamlines showing the effect on airflow of surface ridges: (a) without ridges, (b) with ridges**

In addition, and more to do with physics, the effect of the thick rim is to concentrate a lot of the Frisbee's mass far from the rotation axis. Consequently, it increases the moment of inertia and so contributes significantly to the stability of the Frisbee in flight.

Another effect of the rim is partially to turn the Frisbee into a parachute, albeit a rather poor one, helping it to almost hover in the air by slowing its descent.

On the top surface of some flying disc models there are several concentric ridges, which produce turbulence at the leading edge. The significance of the resultant turbulent layer of air is that it follows the contours of the spinning disc. This keeps the low-pressure air immediately above it close to the disc, enhancing the lift force. Without turbulence, the low-pressure air flowing over the disc no longer stays close to the top surface and some of the lift force is lost (Figure 3). Loss of any lift simply means the Frisbee will not fly so far.

The ridges also allow the Frisbee to be thrown at a higher angle of attack, so increasing its aerodynamic capabilities. They change the aerodynamics of the disc in the same way that the seam of a cricket ball enables swing bowling. All these enhancements to the flight of the Frisbee enable it to be thrown much further than a ball at the same initial velocity.

Record throw

The popularity of the Frisbee is such that it has spawned a host of new ways of using it. There are many trick 'shots' worth seeing on sites such as YouTube, and it has also given rise to a popular team sport called Ultimate. Some public parks and sports centres have disc golf courses, where flying discs are thrown with the object of landing them in small baskets. If you want something slightly different, try an Aerobie flying ring, which is the result of some serious aerodynamic calculations. It is in the *Guinness Book of World Records* for the longest throw of 406 metres, it has even been thrown across Niagara Falls and, remarkably, can stay aloft for over 30 seconds — all thanks to some fantastic physics.

Peter Main is on the academic staff of the University of York and is also a member of the Physics Review editorial board.

Our radioactive environment

Derek Jacobs

Our bodies contain some common radioactive elements whose radiation can damage living tissue. Derek Jacobs discusses some origins of this radiation and why we cannot avoid it

Exam links

The terms in bold link to topics in the **AQA**, **Edexcel**, **OCR**, **WJEC** and **CCEA** A-level specifications, as well as the **IB**, **Pre-U** and **SQA** exam specifications.

Ionising radiation is emitted by **radioactive isotopes**, while cosmic radiation causes some elements to become **radioactive**. Radioactive elements have different **half-lives**, which determine their abundances and **activities**. Radiation **doses** depend on the **activity** and **type of radiation** involved.

We are constantly exposed to ionising radiation from radioactive isotopes in the environment and in the materials that make up our bodies. How do we measure the radiation that our bodies receive and emit, and where does it come from?

Radioactivity and the body

Quantifying the effects of radiation on living tissues is complex, because both the energy and the type of radiation have a role. Box 1 summarises the key points and the units used to measure radiation *dose*.

The average annual background dose our bodies receive from background radiation is 2.7 mSv, of which typically half arises from radon gas coming out of the ground. Other exposure, for example in a medical X-ray, increases the dose. If you have a CT scan, the absorbed dose is typically about 20 mSv.

PhysicsReview*Extras*

Get practice-for-exam questions based on this article at

The body's own radioactivity can be measured. As the total radioactivity of a body is likely to be small, the equipment must be thoroughly screened from the ever-present background radiation. The thick armour plating from sunken Second World War warships is unlikely to be contaminated by the radioactive fallout from atmospheric nuclear bomb testing in the 1950s, so it is much sought after as a screening material.

As you will know, alpha particles have a range of a few centimetres in air and just millimetres in the body. This means that alpha emitters in the body are difficult to detect outside it, which is unfortunate because ingested alpha emitters can cause cancers. Sometimes a gamma ray is emitted following the alpha particle emission and, in general, gamma emitters are more easily detected. The gamma rays escaping the body can be detected by the tiny flashes of light (scintillations) they produce when stopped in special materials (for example, sodium iodide), or by the electrical charge they release in some semiconductors, such as silicon.

Figure 1 shows a whole-body monitor. The monitor is calibrated using a phantom body, fabricated so as to contain radioactive material in the appropriate parts of the 'body'. After calibration, the machine can then be used to monitor any extra radiation introduced into the body for medical diagnosis and treatment.

Radioactive materials around us

Radioactive materials occur in the soil, air and water, so it is impossible to avoid them (Table 1). Uranium is 500 times more common than gold and four times more common than copper. It crops up in plants and, typically, we eat 2 μg of it per day. The

Table 1 Naturally occurring radioactive isotopes in the environment

Isotope	Name	Half-life	Comments
^{14}C	Carbon-14	5730 years	Produced naturally
^{40}K	Potassium-40	1.25×10^{9} years	Two decay routes
^{210}Po	Polonium-210	138 days	From decay chain of U — natural
^{226}Ra	Radium-226	1600 years	Longest lived of 33 Ra isotopes — all radioactive
^{232}Th	Thorium-232	1.39×10^{10} years	Remaining since formation of Earth
^{235}U	Uranium-235	8.52×10^{8} years	<1% of natural uranium
^{238}U	Uranium-238	4.5×10^{9} years	>99% of natural uranium

bulk of this is excreted and the rest accumulates in the hair and fingernails.

The Earth is usually taken as 4.5×10^{9} years old. Therefore, if we assume that the chemical composition of Earth is the same as when it was formed, how can ^{14}C, with its very much shorter half-life, still be around? The half-life of ^{40}K is much longer, as are those of ^{232}Th, ^{235}U and ^{238}U, so it is reasonable to have them still present in the Earth.

Assuming that the isotopes of uranium were present in equal amounts at the formation of the Earth, the different half-lives explain why there is only 0.27% of the ^{235}U isotope now. The ratio of these two isotopes, coupled with a knowledge of their half-lives, enables the age of the Earth to be calculated. Other isotopes of uranium — ^{232}U (half-life of 69.8 years) and ^{234}U (245 000 years) — are not in the table because they have decayed away to negligible amounts by now.

The heavy radioactive elements, thorium, actinium and uranium, are the starting points for three chains of decay. The result of each decay is a *daughter* element that, in turn, often decays into a further radioactive element, and so-on until a different, stable isotope of lead (Pb) is finally reached.

Figure 1 A whole-body radiation monitor

A domestic radon testing kit

Box 1 Radiation, activity and dose

Radiation is measured in several ways.

The *activity* of a source is the number of nuclear disintegrations per second. The SI unit of activity is the becquerel (Bq). An old unit, the curie (Ci), is sometimes used:

$1\,\text{Bq} = 1$ disintegration s^{-1}
$1\,\text{Ci} = 3.7 \times 10^{10}\,\text{Bq} = 3.7 \times 10^{10}\,\text{s}^{-1}$

The energy absorbed per kilogram of material is called the *absorbed dose* (sometimes the *physical dose*), measured in a unit called the gray (Gy):

$1\,\text{Gy} = 1\,\text{J}\,\text{kg}^{-1}$

The amount of damage done by the radiation is called the *equivalent dose* (sometimes the *biological dose*) and is measured in units called sieverts (Sv). The equivalent dose depends on the type of radiation. For example, an absorbed dose of alpha radiation causes 20 times more damage than the same absorbed dose of X-rays:

X-rays, β-particles and γ-rays	1 Gy = 1 Sv
Neutrons	1 Gy = 10 Sv
α-particles	1 Gy = 20 Sv

Humans receive about 2 mSv of radiation per year. Most of this radiation is from cosmic rays, radioactive rocks and radon gas.

The extra radiation from nuclear power accounts for about 0.1% of the normal background radiation.

Some ^{14}C is incorporated into plants, making them slightly radioactive

The thorium chain is shown in Figure 2. Four isotopes of radium crop up in these decay chains so, despite the half-life of the longest lived (^{226}Ra) being only 1600 years, it occurs naturally. There is a similar explanation for the existence of ^{210}Po.

In the early years of the twentieth century scientists disentangled many of these decay chains, with the study being regarded as a branch of chemistry because the concept of isotopes was only developed in 1913. Lord Rutherford, regarded as the father of nuclear physics, was awarded the Nobel prize in chemistry in 1908, rather than in physics. He said that he had studied many rapid changes but never one as rapid as his change from physicist to chemist.

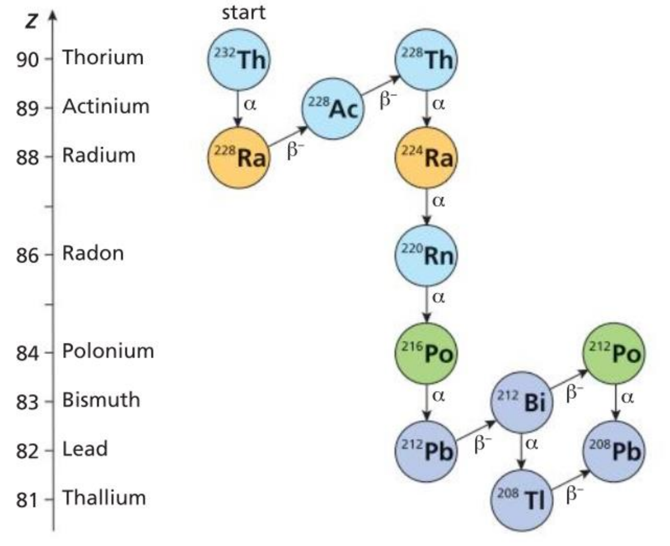


Figure 2 The $^{232}_{90}$Th decay chain

Radon is a very radioactive gas that is produced as radium decays. (Chemically, it is a noble 'inert' gas and is very unreactive.) It can be found in significant amounts in some parts of the UK, especially where granite is the underlying rock. Particularly when combined with cigarette smoke, radon can cause lung cancer and so the design of buildings is modified to reduce the risk in high-risk areas. Mines, too, can be rich in radon if precautions to reduce its levels are not taken.

Radioactive carbon

You probably know that the Earth is bombarded by *cosmic rays* from space. This radiation comprises mainly protons with energies that can exceed 10^{21} eV. In nuclear physics, energies are usually measured in electronvolts ($1\,\text{eV} = 1.60 \times 10^{-19}\,\text{J}$) ('Radiation: not so simple', PHYSICS REVIEW Vol. 27, No. 4, pp. 22–26.) The energy of cosmic rays can be much greater than the 13×10^{12} eV achieved in the Large Hadron Collider at CERN (the European particle physics laboratory near Geneva).

Most of the electrically charged cosmic rays are deflected when they encounter the Earth's magnetic field, but some collide with particles in the upper atmosphere and produce secondary cosmic rays. These secondaries go on to produce more, resulting

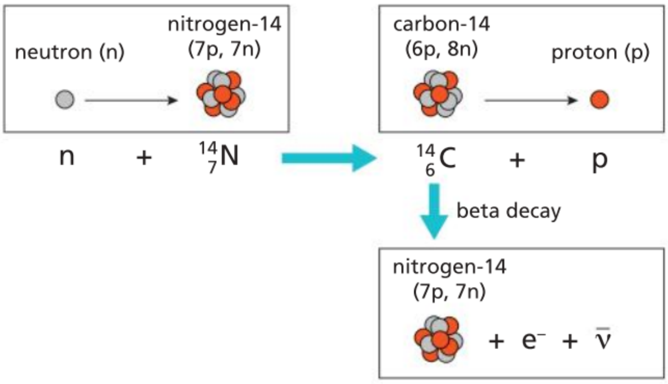


Figure 3 Formation and decay of ^{14}C in Earth's atmosphere

in a cosmic ray shower or cascade, which can cover a large area on the ground. The secondary particles can include positively and negatively charged electrons, muons (particles similar to electrons but heavier) and neutrons. Some of the neutrons interact with nitrogen nuclei in the upper atmosphere and eject a proton, leaving behind a neutron-rich isotope of carbon, ^{14}C, with eight neutrons (Figure 3).

'Everyday' carbon contains 99% ^{12}C nuclei, with six protons and six neutrons, and 1% ^{13}C, with seven neutrons. The ^{14}C isotope is present in trace amounts, and forms CO_2 like ordinary carbon, but the neutron excess means it is radioactive. It decays by emitting an electron (a β-particle) and an anti-neutrino. In atmospheric carbon dioxide, about 1 atom in 10^{12} is the ^{14}C isotope. CO_2 becomes incorporated in plants through photosynthesis, making them slightly radioactive. This means that the porridge, cereal, toast, etc. you had this morning was radioactive. The total activity of this radioactive carbon in our bodies is about 3.7 kBq (Box 1).

In 5730 years, half of the ^{14}C incorporated into plants and products made from them will have decayed. This allows the material to be dated (*radio-carbon dating*), because when living matter dies it ceases to incorporate new ^{14}C.

Radioactive potassium

Because the half-life of ^{40}K is 1.25×10^9 years — a good fraction of the lifetime of the Earth — it is still found in nature. About 120 parts in every million atoms of today's potassium are ^{40}K, and an adult's body has a total activity of about 4.4 kBq arising from it.

Table 2 Some environmental radioactive isotopes arising from human activity

Isotope	Name	Half-life	Comments
^{3}H	Tritium	12.3 years	From nuclear reactors
^{60}Co	Cobalt-60	5.27 years	There are other, shorter-lived, Co isotopes From accidents, e.g. Chernobyl
^{90}Sr	Strontium-90	28.8 years	Chemically similar to calcium From fallout and spent reactor fuel rods
^{131}I	Iodine-131	8.5 days	From atomic bomb tests and accidents
^{137}Ce	Caesium-137	30.2 years	Found in milk and dairy products

Some background radiation is accounted for by atomic bomb testing after the Second World War

Curiously, this isotope can decay in 89% of cases by emitting an electron (beta decay), resulting in the formation of calcium, ${}^{40}_{20}$Ca. In the remainder of the decays, the nucleus reduces its positive charge by one unit by capturing an atomic electron, resulting in the formation of an isotope of argon, ${}^{40}_{18}$Ar, a process known as *electron capture*.

There are about 120 g of potassium present in the body. It is very important as an electrolyte, for carrying electrical signals, as well as being necessary for cell functioning. So do not give up eating foods containing potassium to avoid its radioactivity.

Nuclear accidents and bomb tests

Other radioactive isotopes remain in the soil and water from the fallout following nuclear weapons testing in the atmosphere or from nuclear power plant accidents like Chernobyl, Ukraine (1986) and Fukushima, Japan (2011). After the 400 or more above-ground atomic bomb tests that the USA alone carried out, the strontium-90 (^{90}Sr) levels in baby teeth were found to be 50 times higher than those formed before such tests. This frightening discovery had a lot to do with the Partial Test Ban Treaty signed in 1963.

Some tritium (^{3}H) — an isotope of hydrogen with two neutrons — escapes from nuclear reactors.

Generally speaking, the radiation levels from these isotopes in the environment (Table 2) are much less than those in Table 1 ('Isotopes of hydrogen', PHYSICS REVIEW Vol. 27, No. 4, pp. 6–9).

A final comment

You might be feeling a bit worried about the radiation that your body is exposed to. But remember that most of this background radiation is natural in origin, and animals, including us, have evolved to cope with it over the millennia. They had to because it is inescapable.

Derek Jacobs is a member of the PHYSICS REVIEW editorial board.